

Hannele Niemi
Roy D. Pea
Yu Lu *Editors*

AI in Learning: Designing the Future

OPEN ACCESS



Springer

AI in Learning: Designing the Future

Hannele Niemi • Roy D. Pea • Yu Lu
Editors

AI in Learning: Designing the Future



Springer

Editors

Hannele Niemi
Faculty of Educational Sciences
University of Helsinki
Helsinki, Finland

Roy D. Pea
Graduate School of Education
Stanford University
Stanford, CA, USA

Yu Lu
Advanced Innovation Center for Future
Education, Faculty of Education
Beijing Normal University
Beijing, China



ISBN 978-3-031-09686-0 ISBN 978-3-031-09687-7 (eBook)
<https://doi.org/10.1007/978-3-031-09687-7>

This work was supported by University of Helsinki (7818/31/2018).

© The Editor(s) (if applicable) and The Author(s) 2023. This book is an open access publication.

Open Access This book is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this book are included in the book's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the book's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.

The publisher, the authors, and the editors are safe to assume that the advice and information in this book are believed to be true and accurate at the date of publication. Neither the publisher nor the authors or the editors give a warranty, expressed or implied, with respect to the material contained herein or for any errors or omissions that may have been made. The publisher remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

This Springer imprint is published by the registered company Springer Nature Switzerland AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

Preface and Acknowledgements

This book is entitled *AI in Learning: Designing the Future*. It acknowledges the reality that AI is consequential for societies, organizations, work, and education and that it is becoming more and more interwoven into the cultural activities of everyday life. Artificial intelligence (AI) is changing the world. However, the title also raises the big questions of what is learning with AI, who has the final responsibility for the quality of learning, and who will design the future of learning? AI opens enormous opportunities to education and learning and expands educational settings for learning in and beyond the traditional classroom. However, many innovations are still in their early stages and need much further research and deeper understanding of what the human roles and responsibilities are with respect to AI's integrations into learning environments and educational systems.

For advancing safe and responsible routes to AI in learning and education, the researchers in Finland, the USA, and China have wanted to introduce developments in the latest research on AI in Learning with innovative practices and new solutions. Many chapters provide pedagogical applications and practices demonstrating how to use AI at different levels of education and, in working-life as lifelong learning settings. Cooperation between the three nation's researchers began in a series of joint triangle conferences for Intelligent Digital Tools for Learning and Education, organized at Stanford University in October 2018, the University of Helsinki in February 2019, and Beijing Normal University in June 2019. Thereafter, because of COVID-19, the cooperation has continued virtually.

The book provides cutting-edge research and new scenarios for researchers, companies, policymakers, and all users including teachers and other education stakeholders. It also makes visible that AI has many ethical challenges. The penetration of AI in human life is connected to ethics, security, and human rights and presents important new challenges to research, policymaking, and governance as well as to companies with their AI businesses. Learning and education as fundamental human processes and cultural activities centrally concerned with human values are even more connected with ethical questions than many other more technical applications.

The editors and all authors have contributed significantly to this expansive multidisciplinary and multi-partner journey. Most chapters are based on wide cooperation across disciplines and based on co-work between researchers, technological designers, companies and practitioners, and learners themselves. We take pleasure in extending our sincere gratitude to all participants. Throughout the book preparations, we have been privileged to have valuable practical support in the editing work from Dr. Marianna Vivitsou at the University of Helsinki. Great thanks to Marianna as she has patiently communicated with authors in several rounds of the editors' internal peer-review process and helped authors to keep to their timeline and finalize their chapters.

We also want to thank all funders and supporters—Business Finland as the national funding agency in Finland for financing the AI in Learning project led by Professor Hannele Niemi, the Stanford Institute for Human-Centered Artificial Intelligence, and Sino-Finnish Joint Learning Innovation Institute at Beijing Normal University in China. We also thank the universities, companies, and schools with which authors are affiliated. We are appreciative of the local resources and infrastructure for research on AI which they have contributed.

Helsinki, Finland
Stanford, CA, USA
Beijing, China
January 20, 2022

Hannele Niemi
Roy D. Pea
Yu Lu

Contents

Introduction to AI in Learning: Designing the Future.....	1
Hannele Niemi, Roy D. Pea, and Yu Lu	
1 The Aim and Background of the Book.....	1
2 AI in a Global World: State of the Art	2
3 AI in Education	4
4 The Structure and Contents of the Book	6
5 The Message of the Book	13
References	13

Part I AI Expanding Learning and Wellbeing Throughout Life

Artificial Intelligence Innovations for Multimodal Learning, Interfaces, and Analytics	19
Marcelo Worsley	
1 Introduction	20
2 Prior Literature	20
2.1 Multimodal Learning to Support Twenty-First Century Learning Competencies	21
2.2 Multimodal Interfaces to Facilitate Inclusive Learning	21
2.3 Multimodal Analytics to Enable Novel Measures for Learning	22
3 Multicraft	23
3.1 Overview	23
3.2 Multimodal Learning	25
3.3 Multimodal Interfaces	26
3.4 Multimodal Analytics	27
3.5 Summary	27
4 BLINC	28
4.1 Overview	28
4.2 Multimodal Learning	29
4.3 Multimodal Interfaces	29
4.4 Multimodal Analytics	30
4.5 Summary	31

5	Discussion	31
5.1	Multimodal Learning Deserves Multimodal Assessments	31
5.2	Twenty-First Century Skills Benefit from Twenty-First Century Methods	32
5.3	Be Intentional About Keeping Humans in the Loop	32
5.4	Ethical Considerations	33
6	Conclusion	34
	References	34
Curiosity and Interactive Learning in Artificial Systems.....		37
Nick Haber		
1	Introduction	37
2	AI Successes of the Past Decade	38
3	Artificial Curiosity and Interactive Learning	43
4	Examples of Artificial Curiosity and Interactive Learning	44
5	Artificial Interactive Learning as Models of Human Learning	50
	References	52
Assessing and Tracking Students' Wellbeing Through an Automated Scoring System: School Day Wellbeing Model		55
Xin Tang, Katja Upadyaya, Hiroyuki Toyama, Mika Kasanen, and Katariina Salmela-Aro		
1	Introduction	55
2	The Assessments of Wellbeing	56
3	Artificial Intelligence-Based Wellbeing Assessments and Enhancement...	57
4	School Day Wellbeing Model: A Model for Wellbeing Assessment and Enhancement	59
4.1	Theoretical Foundations for the School Day Wellbeing Model.....	59
4.2	School Day Wellbeing Model	61
4.3	How Does the School Day Wellbeing Model Work?	62
5	Features of the School Day Wellbeing Model	63
6	User Experiences	66
7	The Future Directions	67
8	Conclusion	68
	References	68
Learning from Intelligent Social Agents as Social and Intellectual Mirrors		73
Bethanie Maples, Roy D. Pea, and David Markowitz		
1	Introduction	73
2	Prior Research	74
3	Research Questions	76
4	Method	77
4.1	Replika	77
4.2	Participants and Procedure	77
4.3	Measures and Analysis	78

Contents	ix
5 Results	79
5.1 Participant Profiles	79
5.2 Motivations for Initial Replika Use	80
5.3 Beliefs About Replika	80
5.4 Patterns of Replika Use	81
5.5 Participants' Life Experiences with Replika	82
6 Discussion	85
7 Future Work and Limitations	86
References	87
 An AI-Powered Teacher Assistant for Student Problem Behavior	
Diagnosis	91
Penghe Chen and Yu Lu	
1 Introduction	91
2 Theoretical Framework for System Design	92
2.1 Need Deficiency	93
2.2 Problem Behavior	93
2.3 External Environmental Factors	94
2.4 Individual Factors	95
3 System Design	95
3.1 Diagnosis Module	95
3.2 Question Answering Module	97
3.3 Case Search Module	99
4 Discussion and Conclusion	101
References	102
 Analysis and Improvement of Classroom Teaching Based on Artificial Intelligence	
Zhong Sun, Zi Chun Yu, and Fei Yun Xu	105
1 Introduction	105
2 Literature Review	106
2.1 Classroom Teaching Analysis	106
2.2 Improvement of Classroom Teaching	108
3 Methodology	111
4 Conclusion	118
References	119
 Part II AI in Games and Simulations	
 Perspectives and Metaphors of Learning: A Commentary on James Lester's Narrative-Centered AI-Based Environments	
Marianna Vivitsou	125
1 Introduction	125
1.1 The Key Message of the Keynote and Interview	126
1.2 Key Concepts and Metaphors	126
1.3 Modalities, Narrative, and Metaphors in AI for Learning Purposes ..	127

2	Crystal Island as a Metaphor for Learning with AI.....	129
3	Reversing the Double Narrative Process: The Agency of Students	132
4	Agential Cuts in Narrative-Centered Learning Environments	133
5	Summarizing Remarks	134
	References	135
Learning Career Knowledge: Can AI Simulation and Machine Learning Improve Career Plans and Educational Expectations? 137		
I-Chien Chen, Lydia Bradford, and Barbara Schneider		
1	Introduction	137
2	Making More Informed Career Choices: A Theoretical Framework	138
2.1	Why AI About Career Knowledge?	140
2.2	An Example of Gamified Career Knowledge: Init2Winit, an Overview	140
3	Opportunity for AI and Machine Learning (ML).....	144
3.1	Machine Learning and Decision Trees	145
3.2	Empirical Example: Decision Trees Algorithm in Init2Winit	145
4	Result	147
4.1	Init2Winit Users' Profiles	147
4.2	Init2Winit Users' Classification for Multiple Goals	147
4.3	Alignment Knowledge of Decision Trees and Partition	148
4.4	Regression Decision Trees and Prediction of Educational Expectations	151
5	Strengths and Weakness of Current Design	153
6	Conclusion and Recommendation	153
Appendix A Record of e Daily Active Users		
Appendix B Iterative Dichotomiser 3 (ID3) Algorithm		
	References	156
Learning Clinical Reasoning Through Gaming in Nursing Education: Future Scenarios of Game Metrics and Artificial Intelligence 159		
Jaana-Maija Koivisto, Sara Havola, Henna Mäkinen, and Elina Haavisto		
1	Introduction	159
2	AI in Healthcare Education	161
3	Exploiting AI Through Game Metrics	162
4	A Case Study of CR and the Use of Game Metrics in Nursing Simulation Games	164
5	Directions for Future Work	168
6	Conclusion	169
	References	170

AI-Supported Simulation-Based Learning: Learners' Emotional Experiences and Self-Regulation in Challenging Situations	175
Heli Ruokamo, Marjaana Kangas, Hanna Vuojärvi, Liping Sun, and Pekka Qvist	
1 Introduction.....	175
2 Theoretical Framework	176
2.1 Self-Regulated Learning	176
2.2 Positive and Negative Emotions in Simulation-Based Learning.....	177
2.3 Simulation-Based Learning Situations	179
3 Research Questions	179
4 Method.....	180
4.1 Data Collection.....	180
4.2 Analysis	182
5 Results	183
5.1 Learners' Positive and Negative Emotional Experiences During Simulation-Based Learning	183
5.2 Self-Regulated Learning Operations in Challenging Situations	185
5.3 Toward Developing AI Tutors in Simulation-Based Learning	187
6 Conclusion.....	188
References	189

Part III AI Technologies for Education and Intelligent Tutoring Systems

Training Hard Skills in Virtual Reality: Developing a Theoretical Framework for AI-Based Immersive Learning.....	195
Tiina Korhonen, Timo Lindqvist, Joakim Laine, and Kai Hakkarainen	
1 Introduction.....	195
2 From ITS to AI Tutor	197
2.1 Intelligent Tutoring System (ITS).....	197
2.2 Observability	198
2.3 Modifiability	198
2.4 AI Tutor	199
3 Grounded Cognition	199
4 VR-Native AI Tutor Framework.....	203
4.1 Situated Conceptualizations	203
4.2 Physical Environment Simulation	205
4.3 Specifiers	205
4.4 Learner and Tutor Simulations	206
4.5 Implementing the Framework	206
5 Toward VR-Native Pedagogy	207
5.1 Simulation Environment	207
5.2 Task Sequencing	208
5.3 Scaffolding	208
6 Discussion	210
References	211

Multiple Users' Experiences of an AI-Aided Educational Platform for Teaching and Learning	215
Shuanghong Jenny Niu, Xiaoqing Li, and Jiutong Luo	
1 Introduction	215
2 Study Background and Research Questions	216
3 Description of the AI-Aided SLP Educational Platform	217
4 Methodology	221
5 Findings	222
5.1 Students' Self-Reported Experiences	222
5.2 Teachers' and Principal's Self-Reported Experiences	224
6 Discussion and Learning from This Case Study	226
6.1 Major Functions Favored by Students and Teachers	226
6.2 New Ways and Possibilities in Learning and Teaching	228
6.3 Positive Experiences and Changes	228
6.4 The Importance of Learning Theories Applied in AIED Applications	228
6.5 Continuous Improvements and the Social Nature of Learning	229
7 Conclusion and Recommendation	229
References	230
Deep Learning in Automatic Math Word Problem Solvers	233
Dongxiang Zhang	
1 Introduction	233
2 Methodology and Analysis	236
2.1 Rule-Based Methods	236
2.2 Statistic-Based Methods	236
2.3 Tree-Based Methods	238
2.4 Deep Learning Models	239
2.5 Geometry Problem Solving	242
3 Conclusions	243
References	244
Recent Advances in Intelligent Textbooks for Better Learning	247
Bo Jiang, Meijun Gu, and Ying Du	
1 Introduction	247
2 The Development of Intelligent Textbooks	248
3 Intelligent Tutoring Technologies of Intelligent Textbooks	249
3.1 Domain Modeling Technologies in the Intelligent Textbook	249
3.2 Student Modeling Technologies	252
3.3 Instructional Technologies	253
4 Evaluation of Intelligent Textbooks	255
4.1 Students' Comments on Intelligent Textbooks	255
4.2 The Effectiveness of Intelligent Textbooks	255
5 Discussions and Conclusions	256
References	257

Part IV AI and Ethical Challenges in New Learning Environments

Ethical Guidelines for Artificial Intelligence-Based Learning: A Transnational Study Between China and Finland	265
Ge Wei and Hannele Niemi	
1 Introduction	265
2 AI-Based Learning Needs an Ethical Basis	266
3 Ethics as a Theoretical Concept	268
4 Research Design	269
5 Chinese and Finnish Contexts	270
5.1 AI in Learning and Education in the Chinese Context	270
5.2 AI in Learning and Education in the Finnish Context	270
6 Ethical Guidelines for AI-Based Learning	271
6.1 Inclusion and Personalization	271
6.2 Justice and Safety	273
6.3 Transparency and Responsibility	274
6.4 Autonomy and Sustainability	276
7 Discussion and Conclusions	277
References	280
Artificial Intelligence Ethics from the Perspective of Educational Technology Companies and Schools	283
Päivi Kousa and Hannele Niemi	
1 Introduction	283
2 AI in Education and Learning	284
3 Many Ethical Guidelines and Principles for AI	285
4 Case I: Finnish EdTech Companies' Views on Ethical Challenges	287
5 Case II: Finnish School's Ethical Challenges and Practical Viewpoints on Explicability	289
6 Discussion	291
6.1 Beneficence	292
6.2 Non-maleficence and Justice	292
6.3 Autonomy	293
6.4 Explicability	293
7 Conclusions	294
References	294
Artificial Intelligence in Education as a Rawlsian Massively Multiplayer Game: A Thought Experiment on AI Ethics	297
Benjamin Ultan Cowley, Darryl Charles, Gerit Pfuhl, and Anna-Mari Rusanen	
1 Introduction	298
2 Theoretical Background	299
2.1 Explaining and Evaluating AI	299
2.2 MMOGs, MOOCs and Game-Based Learning	301

2.3	Role of AI in Education	302
2.4	Ethics of AIEd	303
3	Methodology and Analysis	303
3.1	AIEd-MMOG Schematic Technical Definition	304
3.2	Player Models	304
4	Findings	308
4.1	Rawlsian Justice Game	309
4.2	AIEd-MMOG Rawlsian Justice Game	310
4.3	AIEd-RJG for AI Evaluation	311
5	Discussion/Synthesis	311
5.1	Implications	312
5.2	Future Outlook	312
6	Conclusions	313
	References	313

Four Surveillance Technologies Creating Challenges for Education 317

Roy D. Pea, Paulina Biernacki, Maxwell Bigman, Kelly Boles, Raquel Coelho, Victoria Docherty, Jorge Garcia, Veronica Lin, Judy Nguyen, Daniel Pimentel, Rose Pozos, Brandon Reynante, Ethan Roy, Emily Southerton, Miroslav Suzara, and Aditya Vishwanath

1	Introduction	317
2	Surveillance State	318
2.1	Location Tracking	318
2.2	Facial Identification Technologies (FITS)	320
2.3	Automated Speech Recognition	322
2.4	Social Media Mining	323
3	Call to Action	325
3.1	Research	325
3.2	Policy and Law	326
3.3	Practice	326
4	Conclusion	327
	References	328

Reflections on the Contributions and Future Scenarios in AI-Based Learning 331

Roy D. Pea, Yu Lu, and Hannele Niemi

1	Where Are We Now with AI?	331
2	AI Contributions to Different Levels of Education Systems	333
2.1	K-12 Tutoring Systems and Other Adaptive Learning Technologies	333
2.2	Beyond K-12 Disciplinary Curriculum: Whole Child AI Technologies	334
2.3	Higher Education and Lifelong Learning	335
2.4	Enabling Media for the Learning Ecosystem	336
3	Roles of AI for Enhancing the Processes and Practices of Educational Research	337

Contents	xv
4 Advancing the Learning of AI	338
5 Ethical Dimensions of AI Integration into Human Learning Environments and Socio-Technical Systems for Education	339
References	341
Index.....	343

About the Editors and Contributors

About the Editors

Hannele Niemi is Professor and Research Director in education at the University of Helsinki and nominated as UNESCO Chair on Educational Ecosystems for Equity and Quality of Learning 2018–2026. She is also the Chair of the University Board at the University of Lapland (2018). She served as Vice-Rector at the University of Helsinki (2003–2009). She has been invited as honorary doctor or honorary professor in five universities. She has led Finnish national research consortium AI in Learning (2020–2021) <https://blogs.helsinki.fi/ai-in-learning/> that has active cooperation with researchers, companies, and practitioners seeking new solutions how AI can support human learning. The project also has a wide cooperation in China and the USA. She was invited as an education expert in tens of countries. Niemi has been a scientific leader for several large national research projects in Finland, including the Finnable 2020 (finnable.fi) program for advancing educational technology and 21st century skills in schools (2012–2015). She currently serves as an advisor and reviewer for several scientific journals, and she has served as a member of several scientific councils, including the European Science Foundation, the Academy of Finland, and the University of Helsinki, and worked as a reviewer and panel member for research councils in Norway, Portugal, Estonia, and Singapore. She has served as panel member for the evaluations of the quality and effectiveness of more than 10 universities in Europe (2005–2020) and in evaluations of 3 European evaluation councils in Higher Education. Professor Niemi has over 400 publications on teaching, learning, teacher education, and technology-supported learning environments, and she has edited several international books and journal's special issues <https://researchportal.helsinki.fi/fi/persons/hannele-niemi/publications/>

Roy D. Pea is David Jacks Professor of Education and Learning Sciences at Stanford University, Graduate School of Education, and Computer Science (Courtesy). He is also founder and Director of Stanford's PhD program in Learning

Sciences and Technology Design. His studies and extensive publications in the learning sciences focus on advancing theories, research, tools, and social practices of technology-enhanced learning of complex domains. He is co-editor of *Mirrors of Minds: Patterns of Experience in Educational Computing* (1987), *Video Research in the Learning Sciences* (2007), *Learning Analytics in Education* (2018), *Routledge Handbook of the Cultural Foundations of Learning* (2020), and co-author of the National Academy of Sciences book: *How People Learn* (2000) and the 2010 National Education Technology Plan for the US Department of Education. He is a fellow of the National Academy of Education, Association for Psychological Science, the American Educational Research Association, and the Center for Advanced Study in the Behavioral Sciences. In 2004–2005, Roy was President of the International Society for the Learning Sciences. Roy served from 1999 to 2009 as a director for Teachscape, a teacher professional development services company he co-founded with CEO Mark Atkinson. He has been acknowledged as Fellow of American Academy of Arts and Sciences (2019–) and Inaugural Fellow of International Society of the Learning Sciences (2018–).

Yu Lu is Associate Professor with the School of Educational Technology, Faculty of Education, Beijing Normal University (BNU), where he also serves as the Director of the artificial intelligence lab at the advanced innovation center for future education (AICFE). He received the PhD degree from the National University of Singapore. He has published more than 60 academic papers in the prestigious journals and conferences (e.g., IEEE TKDE, TMC, ICDM, AIED, AAAI, CIKM, EDBT, IJCAI, ICDE) and serves as the PC member for multiple international conferences (e.g., AIED, AAAI, EMNLP, CIKM). Before joining BNU, he was a research scientist and principal investigator at the Institute for Infocomm Research (I2R), A*STAR, Singapore. His current research interest sits at the intersection of AI technology and its applications in education.

Contributors

Paulina Biernacki is doctoral student at Stanford University, Stanford Learning Sciences and Technology Design PhD program.

Maxwell Bigman is doctoral student at Stanford University, Stanford Learning Sciences and Technology Design PhD program.

Kelly Boles is doctoral student at Stanford University, Stanford Learning Sciences and Technology Design PhD program.

Lydia Bradford is a third-year PhD student at Michigan State University in Measurement and Quantitative Methods. Her main research interest lies in statistical modeling, computational data analysis, research design, and causal inference.

Lydia currently worked on two science curriculum intervention projects, *Crafting Engaging Science Environments* and *Multiple Literacies in Project-Based Learning*.

Darryl Charles is a senior lecturer of computer science at Ulster University. He is a graduate of Queens Belfast (BEng Electronics), Ulster (MSc Microelectronics), and Paisley (PhD Machine Learning) Universities, UK. His research interests include game-enhanced learning, AI in games, intelligent interactive digital storytelling, and virtual reality-based assistive health technologies.

I-Chien Chen received a PhD in Sociology and worked as a research associate in the College of Education at Michigan State University. Her research is to understand how social contexts, interpersonal relationships, and intervention programs enhance students' and teachers' social-psychological well-being and learning behavior in educational expectations, career pathways, and teaching instruction.

Penghe Chen received a PhD from the National University of Singapore in 2015. He is an assistant professor and principal researcher at the Advanced Innovation Center for Future Education at Beijing Normal University. He has published more than 20 academic papers on artificial intelligence and educational technology. His research interests include educational knowledge graph, educational dialogue system, and data mining.

Raquel Coelho is doctoral student at Stanford University, Stanford Learning Sciences and Technology Design PhD program.

Benjamin Ultan Cowley is Associate Professor of AI in Learning at the Faculty of Educational Sciences, University of Helsinki, Finland. He has a PhD in Computer Science from the University of Ulster, Northern Ireland. Cowley leads HiPerCog group, studying both skilled and impaired performance in cognitively demanding dynamic tasks under uncertainty, using computational cognitive neuroscience methods.

Victoria Docherty is doctoral student at Stanford University, Stanford Learning Sciences and Technology Design PhD program.

Ying Du is a master's student in the Department of Educational Information Technology, East China Normal University (ECNU), Shanghai, China. Her research interests include K-12 computational thinking and intelligent textbooks.

Jorge Garcia is doctoral student at Stanford University, Stanford Learning Sciences and Technology Design PhD program.

Meijun Gu is a master's student in the Department of Educational Information Technology, Zhejiang University of Technology, Hangzhou. Her research interests include intelligent textbook and learning analytics.

Elina Haavisto, PhD, is Professor (Nursing Science) in Tampere University, Faculty of Social Sciences. Her research focuses on seriously ill patients and their families and health care education.

Nick Haber is Assistant Professor at the Stanford Graduate School of Education, and by courtesy, Computer Science. He co-founded the Autism Glass Project, which uses AI and wearable technology in a tool for children with autism. He leads the Stanford Autonomous Agents Lab, which designs artificial intelligence that mimics the ways people learn through interaction and curiosity.

Kai Hakkarainen, PhD, is Professor of Education at the University of Helsinki. He has conducted learning research based on psychology and cognitive science from elementary to higher education. Recently his research has expanded to personal and collective learning processes in communities and networks of experts, knowledge-intensive professional organizations, and academic research communities.

Sara Havola, MNSc, is a doctoral student at the Department of Health Sciences, Tampere University. Her research focuses on nursing students' clinical reasoning skills when using simulation games and game metrics in simulation games.

Bo Jiang, PhD, is Associate Professor at East China Normal University (ECNU), Shanghai, China. Earlier he was Associate Professor at Zhejiang University of Technology, Hangzhou, where he received his PhD degree in 2014. His research interests include educational data mining, learning analytics, and machine learning. He serves as the Editorial Board Member for IEEE Transactions on Learning Technologies.

Marjaana Kangas works as a university lecturer of the Faculty of Education at the University of Lapland. She is an adjunct professor in playful and game-based learning. Her research interests include AI in education, playful and game-based learning, students' and teachers' agency, digital media literacies and competences, out-of-school practices, and creative collaboration.

Mika Kasanen is the Co-founder and CEO of School Day with a driving purpose to make a lasting impact on student well-being, mental health, and social emotional learning (SEL). He received his master's degree in political sciences from the University of Helsinki in 2011. He lives in Helsinki, Finland.

Jaana-Maija Koivisto, PhD, is a principal research scientist at HAMK Smart research unit in Häme University of Applied Sciences. She is also a postdoctoral researcher at the Faculty of Social Sciences, Tampere University. Her research focuses on clinical reasoning skills, serious games, virtual reality, and gamification.

Tiina Korhonen, PhD, is the University Lecturer and Head of Innokas Network (www.innokas.fi/en) at the University of Helsinki. Dr. Korhonen's professional interests lie in the wide landscape of 21st century learning and development of educational practice in the context of the digital society, with special focus on the practical opportunities available through digital tools and processes.

Päivi Kousa (PhD, Science Education) is a teacher educator, researcher, and a project coordinator (AI in Learning) at the Faculty of Educational Sciences at the University of Helsinki, Finland. Her current research focuses on ethical challenges

that schools and EdTech companies have in a context of AI and education. She has specialized in school-company collaboration.

Joakim Laine, M.Ed., is a doctoral student in the doctoral program in school, education, society, and culture (SEDUCE) at the University of Helsinki. Laine is working in Innokas network (www.innokas.fi/en) on various design-based projects that are involved with immersive learning technology. Laine's research interests lie in the facilitation of learning, immersive interfaces, and imagination.

Xiaoqing Li is the Executive Director of Disciplinary Education Laboratory at Advanced Innovation Center for Future Education, Faculty of Education, Beijing Normal University. She received her MEd from Liaoning Normal University. She has more than ten years' experience in working with school leaders and teachers to improve the education quality. Her main research interests are educational technology, big data in education, and education reform with ICTs.

Veronica Lin is doctoral student at Stanford University, Stanford Learning Sciences and Technology Design PhD program.

Timo Lindqvist is the COB and co-founder of Upknowledge, a digital learning company specializing in professional learning. Since 1996, he has advanced training, content creation, and learning management practice in global learning and development organizations. His main interests lie in the development and application of artificial intelligence in the context of professional lifelong learning.

Yu Lu is Associate Professor at Beijing Normal University, where he is the Director of the Artificial Intelligence Laboratory and the Advanced Innovation Center for Future Education. His PhD is in computer engineering from the National University of Singapore in 2012. His research interests include educational data mining, learning analytics, and educational robotics.

Jiutong Luo is Postdoc Fellow at Advanced Innovation Center for Future Education and Center for Educational Science and Technology, Faculty of Education, Beijing Normal University. He received his PhD in Education from the University of Hong Kong. His main research interests are educational technology and psychology, ICTs in education, learning science, and educational neuroscience.

Henna Mäkinen, MNSc, is a project researcher at HAMK Smart research unit in Häme University of Applied Sciences. Her research focuses on virtual reality in healthcare education.

Bethanie Maples is a doctoral candidate at the Stanford Graduate School of Education and a product manager at Google X. She designs embodied artificial intelligence and human-machine interface platforms and systems, with a focus on cognitive development and education. Her master's degree is from Stanford University (2018) and undergraduate degrees are from the University of Auckland, New Zealand.

David Markowitz is Assistant Professor in the School of Journalism and Communication at the University of Oregon. He uses language data to make psychological inferences about people. His work has been published in the *Proceedings of the National Academy of Sciences* and the *Journal of Communication*. His PhD is from Stanford University (2018) and undergraduate and master's degrees from Cornell University.

Judy Nguyen is doctoral student at Stanford University, Stanford Learning Sciences and Technology Design PhD program.

Hannele Niemi, PhD, Professor and Research Director at the Faculty of Educational Sciences, University of Helsinki, Finland. She is also a UNESCO Chair on Educational Ecosystems for Equity and Quality of Learning. She has more than 400 publications on teaching and learning. Many of them focus on learning in digital environments.

Shuanghong Jenny Niu obtained her PhD from the University of Helsinki in the year 2021, and DSc from Aalto University in the year 2009. Her recent publication is on teaching and learning 21st-century competencies. Currently, she is working in the Faculty of Educational Sciences at the University of Helsinki. She is dedicated to the research fields of school leadership and management; teachers' education and training; pedagogical methods; and the development of 21st-century competencies.

Roy D. Pea is David Jacks Professor of Education and Learning Sciences at Stanford University. He is Founder and Director of Stanford's PhD program in Learning Sciences and Technology Design. He is an elected fellow of the American Academy of Arts and Sciences, National Academy of Education, Association for Psychological Science, the American Educational Research Association, and the International Society of the Learning Sciences.

Gerit Pfuhl, PhD, is Professor in Cognitive and Biological Psychology at UiT, the Arctic University of Norway. She graduated in cognitive neuroscience from NTNU, Trondheim, in 2009. She studies rationality and decision-making under uncertainty and is since 2018 member of the ethical committee of the Psychological Science Accelerator.

Daniel Pimentel is doctoral student at Stanford University, Stanford Learning Sciences and Technology Design PhD program.

Rose Pozos is doctoral student at Stanford University, Stanford Learning Sciences and Technology Design PhD program.

Pekka Qvist is User Interface and Usability Engineering Manager in NAPCON, Part of Neste. He is leading R&D for training systems for the process industry, with 20 years of experience in software engineering, and 10 years of expertise in digitalization of training. He has published more than 15 peer-reviewed articles in the fields of simulations, learning analytics, and extended reality.

Brandon Reynante is doctoral student at Stanford University, Stanford Learning Sciences and Technology Design PhD program.

Ethan Roy is doctoral student at Stanford University, Stanford Learning Sciences and Technology Design PhD program.

Heli Ruokamo, PhD, is a professor of media education, a research vice dean of the Faculty of Education, and a director of the Media Education Hub at the University of Lapland. She was Visiting Scholar at Stanford University's School of Medicine and H-STAR institute. She is a member of the Strategic Research Council at the Academy of Finland and a president of the Finnish Educational Research Association.

Anna-Mari Rusanen is a philosopher of artificial intelligence and cognitive sciences, focused on the philosophical foundations of artificial intelligence, computational and algorithmic explanation, and the ethical aspects of algorithmization. She works as a university lecturer in cognitive science (University of Helsinki), and as a senior specialist on AI (Ministry of Finance, Finnish Government).

Katriina Salmela-Aro is Academy Professor at the Faculty of Educational Sciences, University of Helsinki, Finland, and Visiting Professor, Institute of Education, University College London, UK. She received her doctorate in psychology in the University of Helsinki. Her major interest includes motivation, well-being, and educational transitions.

Barbara Schneider is the John A. Hannah University Distinguished Professor in the College of Education and Department of Sociology. A sociologist of education, her major interest is understanding how the social and emotional states and actions of individuals are influenced by the contexts in which they inhabit.

Emily Southerton is doctoral student at Stanford University, Stanford Learning Sciences and Technology Design PhD program.

Liping Sun received her MS in Learning, Education, and Technology from the University of Oulu, 2014. She is a PhD candidate and researcher at the Media Education Hub, Faculty of Education, University of Lapland. Her research interests include digital game-based learning and teaching, primary education pedagogy, media education, collaborative learning, and self-regulated learning.

Zhong Sun is Professor at the Capital Normal University, China. Her PhD is in the Educational Technology Institute from Beijing Normal University in 2008. Her research interests include technology-enhanced teacher professional development, artificial intelligence in education, classroom teaching quality, interaction, and course design in technology-based environments.

Miroslav Suzara is doctoral student at Stanford University, Stanford Learning Sciences and Technology Design PhD program.

Xin Tang, PhD, Docent, is a university researcher at the Faculty of Educational Sciences, University of Helsinki, Finland. He received his doctorate in psychology

from the University of Jyväskylä, Finland in 2017. His research interests include motivation, engagement, social emotional skills (e.g., grit and curiosity), and classroom practices.

Hiroyuki Toyama is a postdoctoral researcher at the Faculty of Educational Sciences, University of Helsinki, Finland. He received his doctorate in psychology from the University of Jyväskylä, Finland in 2018. His research interest includes employees' proactivity and well-being.

Katja Upadyaya, PhD, Docent of Educational Psychology, is a researcher at the Faculty of Educational Sciences, University of Helsinki, Finland. Her research interests include student engagement, academic motivation, and lifelong learning. Currently, she is conducting research on students' situational experiences while learning, socio-emotional skills, and parental burnout.

Aditya Vishwanath is doctoral student at Stanford University, Stanford Learning Sciences and Technology Design PhD program.

Marianna Vivitsou, PhD, is postdoctoral researcher at the Faculty of Educational Sciences, University of Helsinki. Her research and scholarly interests focus on hybrid online pedagogy and the ethical challenges of transformative and sustainable pedagogies.

Hanna Vuojärvi works as University Lecturer of the pedagogy of adult education at the University of Lapland. Her current research focuses on higher education pedagogy, adults' and older people's media literacy education, and developing adult education teachers' studies.

Ge Wei, PhD, is Associate Professor at College of Elementary Education, Capital Normal University, Beijing, China. He is also the Director of the Research Centre for Children and Teacher Education in Capital Normal University. His research focus centers on teacher education and human development. His recent publications appear in *Teaching and Teacher Education* and *Journal of Curriculum Studies*.

Marcelo Worsley is Assistant Professor of Computer Science and Learning Sciences at Northwestern University. His research integrates artificial intelligence and data mining with multimodal interfaces. He directs the technological innovations for inclusive learning and teaching lab which works with community and industry partners around the world to empower people and organizations through novel analytic tools.

Fei Yun Xu studied at the School of Information Technology of Hebei Normal University 2015–2019. He is currently a Master's student at Capital Normal University. His research interests include teacher education and learning analysis. His main research contributions were published in the 25th Global Chinese Conference of Computer in Education and the 13th International Conference on Educational Technology and Computers.

Zi Chun Yu studied at the College of Information Engineering of Capital Normal University from 2016 to 2020. She is currently a master's student at Capital Normal University, in Curriculum and Teaching Theory. Her major research interests include artificial intelligence in education, teacher education, and learning analytics.

Dongxiang Zhang received his PhD in computer science from National University of Singapore in 2012. He is now a researcher in Zhejiang University and was a professor in the University of Electronic Science and Technology of China from 2016 to 2019. His research interests include AI+education and smart city. He has won the rising star award of ACM SIGMOD China.

Introduction to AI in Learning: Designing the Future



Hannele Niemi, Roy D. Pea, and Yu Lu

Contents

1 The Aim and Background of the Book	1
2 AI in a Global World: State of the Art	2
3 AI in Education	4
4 The Structure and Contents of the Book	6
5 The Message of the Book	13
References	13

1 The Aim and Background of the Book

Artificial intelligence (AI) is changing the world radically. It impacts societies, organizations, work, and education, and it is becoming more and more part of everyday life. The surge of AI requires analysis and foresight to determine what it may mean in education and for learning. This book is based on contemporary research with Artificial Intelligence in educational settings (AIED) in educational settings. The major questions are: (1) How is learning changing when human learning and machine learning are connected and what consequences does this conjunction have for education, also for working life as lifelong learning and (2) what kind of ethical issues are emerging with AI in education from the viewpoints of schools and other learning environments. The core aim is to discover how AI-based intelligent tools and environments can augment and support human learning.

H. Niemi (✉)

Faculty of Educational Sciences, University of Helsinki, Helsinki, Finland
e-mail: hannele.niemi@helsinki.fi

R. D. Pea

Graduate School of Education, Stanford University, Stanford, CA, USA
e-mail: roypea@stanford.edu

Y. Lu

Advanced Innovation Center for Future Education, Faculty of Education, Beijing Normal University, Beijing, China
e-mail: luyu@bnu.edu.cn

In this volume, over 60 researchers in universities in China, USA, and Finland have introduced their recent research concerning how they see the potentialities of AI for education and learning. Many authors provide evidence of new applications and consequences. Many chapters also provide reflections on the newest trends in AI development and what kinds of changes they may require in adaptations by schools and working life contexts.

Our authors leap forward to share the ways AI may contribute to redesigning our future when it is applied in education and learning. This introduction has two tasks. It first draws a general picture of the state of the art of AI's role globally and summarizes how education is a fundamental part of these processes. Secondly, the introduction summarizes the contributions of chapters. The book has four parts, each of them giving a special viewpoint to AI with meanings, relevance, and challenges when applying AI in learning and education.

2 AI in a Global World: State of the Art

The definition of AI has been in discussion since its origins. Starting in the 1950s, the core idea of most definitions has been that a machine can be intelligent because it embodies some performance elements that human brains enact such that computer systems can perform tasks that normally require human intelligence (e.g., Stone 2016; Roschelle et al. 2020; UNESCO 2021a). Based on huge developments in technology and computing sciences, we can see that AI has become a more and more complex, cross-subject and cross-disciplinary, multipurpose, global endeavor, and it is in an ongoing development process. The intelligent features have increased with advanced computer programming, for example, through neural networks in deep learning. Still many researchers remark that it is still a long way from achieving the flexibility, width of task performance, and progress in competences to reflect on and give reasons for decisions made that are typical qualities of the human mind. Nonetheless, new technologies have made AI useful for industry and business, health and medicine, transportation, and logistics as well as in many service sectors. AI has brought additional value to design, manufacturing, and products, robotics, chatbots, and automatic mobile device log in and face recognition as typical examples. All the same, researchers still observe that we are only in the spring of AI applications and much more research and development will be needed to achieve the full potential of AI for all the seasons (Stone 2016).

At the policy-making level concerning AI and human affairs, the last 5 years have demonstrated exponential growth. In China, the USA, and the European Union, many strategic plans have been published since the middle of the last decade. The recent trends can be summarized:

- In China, a discussion paper from the McKinsey Global Institute (2017), originally presented at the 2017 China Development Forum, explored AI's potential to fuel China's productivity and growth – and to disrupt the nation's workforce.

- In the USA, the National AI Initiative Act of 2020 (House of Representatives 2020) became law on January 1, 2021, providing for a coordinated program across the entire federal government to accelerate AI research and application for the nation's economic prosperity and national security.
- The European Commission's (2020) white paper on AI sets strategic plans for how European countries will use AI in different sectors of society. AI should work for people and be a force for good in society. The European union has a strong emphasis on ethical issues (Europen Union 2020). AI is recognized for the likelihood that it will change the whole society, and the European documents correspondingly emphasize, in addition to the topic of work disruption due to job automation, key concerns for advancing trustworthy AI and its ethical requirements.

The Organization for Economic Co-operation and Development (OECD) launched 2020 "AI Policy Observatory" (OECD 2019). It tracks policy areas where AI is driving changes in the workforce, transportation, and healthcare sectors. It follows up trends and AI data use and provides a forum for national AI policies and global initiatives of different stakeholders including business, academia, and civil society. The OECD highlights that its observatory project aims to help countries to encourage, nurture, and monitor the responsible development of trustworthy AI systems for the benefit of society. OECD AI principles also recommend governments and the private sector combine their investments for research, including interdisciplinary efforts and development of AI. The future emphasis is that innovations should focus on challenging technical issues and on AI-related social, legal, and ethical implications and policy issues.

In addition to the policy-level strategies, AI is also seen as a tool for sustainable development. In April 2021, the United Nations (UN) published its Resource Guide on Artificial Intelligence AI Strategies (UN 2021). It introduces how AI can provide resources to achieve sustainable development goals (SDGs) that are related to big challenges such as climate change, hunger, poverty, inequalities, and other severe global threats. The volume has collected existing AI-based resources as well as examples from policies, strategic plans, and ethical guidelines of governments, private sectors, and other stakeholders. It also warns that AI will have unanticipated consequences that will exacerbate inequalities and negatively impact individuals, societies, economies, and the environment.

UNESCO, as a United Nations agency, has a special mandate for education and culture. UNESCO published its 2021 AI guidelines for policymakers, introducing and reviewing AI technologies in educational and their ethical challenges (UNESCO 2021a). In November 2021, UNESCO launched a report *Reimagining our futures together: a new social contract for education* (UNESCO 2021b). The report provides a strong appeal for the importance of education and the strategic goals in education (SDG 4) that is one of strategic goals of UN. The report has a strong message: Quality education and access to learning must be guaranteed for everyone and throughout the life course. We need a new social contract to develop education globally. Access to school alone is not enough. Currently, the biggest problem is the quality of education; what and how to learn in schools. Inequality in quality

education is growing exponentially and sustainable development is also based on education. For AI, UNESCO has a double message (Niemi 2020). On the one hand, we need new technology that helps to increase access to education and increase the quality of education. And on the other hand, AI must not increase the digital gap and deepen inequities in education. Based on over 2 years of joint and wide cooperation with its member states, the recommendations for AI policy were adopted by UNESCO's General Conference on 24 November 2021 (UNESCO 2021c). The consensus reaffirms a humanistic approach to the use of AI with a view towards protecting human rights and preparing all people with the appropriate values and skills needed for effective human-machine collaboration in life, learning, and work, and for sustainable development. It advocates for human-controlled and human-centered AI development, where the deployment of AI should be in the service of people and to enhance human capacities. It recommends that the impact of AI on people and society should be monitored and evaluated throughout value chains. The key principles emphasize that digital technologies should aim to support—and not replace—schools. We should leverage digital tools to enhance student creativity and communication. When AI and digital algorithms are brought into schools, we must be vigilant to ensure that they do not simply reproduce existing stereotypes and systems of exclusion.

3 AI in Education

AI is part of fundamental global changes and its power is increasing. Most policy-level strategic plans draw a picture at the global or whole societal level. References to education comes mainly from a perspective of changes in work and new competences needed in working life. Otherwise, education and learning are rather invisible in the policy-level documents. This concerns also ethical principles of AI that set general guidelines for trustworthy AI. Only UNESCO's guidelines and ethical principles have focused directly on education.

However, we have reviews and ongoing research how AI has been implemented in education and learning (Bransford et al. 2006; Niemi 2021). AI has already entered education and schools in different forms. Learning Sciences has published for decades how learning analytics can help to recognize and facilitate learning processes with intelligent tools (Baker and Inventado 2014; Fischer et al. 2020; Niemi et al. 2018). Chen et al. (2020) reviewed research on AI published in education in high-quality international journals between 2009 and 2019. The review provides evidence that AI has been extensively adopted and used in administration, instruction, and learning. In administration, AI applications such as reviewing and grading students' assignments were seen as very useful and, in some cases, even more accurate than human-based assessments. Important implementations were also applications for teachers which help them improve instruction with more knowledge about students' learning and with interactive tools for learners' knowledge construction and sharing. For students' learning, AI could help them by

tutoring and personalization. New technological systems have leveraged machine learning and adaptability, and curriculum and contents can be customized and personalized in line with students' needs. Reviews and analyses of current state of the art (Chen et al. 2020; Stone et al. 2016; Timms 2016; Roschelle et al. 2020) also reveal that a transformation has happened from computer and computer-related technologies to web-based and online intelligent education systems. Often with the use of embedded computer systems but also together with other technologies, we also note the use of humanoid robots and web-based chatbots to perform instructors' duties and functions independently or jointly with instructors. AI-related themes, such as teaching robots, intelligent tutoring systems (ITS), online learning, and learning analytics, have become common over the past several years. In many studies, big data, learning analytics, and data mining techniques have become major tools for personalized learning.

Recent AI technologies provide several options for learning and educational services which can be summarized (UNESCO 2021a; Roschelle et al. 2020):

- Natural language processing (NLP) for the use of AI to automatically interpret texts, including semantic analysis, used in translations, and for generating texts of learning contents, and supporting personalization processes.
- Speech recognition covers the application of NLP to spoken words, including smartphones, and provides AI personal assistants within games and intelligent tutoring systems, and for conversational bots in learning platforms.
- Image recognition and processing employs AI for facial recognition (e.g., for electronic documents and processes in classroom situations), handwriting recognition, text analysis (e.g., to detect plagiarism), image manipulation (e.g., for recognizing deepfakes), and for autonomous scoring and grading.
- Autonomous agents use AI in computer game avatars, software bots, virtual learning spaces, smart robots.
- Affect detection employs AI to analyze sentiment in text, behavior, and faces.
- AI underlies data mining algorithms for predictive learning diagnoses, progress forecasting, socio-emotional well-being analysis, financial predictions, and fraud detection.
- Artificial creativity uses AI in systems that can create new kinds and exemplars of photographs, music, artwork, or stories.

In the last 10 years, AI has taken big steps in education and learning with a new method of computing and advanced technology for using and integrating multimodal data. The multisector expert group (Roschelle et al. 2020) convened by the nonprofit organization Digital Promise drafted scenarios for how AI will influence education. They foresee that AI-based learning goes far beyond what was earlier possible with tracing users' learning paths through keyboard strokes or eye movements in learning analytics. The advanced human-machine interface provides AI-related functions including natural language interaction, speech recognition, and detecting learners' emotions. AI allows sensing, recognizing patterns, representing knowledge, making and acting on plans, and supporting naturalistic interactions with people and support learners with varied strengths and needs, allowing students to use handwriting, gestures, or speech as input in addition to more traditional

keyboard and pointer input. The expert group also sees that AI can support learning in terms of orchestrating complex learning activities with multiple people and resources, augmenting human abilities in learning contexts, expanding naturalistic interactions among learners and with artificial agents. It broadens the competencies that can be assessed and reveals learning connections that are not easily visible. These approaches go beyond familiar design concepts for individualized, personalized, or adaptive learning. All these new opportunities bring many ethical challenges and these should be urgently investigated.

As a conclusion of our brief survey of the current state of art in AI for education and learning, we can see that AI is massively applied already in societies and globally. In education and learning, many advanced techniques are already available, and we have tentatively promising findings (e.g., Niemi 2021). However, the accelerating pace of development of technology expands AI's potentialities in education, so we need extensive new research about educational implementations and their effects on human learning and people's lives. The more AI is applied in education and learning, the more we need reflections on and solid grounds for ethical use of AI.

4 The Structure and Contents of the Book

The book is based on the most recent research on AI in learning and education in Chinese, European, and American contexts. The articles introduce how new intelligent tools and machine learning can support human learning and well-being and what kinds of consequences it has for education and learning environments. The articles provide insights into the state of the art of AI when used in education systems and for learning environments.

The book has four parts:

- (i) AI expanding learning and well-being throughout the life
- (ii) AI in games and simulations
- (iii) AI technologies for education and intelligent tutoring systems
- (iv) AI and ethical challenge in new learning environments

Part I: AI Expanding Learning and Well-Being Throughout Life

The articles cover the methods for how human learning can be supported through AI-based tools and environments in school contexts and informal settings. The articles introduce new methods for how AI-based tools and services can support students' learning and help them to become more engaged, curious, and in positive social-emotional well-being states. Articles also describe how teachers can be assisted by AI-based tools and environments in

(continued)

diagnosing students' behavioral and learning difficulties and how researchers can see more deeply into what is happening in classrooms with multimodal data collection.

Part I starts with the chapter “[Artificial Intelligence Innovations for Multimodal Learning, Interfaces, and Analytics](#)” of Marcelo Worsley. It describes how the twenty-first century has brought a growing variety of authentic and engaging learning environments. The chapter discusses artificial intelligence-based tools and technologies that can help researchers and practitioners navigate and enact these novel approaches to learning, while also providing a meaningful lens for student reflection and inquiry. The chapter includes technologies that offer insights for using audio/video information and resources for studying learner electrodermal activity, and it provides analytic techniques and interfaces for helping researchers collect and analyze different types of multimodal data across contexts.

Nick Haber underlines in his chapter “[Curiosity and Interactive Learning in Artificial Systems](#)” the fact that human learning is interactive, and we learn through curiosity, and we interact with both physical objects and the people around them. This flexible capacity to learn about the world through intrinsically motivated interaction continues throughout life. He asks how we would engineer an artificial, autonomous agent that learns in this way – one that flexibly interacts with its environment, and others within it, in order to learn as humans do. The chapter first motivates this question by describing important advances in artificial intelligence in the last decade, noting ways in which artificial learning within these methods are and are not like human learning. Nick Haber also gives an overview of recent results in artificial intelligence aimed at replicating curiosity-driven interactive learning. Finally, he speculates on how AI that learns in this fashion could be used as fine-grained computational models of human learning.

In the chapter “[Assessing and Tracking Students' Well-Being Through an Automated Scoring System: School Day Well-Being Model](#)”, the research group Xin Tang, Katja Upadyaya, Hiroyuki Toyama, Mika Kasanen, and Katariina Salmela-Aro introduces the model for automated scoring system for modelling students' well-being. Students' well-being is critical as it influences their positive development in school life and ensures their future growth. The assessment of well-being has been often static, lagging behind for diagnostic and intervention purposes. In this research, the authors introduce an automated scoring well-being system, School Day Well-Being Model, that is featured as dynamic and real time. User experiences are collected to show the utility of the model. The findings were consistent across the globe.

In the chapter “[Learning from Intelligent Social Agents as Social and Intellectual Mirrors](#)”, Bethanie Maples, Roy D. Pea, and David Markowitz introduce the concept of Intelligent Social Agents (ISAs) which are conversational agents that leverage emergent machine learning techniques to present as sufficiently anthropomorphized

to pass Turing tests in short exchanges. The interaction capabilities of these agents made possible by advances in artificial intelligence lead to deep emotional bonding with users, leading researchers to reexamine the impact and potential uses of these human-machine relationships in education. In this work, they examined the technical advances that made a new breed of ISA possible, and dive into how one best-in-class ISA, Replika, might be affecting users socially, emotionally, and cognitively. A small, mixed-method study of Replika users explored relationships between user loneliness, use motivations, use patterns, and user outcomes. Their results seem to indicate that the confluence of new functionality, product narrative, and user life stressors make ISAs an emerging tool for cognitive and emotional support, filling a gap in users' needs which humans do not fill.

Penghe Chen and Yu Lu describe in their chapter "[An AI-Powered Teacher Assistant for Student Problem Behavior Diagnosis](#)" a novel interactive technology to diagnose students' behavioral difficulties in schools. The chapter describes the process of designing and implementing an intelligent teacher assistant, which could advise teachers and help them to diagnose the student problem behavior. Technically, it utilizes a task-oriented dialogue system to help identify the underlying reasons (i.e., the student need deficiency) behind their problem behaviors, and accordingly provides advice to teachers. It also employs the semantic search technology to find the similar cases that have been well resolved by the experienced teachers.

In the chapter "[Analysis and Improvement of Classroom Teaching Based on Artificial Intelligence](#)", Zhong Sun, Zi Chun Yu, and Fei Yun Xu discuss on classroom research and how new AI-based techniques can improve our understanding what happens in classrooms. Common classroom teaching analysis, which focuses on counting and coding teacher-student behaviors and discourse interactions, faces many difficulties as content-free, low efficiency, and small scale in analysis. To overcome the shortcomings of recent research methods, and to foster high-quality classroom teaching, they propose a human and AI technologies blended analysis framework named as TESTII for classroom teaching. It consists of five steps identifying teaching events, sequencing the pedagogies of classroom teaching structure, analyzing teacher-student interaction, interpreting teaching meaning, and providing improvement strategies for high-quality classroom teaching.

Part II: AI in Games and Simulations

This part introduces cross-scientific and multi-method research with cases, pedagogical models for artificial intelligence-supported gaming and simulation-based learning. It starts with an interview of Professor James Lester on narrative-centered learning environments which can be designed as engaging games for students.

In chapter, “[Perspectives and Metaphors of Learning: A Commentary on James Lester’s Narrative-Centered AI-Based Environments](#)” is a special chapter by Marianna Vivitsou. It is based on Professor James Lester’s keynote presentation of narrative-centered learning environments. The commentary aims to discuss perspectives on narrative-centered learning and metaphors of AI-based learning. The chapter focuses on the narrative elements that underlies the use of AI in Learning. One example of such environments is Crystal Island, an AI-based game for K-12 students learning science. Vivitsou uses Paul Ricoeur’s narrative and metaphor theories to reflect on the role of characters and the narrative plot in relation to Lester’s visualization of the future of learning with AI-based technologies. In this process, new roles in AI-based learning are introduced. One such example is the role of drama manager. The drama manager is a novel metaphor in game-based learning. In addition, more conventional metaphors, such as the tutorial dialogue, are brought forward as well as technological metaphors. The multiplicity of metaphors have agency at their core. As technological advancement shakes the boundaries of thinking about agency nowadays, new dynamic metaphors are needed in AI-based learning. Toward this direction, the commentary draws from new materialist and post-humanist thinkers to raise these issues and the need to take the narrative further.

In the chapter “[Learning Career Knowledge: Can AI Simulation and Machine Learning Improve Career Plans and Educational Expectations?](#)” I-Chien Chen, Lydia Bradford, and Barbara Schneider introduce a game simulation for young adults and those who have lost their jobs. In these life situations, the employment landscape is characterized by ambiguity and insecurity. They introduce the game Init2Winit which integrates data-based analytics with occupational information algorithms that allows users to make choices with respect to their education planning and salary projection in visualizing themselves in a dream job. Their results show promise in terms of the prediction accuracy of educational expectations and users’ behavioral classifications. Init2Winit can be an informational channel for students who lack informal networks in career planning. It also serves as a supplementary network supporting career/ college planning knowledge for students to make better education and employment decisions. Beyond this, the authors propose that machine learning could incorporate a game designed to measure students’ strengths and weaknesses to give career recommendations and pathways.

In the chapter “[Learning Clinical Reasoning Through Gaming in Nursing Education: Future Scenarios of Game Metrics and AI](#)”, the research group Jaana-Maija Koivisto, Sara Havola, Henna Mäkinen, and Elina Haavisto introduce how healthcare professionals can improve their clinical reasoning through AI and how AI techniques can be used in healthcare education and training. Previously simulation games have been proven effective for learning clinical reasoning skills. However, game metrics have not been utilized much in nursing simulation games, although research in other disciplines shows that game metrics are suitable for demonstrating learning outcomes. This chapter discusses the possibilities to exploit game metrics in developing adaptive features for nursing simulation games, especially difficulty adoption based on students’ knowledge and skills. Personalization and adaptivity in simulation games can enable meaningful learning experiences and enable nursing

students to achieve good CR skills for their future work in constantly challenging clinical situations.

In the chapter “[AI-Supported Simulation-Based Learning: Learners’ Emotional Experiences and Self-Regulation in Challenging Situations](#)”, Heli Ruokamo, Marjaana Kangas, Hanna Vuojärvi, Liping Sun, and Pekka Qvist explore learners’ emotional experiences and self-regulation (SRL) and how to overcome stressful situations in a simulation-based learning environment (SBLE). In the experiment, data was collected from the trainees of a basic training phase at Oil Company Neste by online observations, video recordings, and delayed stimulated recall interviews. The findings evidence that SBLE was generally a positive experience to the learners. However, the trainees met several challenging situations with topics related to chemical engineering and process operation. These tasks were often experienced as stressful, and emotional regulation was needed. The trainees used the following SRL operations: metacognitive monitoring, social scaffolding, cognitive operations, and emotional regulation. According to the results, an AI tutor can provide help for decision-making and visualizing critical points of learning processes.

Part III: AI Technologies for Education and Intelligent Tutoring Systems

This part focuses on new systems in which AI technology is used for professional training situated in virtual reality (VR). The articles also describe VR-based learning technology for contextual learning and how scaffolding can be provided by an AI Tutor within VLE. Automatic scoring and e-books are also introduced as tools for improve teaching and learning.

In the chapter “[Training Hard Skills in Virtual Reality: Developing a Theoretical Framework for AI-Based Immersive Learning](#)”, the research group Tiina Korhonen, Timo Lindqvist, Joakim Laine, and Kai Hakkarainen develops a theoretical frame for pedagogical settings for an immersive virtual reality-based hard-skills training guided by an artificial intelligence software agent. They suggest the theoretical assumptions of embodied, embedded, enacted, and extended (4E) cognition to fully consider learner epistemology in a virtual world, and to account for and make full use of the unique opportunities afforded by the synthetic nature of the immersive virtual learning environment. They outline a theoretical framework for a virtual reality AI tutor and propose pedagogical principles for such a framework that could inform follow-on research.

The chapter of Shuanghong Jenny Niu, Xiaoqing Li, and Jiutong Luo “[Multiple Users’ Experiences of an AI-Aided Educational Platform for Teaching and Learning](#)” provides new knowledge for how AI technology can be used to assist in teaching and learning at schools through The Smart-Learning Partner (SLP) educational platform. This learning environment is based on AI technology to provide new possibilities for individualized learning and more educational resources. The chapter introduces a case study of how the AI-aided SLP platform helped in teaching

and learning from students', teachers', and a principal's perspectives at a Chinese school. The platform provided them with diagnostic feedback and assessments, and information about the learning progress. In addition, students had access to various microlectures according to their interests. Teachers got real-time learning reports. They could follow progress at the individual or class level and adjust better their teaching according to students' needs. The principal used the information in resource allocating and in curriculum planning.

In the chapter "[Deep Learning in Automatic Math Word Problem Solvers](#)", Dongxiang Zhang introduces a new innovative automatic solver for mathematical word problems (MWPs) dated early back to the 1960s. Revolutionary advances of deep learning (DL) have opened new ways to parse the human-readable word problems into machine-understandable logical expressions. The problem is challenging due to the existence of a substantial semantic gap. The chapter introduces various attempts that have been made to bridge the gap, from rule-based pattern matching to semantic parsing with statistical machine learning, and to the recent end-to-end deep learning (DL) models. Despite the great success achieved by applying DL models to solve MWPs, the current status in this research domain still has room for improvement. MWPs have also been recognized as good testbeds to evaluate the intelligence level of agents in terms of natural language understanding and automatic reasoning. The successful solving of MWPs can benefit online tutoring significantly.

The chapter "[Recent Advances in Intelligent Textbooks for Better Learning](#)" by Bo Jiang, Meijun Gu, and Ying Du emphasizes that understanding how people read and interact with e-textbooks could not only promote our understanding of how people learn, but also benefit us in providing intelligent learning support to learners. This chapter offers a state-of-the-art overview of intelligent textbooks. It introduces the history of intelligent textbooks and describes the technologies behind these books and what mechanism makes a textbook intelligent. The analysis consists of student modeling approaches from three aspects: the learners' knowledge state model, the learners' learning behavior model, and the learners' psychological characteristic model. The chapter also describes domain modeling technologies. The chapter also summarizes what effects intelligent textbooks provide to students' learning. The last section discusses the future and challenges of intelligent textbooks.

Part IV: AI and Ethical Challenges in New Learning Environments

This part overviews ethical challenges from Chinese and European perspectives. It also opens up the complex picture of ethical challenges from teachers' and companies' perspectives. Games and their algorithms include many ethical questions about transparency and explicability, and these will be reflected upon through a multiplayer game simulation. The part includes also a serious message of risks if AI is used for surveillance.

In the chapter “[Ethical Guidelines for Artificial Intelligence-Based Learning: A Transnational Study Between China and Finland](#)”, Ge Wei and Hannele Niemi have reviewed ethical guidelines in China and in Europe where Finland is one member state. The chapter, taking China and Finland as two contextual cases, analyzes how AI-related policies at the national level have focused on educational themes and set aims for improving the quality of learning and education. The references to education are mainly general and indirect, but four themes for AI ethics in education emerged: (1) inclusion and personalization, (2) justice and safety, (3) transparency and responsibility, and (4) autonomy and sustainability. Although both China and Finland recognize the importance of AI ethics, the differences are manifested as policy approaches, properties, and strategies due to sociocultural variation. The authors emphasize the importance of international and transnational dialogue from ethical perspectives to foster our reciprocal understanding of AI and the human-centered stance on education.

In the chapter “[Artificial Intelligence Ethics from the Perspective of Educational Technology Companies and Schools](#)”, Päivi Kousa and Hannele Niemi discuss opportunities and challenges that AI is bringing to learning in schools and working life contexts. Ethical issues are viewed from the perspectives of companies who produce educational AI-based tools and services, and from those who use them in schools and workplaces for learning. From companies’ viewpoints, ethical challenges are related to regulations, equality and accessibility, machine learning, and society. From schools’ perspectives, the major critical questions are who has the power to decide which educational services the school can use and who is responsible for the ethical issues of those services, for example, student privacy. In addition, schools are concerned with how to ensure that AI-based services and tools are equally accessible to all and genuinely useful in supporting teaching and learning.

The chapter “[Artificial Intelligence in Education as a Rawlsian Massively Multiplayer Game: A Thought Experiment on AI Ethics](#)” by Benjamin Ultan Cowley, Darryl Charles, Gerit Pfuhl, and Anna-Mari Rusanen reflect on the deployment of Artificial Intelligence as a pedagogical and educational instrument, and the challenges that arise to ensure transparency and fairness to staff and students. They apply a Rawlsian justice game, played within the Massively Multiplayer Game: to facilitate transparency and trust of the algorithms involved, without requiring algorithm-specific technical solutions to, for example, “peek inside the black box.” The chapter suggests solutions for the well-known challenges of explainable AI and distributive justice.

The Part IV of ethical issues of AI ends with the chapter “[Four Surveillance Technologies Creating Challenges for Education](#)” by Roy D. Pea and doctoral students of Stanford’s Learning Sciences and Technology Design PhD program: Paulina Biernacki, Maxwell Bigman, Kelly Boles, Raquel Coelho, Victoria Docherty, Jorge Garcia, Veronica Lin, Judy Nguyen, Daniel Pimentel, Rose Pozos, Brandon Reynante, Ethan Roy, Emily Southerton, Miroslav Suzara, and Aditya Vishwanath. They summarize four core surveillance technologies that are entering as common practices to universities as well as preK-12 schools: Location Tracking, Facial

Identification, Automated Speech Recognition, and Social Media Mining. The authors make several critical questions about how these technologies are shaping human development and learning and how current algorithmic biases increase inequities. They also emphasize that the need for learners' critical consciousness concerning their data privacy should be taken as a serious task in education. All these challenges need collaboration of government, industry and the public sector.

The final chapter “[Reflections on the Contributions and Future Scenarios in AI-Based Learning](#)” by Roy D. Pea, Yu Lu, and Hannele Niemi summarizes the importance of the contribution of all chapters and how they deepen our understanding of what possibilities and challenges exist when AI is applied in education. Seven categories provide perspectives to reflections. Four of them are connected to different levels of the educational system, others are opening scenarios to research on education and learning with AI, and finally the last category is devoted to ethical challenges of AI in education and learning. AI will be the powerful tool in education and learning but ethics of AI in education is a keystone issue which will ramify throughout future inquiries into the future of AI-augmented learning.

5 The Message of the Book

The book is based on interdisciplinary cooperation. Technology and human learning in educational settings are integrated. The book provides examples of the most recent AI research at the nexus of computing sciences, learning sciences, and educational technologies. Much is going on – yet longitudinal studies of emerging and long-term effects are very much needed to understand the dimensions of societal change that education and learning transformed by AI will reveal. The chapters point to the future and give evidence that AI will have significant consequences for education and learning. The book opens up inquiries into how AI supports both students and teachers through interactive, intelligent tutoring, multimodal data and feedback systems incorporating speech, images, and other behavioral data. Many challenges are ethical and related to trustworthy AI and issues of equity in AI applications such as face recognition, games and simulations, personalizing learning, and data mining. It is evident that we will collectively need to continue to develop and report research-based evidence for designing the future toward the benefits of all individuals and their societies.

References

- Baker, R.S., & Inventado P.S. (2014). Educational data mining and learning analytics. In J. Larusson, & B. White (Eds.), *Learning Analytics* (pp. 61–75). New York, NY: Springer. https://doi.org/10.1007/978-1-4614-3305-7_4

- Bransford, J., Stevens, R., Schwartz, D., Meltzoff, A., Pea, R., Roschelle, J., Vye, N., Kuhl, P., Bell, P., Barron, B., Reeves, B., & Sabelli, N. (2006). Learning theories and education: Toward a decade of synergy. In P. A. Alexander & P. H. Winne (Eds.), *Handbook of educational psychology* (pp. 209–244). Mahwah, NJ: Lawrence Erlbaum Associates Publishers.
- Chen, L Chen, P. Z. Lin (2020) Artificial Intelligence in Education: A Review. *IEEEAI Access.* 8, 75264-75278. <https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=9069875>
- European Commission. (2020). *A European approach to artificial intelligence*. Retrieved from <https://digital-strategy.ec.europa.eu/en/policies/european-approach-artificial-intelligence>
- European Union (2020). *European framework on ethical aspects of artificial intelligence, robotics and related technologies*. Retrieved from: [https://www.europarl.europa.eu/RegData/etudes/STUD/2020/654179/EPRS_STU\(2020\)654179_EN.pdf](https://www.europarl.europa.eu/RegData/etudes/STUD/2020/654179/EPRS_STU(2020)654179_EN.pdf)
- Fischer, C., Pardos, Z. A., Baker, R. S., Williams, J. J., Smyth, P., Yu, R., Slater, S., Baker, R., & Warschauer, M. (2020). Mining big data in education: Affordances and challenges. *Review of Research in Education*, 44(1), 130–160. <https://doi.org/10.3102/0091732X20903304>
- House of Representatives (2020). *Report 116–617*. Retrieved from <https://www.congress.gov/116/crp/116hrpt617/CRPT-116hrpt617.pdf#page=1210>
- McKinsey Global Institute (2017). *Artificial intelligence: Implications for China, April 2017, discussion paper*. Presented at the 2017 China development forum. Retrieved from <https://www.mckinsey.com/-/media/mckinsey/featured%20insights/china/artificial%20intelligence%20implications%20for%20china/mgi-artificial-intelligence-implications-for-china.pdf>
- Niemi, D., Pea, R., Saxberg, B., & Clark, R.E. (2018). (Eds.). *Learning Analytics in Education*. Charlotte, NC: Information Age Publishing (Published in Chinese Translation, 2020).
- Niemi, H. (2020). Artificial intelligence for the common good in educational ecosystems. In *Humanistic futures of learning: Perspectives from UNESCO Chairs and UNITWIN Networks* (pp. 148–152). Paris: Ediciones UNESCO. Retrieved from <https://researchportal.helsinki.fi/en/publications/artificial-intelligence-for-the-common-good-in-educational-ecosys>
- Niemi, H. (2021) AI in learning: Preparing grounds for future learning. *Journal of Pacific Rim Psychology*, 15.<https://doi.org/10.1177/18344909211038105>
- OECD (The Organisation for Economic Co-operation and Development) (2019) OECD.AI. Policy observatory. OECD AI Principles overview. <https://oecd.ai/en/ai-principles>
- Roschelle, J., Lester, J., & Fusco, J. (Eds.) (2020). *AI and the future of learning: Expert panel report*. Digital Promise. Retrieved from <https://circls.org/reports/ai-report>
- Stone, P., & 16 Study Panel members (2016). *Artificial intelligence and life in 2030, one hundred year study on artificial intelligence report of the 2015*. Stanford, CA: Stanford University. Retrieved from https://ai100.stanford.edu/sites/g/files/sbiybj9861/f/ai100report10032016fnl_singles.pdf
- Timms, M.J. (2016). Letting artificial intelligence in education out of the box: Educational cobots and smart classrooms“, *Int. J. Artif. Intell. Edu.*, 26, (2), 701-712. <https://link.springer.com/content/pdf/10.1007/s40593-016-0095-y.pdf>
- UN (United Nations). (2021). *Resource Guide on Artificial Intelligence (AI) Strategies*.<https://sdgs.un.org/documents/resource-guide-artificial-intelligence-ai-strategies-25128>
- UNESCO (United Nations Educational, Scientific and Cultural Organization). (2021a). *AI and education: guidance for policy-makers* (2021): Paris. <https://unesdoc.unesco.org/ark:/48223/pf0000376709>
- UNESCO (United Nations Educational, Scientific and Cultural Organization). (2021b). *Reimagining our futures together: a new social contract for education*. Paris. <https://unesdoc.unesco.org/ark:/48223/pf0000379707>
- UNESCO (United Nations Educational, Scientific and Cultural Organization). (2021c). *Report of the Social and Human Sciences Commission (SHS)*.UNESCO. General Conference, 41st, 2021Item 8.2, the Ethics of Artificial Intelligence, Adopted. 2411.2021. 41 C/73Annex. Document code:41 C/73. Paris.<https://unesdoc.unesco.org/ark:/48223/pf0000379920.page=14>

Open Access This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.



Part I

AI Expanding Learning and Wellbeing

Throughout Life

Artificial Intelligence Innovations for Multimodal Learning, Interfaces, and Analytics



Marcelo Worsley

Contents

1	Introduction	20
2	Prior Literature	20
2.1	Multimodal Learning to Support Twenty-First Century Learning Competencies	21
2.2	Multimodal Interfaces to Facilitate Inclusive Learning	21
2.3	Multimodal Analytics to Enable Novel Measures for Learning	22
3	Multicraft	23
3.1	Overview	23
3.2	Multimodal Learning	25
3.3	Multimodal Interfaces	26
3.4	Multimodal Analytics	27
3.5	Summary	27
4	BLINC	28
4.1	Overview	28
4.2	Multimodal Learning	29
4.3	Multimodal Interfaces	29
4.4	Multimodal Analytics	30
4.5	Summary	31
5	Discussion	31
5.1	Multimodal Learning Deserves Multimodal Assessments	31
5.2	Twenty-First Century Skills Benefit from Twenty-First Century Methods	32
5.3	Be Intentional About Keeping Humans in the Loop	32
5.4	Ethical Considerations	33
6	Conclusion	34
	References	34

M. Worsley (✉)
Northwestern University, Evanston, IL, USA
e-mail: marcelo.worsley@northwestern.edu

1 Introduction

One hallmark of the twenty-first century has been an expansion in the places where meaningful learning takes place. While many discussions of learning had primarily been confined to traditional classrooms and other formal spaces, recent work has reemphasized the important learning that takes place outside of traditional learning settings (Barron and Bell 2015; Pinkard 2019; Vossoughi and Bevan 2014). Some of these spaces involve after-school enrichment programs, open-ended science laboratories, community-based learning experiences, and makerspaces. These spaces can provide learners with authentic and locally situated learning experiences. They can also be used to facilitate learning of a broader set of competencies: critical thinking, collaboration, communication, and creativity, for example. These and other twenty-first century skills have received increased recognition as essential for addressing future societal needs. For example, much research has been conducted to study learner development of twenty-first century skills (Dede 2009), the 4Cs (critical thinking, communication, collaboration, and creativity), and soft skills (Touloumako 2020). These additional learning contexts and constructs represent important advances in the educational experiences available for today's learners. However, supporting these new types of learning and contexts introduces significant challenges for both learners and educators. Whereas researchers and practitioners have spent decades developing learning experiences and associated measures for competencies like literacy and numeracy, these new contexts and competencies necessitate further research and development. Fortunately, recent advances in the low-cost multimodal sensors can be used to foster new forms of interaction and novel approaches for studying learning that might enable our ability to study, measure, and support these new contexts and competencies.

This chapter will explore the use of multimodal technologies to simultaneously support student learning in nontraditional learning environments and study student learning of these newly emphasized constructs. Two recently developed platforms, Multicraft (Worsley et al. 2021c) and BLINC (Building Literacy in In-Person Collaboration) (Worsley et al. 2021a) will be used to demonstrate how to integrate multimodal interfaces and analytics in K-12 and higher education settings. Each platform supports learners as they practice relatively newly recognized competencies and include a host of multimodal analytics. The two platforms also allow for users to engage in multimodal interactions that utilize speech, eye gaze, tangible blocks, electroencephalography, body pose, and/or facial expressions.

2 Prior Literature

Before moving into a discussion of each platform, this chapter will highlight some pertinent prior research in multimodal learning, multimodal analytics, and multimodal interfaces.

2.1 Multimodal Learning to Support Twenty-First Century Learning Competencies

Within this chapter, we will refer to multimodal learning as being associated with experiences that allow users to (1) engage in learning relevant concepts and ideas through a variety of modalities (e.g., images, videos, text, embodied experiences) and (2) demonstrate their knowledge using a combination of modalities (e.g., speech, written text, drawings, gestures, physical artifacts). The idea of multimodal learning has been a guiding principle within the hands-on, project-based, makerspace, and embodied cognition communities. At the same time, prior research has frequently coupled learning twenty-first century skills, with hands-on, collaborative learning environments that are often supported by computational tools and interfaces. Simply put, many of these contexts emphasize skills of real-world, collaborative problem-solving that are difficult to replicate within a traditional, individual-oriented learning experience. For instance, the process for learning collaboration typically necessitates working in close contact with other individuals and is often situated around a specific unifying real-world problem. Students interact with one another using text, speech, physical artifacts, and gestures, in either colocated or remote settings, for example. Frequently, the means for assessing learning is embedded within the artifact or project that the team creates as opposed to being determined by a written or verbal exam. In summary, attention to learning as multimodal is in alignment with previous calls for epistemological pluralism, equity, accessibility, and inclusion. More generally, researchers have documented the shortcomings of not allowing learners to explore a full set of modalities within a given learning scenario, and the problems with limiting the modalities students are permitted to use to demonstrate their knowledge or learning (Kress 2001; Worsley et al. 2021b).

2.2 Multimodal Interfaces to Facilitate Inclusive Learning

While multimodal learning experiences need not occur through digital technologies, artificial intelligence-enabled multimodal interfaces are becoming an increasingly common strategy for supporting naturalistic interactions between humans and computers (Martinez-Maldonaldo et al. 2017). These interfaces use things like speech-recognition, gesture recognition, and eye tracking, for example, to intelligently interpret the user's intended action. Near the turn of the century, researchers became increasingly intrigued by opportunities to interact with computers using a wide variety of modalities (e.g., speech, eye gaze, gesture, and pen) that typically require some level of artificial intelligence to determine user intent based on an individual modality, or a combination of modalities. Significant decreases in the cost and availability of these multimodal technologies, coupled with the relatively high accuracy of these new tools, fueled considerable advancements in both hardware

and software for capturing and analyzing multimodal data. Developments in video game technology were particularly important contributors to this growth as many computer science researchers explored opportunities to implement multimodal interfaces using the Nintendo Wiimote, Xbox Kinect sensor, and Oculus Rift, for example. The Xbox Kinect sensor included a microphone array for collecting directional audio (to determine who is talking), a depth camera (to estimate object distances), skeletal tracking for up to six individuals (to detect body poses and gestures), and open-source libraries to program the sensors. More recently, researchers have created algorithms that can realize many of those capabilities using a standard web camera, which provides immense opportunities for innovative, low-cost, multimodal interfaces. Researchers and developers create these multimodal interfaces with differing objectives. At times, the interfaces are created to promote accessibility, while in other instances they are developed to enable users to complete their desired tasks more easily. Some common interfaces that feature speech and/or gesture-based input include the smart home technologies available in Amazon Alexa and Google Home, and the touchscreens that are standard within smartphones, tablets, and computers.

2.3 Multimodal Analytics to Enable Novel Measures for Learning

Alongside novel developments in multimodal interfaces, researchers are also developing novel ways to use multimodal data to assess student learning. This specific area of scientific inquiry is called Multimodal Learning Analytics (MMLA) (Blikstein and Worsley 2016; Worsley et al. 2016, 2021b) and refers to ways that multimodal data and computational tools can be employed to model and represent learning within a given environment. The need to study complex learning environments is among the driving motivations for establishing this subfield of learning analytics. Researchers frequently utilize modalities of video, audio, eye gaze and electrodermal activity to look for patterns and forms of interaction that may be hard to identify using traditional learning assessments or through human observation. Additionally, research in MMLA is often concerned with constructs of communication (Ochoa and Dominguez 2020; Ochoa et al. 2018), collaboration (Cukurova et al. 2018; Schneider and Pea 2015; Worsley et al. 2021a), critical thinking (Di Mitri et al. 2020; Oviatt et al. 2015), and creativity (Schneider and Blikstein 2015; Worsley and Blikstein 2018). Across these studies, researchers focus on the combination of audio, gesture, and human-technology interactions to advance theory about collaborative problem solving, communication, creativity and more. MMLA encompasses a broad set of analytic techniques that involve differing levels of human-machine collaboration. In some cases, MMLA analyses involve applying computational techniques to human labelled data. In other cases, researchers might utilize the output from one or more machine learning classifiers to draw inferences

about human learning. In other instances, the analyses may almost exclusively be conducted using machine learning. The unifying perspective across these types of analyses is the realization that multimodal data is essential for supporting the types of inferences that researchers wish to make, and that computational techniques can assist them in providing interpretations of the learning experience.

Prior studies in multimodal learning, multimodal interfaces, and multimodal analytics have individually spurred meaningful contributions to the research community. However, seldom has research from these different areas been integrated with one another. For example, much of the prior work on multimodal learning has tended to rely on traditional measures of student learning. Similarly, work on multimodal interfaces has principally looked at the quality of the user experience, but rarely considered using that same multimodal data to support rich analytics about student learning. Finally, multimodal learning analytics has tended to focus on analyzing data and only seen a select few projects that involve simultaneously using multimodal interfaces together with multimodal analytics. Instead, the multimodal technology has typically only been used to capture data. Intersecting these different areas likely represents the future of learning technologies. This book chapter will describe two examples of tools that sit at the intersection of these three areas. The first, Multicraft, is a multimodal interface for Minecraft that supports collaboration, creativity, computational thinking, and spatial reasoning. The second, BLINC (Building Literacy in In-Person Collaboration) is a platform that uses AI to support real-time collaboration in active learning classrooms, and includes rich, context-specific collaboration analytics. The sections to follow describe each platform in detail and outline their connections to multimodal learning, multimodal interfaces, and multimodal analytics.

3 Multicraft

3.1 Overview

Multicraft is a multiplayer experience for Minecraft that allows for various types of multimodal input. Minecraft is a virtual sandbox game where users can individually or collaboratively design and create buildings, cities, and entire worlds. The platform is sometimes described as a virtual reality space for Legos that has been augmented with some computer programming functionality. Figure 1 includes a picture of a Minecraft world collaboratively created by youth that consists of various puzzles and games. Figure 2 shows a professionally created world that replicates significant portions of Florence, Italy. This particular world aims to allow youth to explore Florence through an interactive virtual reality experience.

Within the current version of the Multicraft platform, users can interact with Minecraft using speech, gestures, eye gaze, tangibles, and even electroencephalography (EEG). The platform was developed to support children with disabilities to

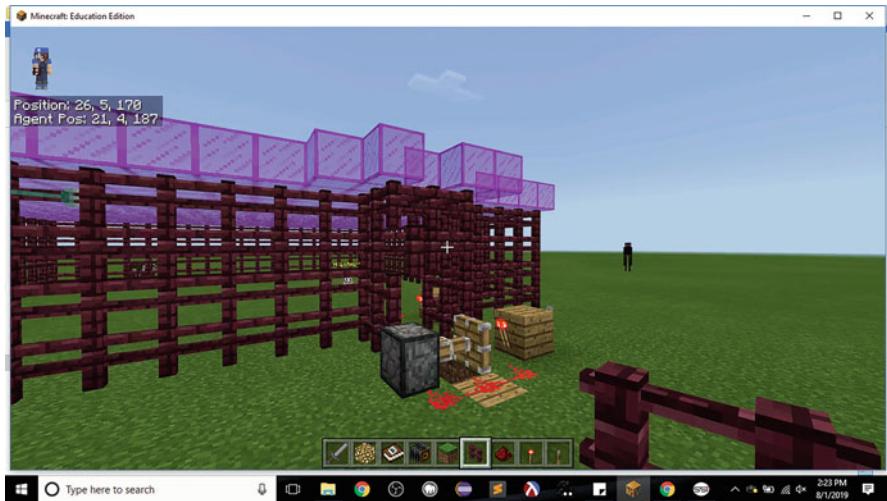


Fig. 1 Picture of Minecraft world created by youth



Fig. 2 Picture of professionally created Minecraft world that replicates Florence, Italy



Fig. 3 An early prototype of the tangible interface used within Multicraft

equitably participate in the Minecraft learning experience. Figure 3 shows an early prototype of the tangible interface used within Multicraft (Bar-El et al. 2018).

3.2 *Multimodal Learning*

As previously noted, Multicraft is designed to be utilized in conjunction with Minecraft, a virtual learning and gaming environment that is popular among youth. The Minecraft learning space allows users to practice several important competencies. Some of these competencies include creativity, problem-solving, spatial reasoning, and computational thinking. Furthermore, it provides the type of virtual world where youth can naturally, and collaboratively, interact with phenomena that connect to any number of disciplines. For example, youth can use Minecraft to create the logic for a computer or use it to create entire cities. Furthermore, the platform is designed to effectively engage and support relative novices, while also being sufficiently generative to allow experts ample opportunities to engage with complex concepts and interactions.

Another hallmark of Minecraft is the opportunity for participants to collaboratively mine, craft, and build within the same virtual world. For example, a group of friends could enter a shared Minecraft world and collectively work on designing a sustainable city over the course of several weeks. Within the game environment, participants are encouraged to communicate with one another through in-game chat,

and control virtual avatars that can interact with one another. Furthermore, educators and computer scientists have developed hundreds of free publicly available lessons that include design challenges, virtual field trips, and more traditional STEM content. These affordances come together to position Minecraft as a learning platform that can advance various twenty-first century competencies.

3.3 Multimodal Interfaces

From a multimodal interface perspective, Minecraft was originally designed to be played with a keyboard and mouse, or a standard gaming controller. In many youth classrooms, it is common to see players use one hand to control the keyboard and the other hand to control the mouse. The Multicraft platform augments the keyboard and mouse-based input, to also include speech, eye gaze, EEG, gestures, and tangibles. Users can select which modalities they wish to employ to complete a given action. An important design principle for Multicraft, however, is to do more than simply replace the existing input modalities using multimodal interfaces. Instead, the platform aims to foster equitable play and leverages computer programming to accelerate some aspects of the gameplay experience. For example, users can say “build a five by ten by eight wood structure here” and Multicraft can utilize a combination of speech recognition, natural language understanding, and eye tracking to instantly build the desired structure where the user is looking. The platform also includes block-based, tangibles input in which a user, or group of users, can manipulate wooden blocks and have their design uploaded to the game in real-time. The tangible block-based input is accomplished using computer vision and relies on a combination of contour detection and color-based tracking. Recent prototypes of the platform also include use of simple hand gestures and EEG. Both approaches are based on machine learning algorithms that can be trained for user-specific gestures or brain activity. The data used to identify hand gestures are from a standard web camera. The EEG data comes from the Muse S headband and includes features from participant brain wave activity. Broadly speaking, Multicraft includes a wide collection of modalities to encourage participants to engage in gameplay using the modalities that best suit them.

These different modalities are important for fostering more equitable and inclusive gameplay and are being researched for their ability to also facilitate improved spatial reasoning and computational thinking. As an example, prior research in spatial reasoning suggests that using spatial language can be a meaningful way to improve spatial reasoning. By encouraging participants to talk to the game using spatial language, we hope to leverage this finding in ways that will result in significant improvements in spatial reasoning. The tangible-based input modality can also confer learning of spatial reasoning. Namely, the use of wooden blocks that exist within the material world, and that are subsequently translated into a 2D representation of the 3D world, can support learners as they practice this process of translating between 2D and 3D representations. Hence, the incorporation of

a multimodal interface can substantively contribute to the goals of multimodal learning of new competencies. Additionally, as we see in the next section, analytics can also help expand how we think about these different competencies and support researchers as they identify and chronicle learner growth with these competencies.

3.4 Multimodal Analytics

The wealth of multimodal data available through Multicraft is also instrumental in supporting analyses of student learning. As an example, this research project includes several hours of data from participants as they engage in Minecraft-focused summer camps and after-school programs. One way for researchers to more tractably navigate human analysis is through the use of computational analyses. Worsley and Bar-El (2019) used log data from the Multicraft server, together with screen recordings of user gameplay, to determine segments in which learners with differing spatial reasoning performance, significantly differed in their in-game interactions. Using this reduced set of data, the authors were able to surface some novel spatial reasoning practices. Worsley and Bar-El describe various ways that students use a combination of explicit and implicit attentional anchors to support the building process within Minecraft. Using eye tracking data, researchers have also highlighted ways that students may practice common spatial reasoning skills within Minecraft, such as perspective-taking and constructing mental representations. At the same time researchers also proposed some spatial reasoning practices that are unique to virtual environments, some of which are based on combinations of well-documented spatial reasoning practices (Andrus et al. 2020). One such practice was error checking, which combines aspects of constructing mental representations and perspective-taking. This project has also used eye tracking data to investigate spatial reasoning practices and identify eye tracking behaviors of learners that exhibit differential performance on common spatial reasoning tasks. Many of these insights are made possible because of the combination of a generative, multimodal learning environment, the utilization of multimodal interfaces, and the computational tools for analyzing data across different modalities.

3.5 Summary

Multicraft is an example of a platform which highlights some of the possibilities for connecting across multimodal learning, multimodal interfaces, and multimodal analytics. Each of these areas is central to the goals and implementation of the platform. Furthermore, the three approaches are integrated to support one another. The next section will present an example designed for the higher education context.

4 BLINC

4.1 Overview

Collaboration is among the most regularly discussed competencies for learners to develop. However, learning institutions seldom offer their students explicit instruction in how to collaborate, or meaningful data around how they are collaborating. A primary goal of the BLINC platform is to provide students with useful insights about how they are collaborating within different contexts. This is achieved by giving users real-time information about how a collaboration is progressing. At a high level, this includes data about how much the group is talking, asking questions, or remaining silent, and the relative distribution of talk among different participants (Fig. 4). The data also includes tracking of user-specified keywords and sentiment classes (Fig. 4). The interface also includes a searchable history of spoken utterances that users can look through for reference. Finally, users can look at discussion content across all groups within the same view and get a summary of verbal contribution frequencies (see Fig. 5).

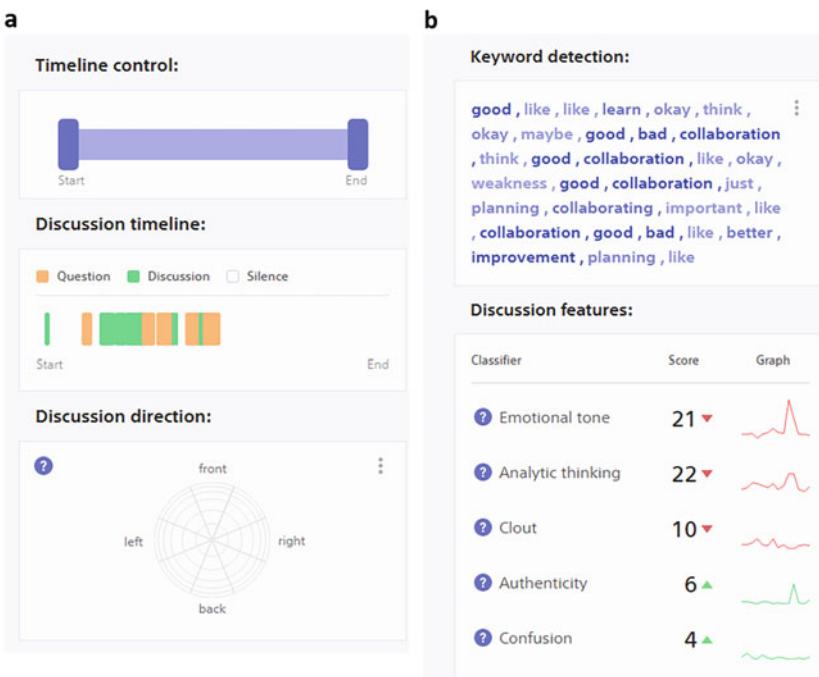


Fig. 4 (a) View from BLINC that shows timeline control, portions of questions, discussion, and silence, and the Discussion direction components. (b) View from BLINC that shows keyword detection and sentiment analysis

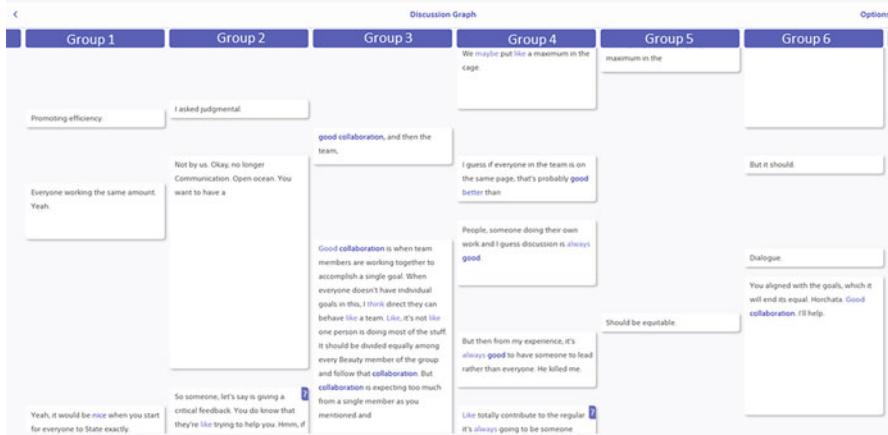


Fig. 5 View from BLINC that shows discussion content for six groups simultaneously

4.2 Multimodal Learning

The BLINC platform was developed amidst growing interest in active learning within institutions of higher education. The term active learning describes a learning environment that contrasts the common practice of learners passively sitting through lectures (Lombardi et al. 2021). Instead, active learning spaces are typified by small group discussions, student-teacher interaction, and limited lecturing. Engaging students in this way can have substantive benefits for student knowledge construction, collaboration, communication, and various other skills that receive significantly less emphasis in traditional lecture-based courses. While this approach is grounded in formative theories from the education research community, instantiating and supporting these types of active learning experiences can present challenges to students and instructors. Instructors may struggle to know how best to support their students within such a format, as it can be difficult to simultaneously have a clear window into all of the small group discussions. At the same time, it can be difficult for learners to get constructive and contextualized feedback from a faculty member who leads a class of more than 50 students. BLINC addresses these challenges through the use of multimodal technologies.

4.3 Multimodal Interfaces

Whereas Multicraft includes a host of multimodal input devices, BLINC primarily uses audio, with an option for video-based input. Users primarily interact with the BLINC system using a web browser which provides them with password-protected access to their current and previous collaboration sessions. Within the

current implementation, audio from collaboration sessions can be captured using two different types of devices. The first is a commercial microphone array called the ReSpeaker Core v2.0. The ReSpeaker includes six microphones to capture audio from up to 5 meters away from the device. The audio capture can be augmented with video from a USB web camera. The BLINC system can accommodate any number of different types of microcomputers through an API that exposes the necessary components for facilitating encrypted data transfer between the microcomputer and the BLINC backend. The second mode for data capture is the microphone from a standard, web-enabled smartphone. Users can access the BLINC webpage and enter a join code for the current discussion. This will subsequently allow them to include their smartphone as one of the audio data collection devices for the group discussion. This feature is particularly salient for higher education contexts where students regularly collaborate outside of class sessions.

In terms of additional interfaces, the platform includes various customizable visualizations and data representations that can support participant sensemaking around their data. The specific time ranges can be adjusted using a slider, and nearly all of the visualizations provide drill down capabilities that take the user to the underlying text associated with a given data point or data segment.

4.4 Multimodal Analytics

The various capabilities offered through the BLINC platform are heavily dependent on multimodal analytics. Even though most of the data being analyzed comes through a single modality (i.e., audio), computational tools and techniques allow for that data to be transformed into several meaningful data points. This section will outline some of those capabilities.

The analytic pipeline begins with the collection of multichannel audio. Each of the six microphones captures audio from the surrounding area. That multichannel audio is used to compute the direction of arrival based on differences in the amount of time it took for a given utterance to reach each of the different microphones. The audio data subsequently undergoes speech recognition. Speech recognition translates from audio into text. The text is later used for various text processing tasks. BLINC also includes speaker diarization. Speaker diarization provides an estimation of who said each utterance. The utterances are labelled with generic titles (e.g., Speaker 1, Speaker 2, etc.). While the platform can support direction of arrival to an accuracy of 20–30 degrees, speaker diarization offers an important augmentation in settings where participants are not stationary, and when users are collecting data through their smartphones. The results from speech recognition also include timestamps on a per utterance basis, and estimated punctuation. Both pieces of information are useful in quantifying the distribution of talk among different team members and the relative timing and distribution of silence, questions, and discussion. As previously noted, the primary output from speech recognition is an estimated transcript of what group participants said. That transcript is used to

support keyword detection. For example, in a class on educational technology, an instructor could specify a collection of keywords: creativity, innovation, technology, ethics, and data. The system would annotate each utterance containing one of those keywords and keep a count of each keyword that appears in the transcript. Furthermore, the system has integrated topic modeling (McCallum 2002). Users can, provide a custom set of documents to train a course- or context-specific topic model and subsequently use that model to examine and chronicle how much group discussion aligns with the different topics. It can also represent how groups are transitioning between the different topics.

4.5 Summary

The BLINC platform sits on top of several computational techniques for analyzing and extracting meaning from audio. While audio is the primary modality, the platform finds several ways to deconstruct that data into useful insights for learners and educators. In so doing, the platform fills an important practical gap of supporting active learning in large enrollment classes and allowing users to explore their collaboration literacy outside of the classroom. Hence, the platform aims to bring together the need for collaborative, active learning, the challenge of facilitating such learning, and the opportunities for utilizing multimodal data and analytics in ways that can support researchers, learners, and educators.

5 Discussion

Multicraft and BLINC provide a glimpse of potential innovations that integrate multimodal learning, interfaces, and analytics. Each platform provides tangible benefits for both users and researchers. At the same time, the pair of projects also highlight a few commonalities that are described in the subsequent sections.

5.1 *Multimodal Learning Deserves Multimodal Assessments*

The design of Multicraft and BLINC are both informed by the realities of new types of learning experiences. BLINC is designed to support collaborative learning environments where students are actively engaged in discussions with their peers and the course instructors. BLINC also supports student collaboration in out-of-school contexts, through the “bring your own device” (BYOD) feature. Both features speak to the idea of students engaging in what we are loosely calling multimodal learning. Similarly, Multicraft, or Minecraft more broadly, is a virtual learning environment where players can collaboratively engage in hours of creative

designing, mining, crafting, and exploring. While researchers have looked at these types of learning environments through traditional assessments and constructs, those constructs fail to do justice to the types of learning and competencies that the spaces make available. Furthermore, asking students to learn and practice material through a variety of modalities, and subsequently restricting assessments to a single modality represents a contradiction to the design and motivation of multimodal learning experiences.

5.2 Twenty-First Century Skills Benefit from Twenty-First Century Methods

Some of the competencies supported through BLINC and Multicraft include collaboration, communication, spatial reasoning, and computational thinking. Researchers have explored various methods for studying these, with many relying on traditional techniques from quantitative and qualitative research traditions. These have been beneficial in furthering our understanding of these constructs, but part of what we see with these two platforms is the need for novel methods for examining these different skills. For Multicraft, while we could administer a typical mental rotation test, such a test becomes highly decontextualized and lacks authenticity and contextual validity. Instead, leveraging computational techniques from eye-tracking data, for instance, can surface the visual spatial anchors that participants may use as part of the building process. Similarly, EEG data might highlight aspects of student concentration and focus that go undetected using most traditional tests and analytic approaches. In the case of BLINC, the platform can support temporal and group-level inferencing about how a group is collaborating. This goes well beyond what one might get from simply having participants complete pre- and post-tests about their collaboration preferences, for example.

5.3 Be Intentional About Keeping Humans in the Loop

A final unifying idea to discuss with regard to Multicraft and BLINC is their intentionality in keeping humans in the loop. Many discussions of artificial intelligence gravitate towards fully automated systems that seemingly replicate human reasoning. Neither Multicraft nor BLINC follow this paradigm. Instead, the platforms reflect inclusion of human decision-making and inference throughout their design and use. They are also intentional about avoiding explicit prescriptions or labelling of individuals and make an effort to present data in context. Many of these approaches are most readily apparent in BLINC. First, the BLINC platform includes considerable customization that can cater the data representations to the specific keywords that the students or instructor wish to focus on, for example. BLINC also avoids generating prescriptions or recommendations around an ideal collaboration

style. For instance, the data representations concerning verbal contributions do not include suggested target values. Instead, instructors and participants are encouraged to use the data in conjunction with their knowledge of the specific learning context and group. This combination of information can help them reflect upon and modify their collaboration practices. Additionally, the ability to drill down into the specific utterances that underlie the visualizations means that humans have an opportunity to interrogate the representations and determine which pieces of data necessitate significant user action. In these ways, these systems aim to simultaneously take advantage of the power of artificial intelligence and the complex reasoning patterns that humans exhibit. Certainly, as society moves into scenarios where people are practicing and evaluating new competencies, it will be beneficial to leverage both of these forms of intelligence, or as Doug Engelbart would say, to “co-evolve” human-computer intelligent systems.

5.4 Ethical Considerations

As society continues to explore the various innovations that might be had through integrating multimodal learning, interfaces, and analytics, it is important to touch on some ethical considerations that can be used to protect participants. Worsley, Martinez-Maldonado, and D’Angelo (Worsley et al. 2021b) include a detailed discussion of 12 core MMLA commitments that span the research pipeline. Their discussion outlines commitments related to data collection, data analysis, and data dissemination. Most salient under the idea of data collection is being circumspect and transparent about what multimodal data is being collected and providing ways for participants to control when that data is being collected. Within the data analysis portion, two commitments that stand out are related to thorough, consistent, and transparent data modeling, and creating opportunities for participants to provide feedback and reflection within the data analysis process. Broadly speaking these two commitments aim to minimize researcher or algorithmic bias. Finally, with regard to dissemination, the authors argue for researchers to develop multimodal systems that provide tangible benefits to research participants. This commitment is not intended to undercut the overall value of research, but to instead advocate for researchers to embark on studies that can potentially confer meaningful benefits to participants, whenever possible. Researchers and designers of multimodal systems should elevate the needs of users. Moreover, the field must carefully consider how this work might feasibly be integrated into ecological settings and how it might scale from classrooms, to schools, to entire districts. These points of integration cannot merely be about the technologies, but must also center ethics.

6 Conclusion

Artificial Intelligence is quickly becoming an integral part of our lived experiences. From speech recognition to computer vision and natural language processing, AI is poised to make a significant impact on the future of learning. One particularly impactful point of integration could be in bridging among multimodal learning, multimodal interfaces, and multimodal analytics. This chapter explored some examples that effectively merge these three areas in ways that support student learning of novel competencies. Notwithstanding, this chapter suggests that truly fomenting student growth in these newly dubbed competencies may require expanding the modalities and analytic techniques that researchers employ.

References

- Andrus, B., Bar-el, D., Msall, C., Uttal, D., & Worsley, M. (2020). Minecraft as a Generative Platform for Analyzing and Practicing Spatial Reasoning. *Spatial Cognition XII*.
- Bar-El, D., Davison, L., Large, T., & Worsley, M. (2018). Tangicraft: A Multimodal Interface for Minecraft. *20th International ACM SIGACCESS Conference on Computers and Accessibility*.
- Barron, B., & Bell, P. (2015). Learning environments in and out of school. *Handbook of Educational Psychology*, 323–336.
- Blikstein, P., & Worsley, M. (2016). Multimodal Learning Analytics and Education Data Mining: using computational technologies to measure complex learning tasks, 3(2), 220–238. <https://doi.org/10.18608/jla.2016.32.11>
- Cukurova, M., Luckin, R., Millán, E., & Mavrikis, M. (2018). The NISPI framework: Analysing collaborative problem-solving from students' physical interactions. *Computers & Education*, 116, 93–109. <https://doi.org/10.1016/j.compedu.2017.08.007>
- Dede, C. (2009). Comparing Frameworks for "21st Century Skills," 1–16.
- Di Mitri, D., Schneider, J., Trebing, K., Sopka, S., Specht, M., & Drachsler, H. (2020). Real-Time Multimodal Feedback with the CPR Tutor. In I. I. Bittencourt, M. Cukurova, K. Muldner, R. Luckin, & E. Millán (Eds.), *Artificial Intelligence in Education* (pp. 141–152). Cham: Springer International Publishing.
- Kress, G. (2001). Multimodal teaching and learning: The rhetorics of the science classroom. Retrieved from <https://doi.org/10.1002/9781405198431.wbeal0815/full>
- Lombardi, D., Shipley, T. F., Bailey, J. M., Bretones, P. S., Prather, E. E., Ballen, C. J., ... Docktor, J. L. (2021). The Curious Construct of Active Learning. *Psychological Science in the Public Interest*. <https://doi.org/10.1177/1529100620973974>
- Martinez-Maldonado, R., Buckingham-Shum, S., Schneider, B., Charleer, S., Klerkx, J., & Duval, E. (2017). Learning Analytics for Natural User Interfaces. *Journal of Learning Analytics*, 4(1), 24–57. <https://doi.org/10.18608/jla.2017.41.4>
- McCallum, A. K. (2002). MALLET: A Machine Learning for Language Toolkit. Retrieved from <http://mallet.cs.umass.edu>
- Ochoa, X., & Dominguez, F. (2020). Controlled evaluation of a multimodal system to improve oral presentation skills in a real learning setting. *British Journal of Educational Technology*.
- Ochoa, X., Dominguez, F., Guamán, B., Maya, R., Falcones, G., & Castells, J. (2018). The RAP System: Automatic Feedback of Oral Presentation Skills Using Multimodal Analysis and Low-cost Sensors. In *Proceedings of the 8th International Conference on Learning Analytics and Knowledge* (pp. 360–364). New York, NY, USA: ACM. <https://doi.org/10.1145/3170358.3170406>
- Oviatt, S., Hang, K., Zhou, J., & Chen, F. (2015). Spoken interruptions signal productive problem solving and domain expertise in mathematics. *ICMI 2015 – Proceedings of the 2015*

- ACM International Conference on Multimodal Interaction, 311–318. <https://doi.org/10.1145/2818346.2820743>
- Pinkard, N. (2019). Freedom of movement: Defining, researching, and designing the components of a healthy learning ecosystem. *Human Development*<https://doi.org/10.1159/000496075>
- Schneider, B., & Blikstein, P. (2015). Unraveling students' interaction around a tangible interface using multimodal learning analytics. *Journal of Educational Data Mining*, 7(3), 89–116. Retrieved from <https://jedm.educationaldatamining.org/index.php/JEDM/article/view/JEDM102>
- Schneider, B., & Pea, R. (2015). Does Seeing One Another's Gaze Affect Group Dialogue? A Computational Approach. *Journal of Learning Analytics*, 2(2), 107–133. <https://doi.org/10.18608/jla.2015.22.9>
- Touloumakos, A. K. (2020). Expanded Yet Restricted: A Mini Review of the Soft Skills Literature. *Frontiers in Psychology*. <https://doi.org/10.3389/fpsyg.2020.02207>
- Vossoughi, S., & Bevan, B. (2014). *Making and tinkering: A review of the literature*. National Research Council Committee on Out of School Time STEM. Washington, DC, DC.
- Worsley, M., Abrahamson, D., Blikstein, P., Grover, S., Schneider, B., & Tissenbaum, M. (2016). Situating Multimodal Learning Analytics. *International Conference for the Learning Sciences 2016*, 2, 1346–1349.
- Worsley, M., Anderson, K., Melo, N., & Jang, J. (2021a). Designing Analytics for Collaboration Literacy and Student Empowerment. *Journal of Learning Analytics*. Retrieved from <https://northwestern.box.com/s/lssb45coayr4h76e09gxk6ktj1qhka6p>
- Worsley, M., & Blikstein, P. (2018). A multimodal analysis of making. *International Journal of Artificial Intelligence in Education*, 28(3), 385–419. <https://doi.org/10.1007/s40593-017-0160-1>
- Worsley, M., Martinez-Maldonado, R., & D'Angelo, C. (2021b). A New Era in Multimodal Learning Analytics: Twelve Core Commitments to Ground and Grow MMLA. *Journal of Learning Analytics*, 1–18. <https://doi.org/10.18608/jla.2021.7361>
- Worsley, M., Mendoza, K., Mwiti, T., Zhen, M., & Jiang, M. (2021c). Multicraft: A multimodal interface for supporting and studying learning in Minecraft. In *2021 Human Computer Interaction International*.

Open Access This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.



Curiosity and Interactive Learning in Artificial Systems



Nick Haber

Contents

1 Introduction	37
2 AI Successes of the Past Decade	38
3 Artificial Curiosity and Interactive Learning	43
4 Examples of Artificial Curiosity and Interactive Learning	44
5 Artificial Interactive Learning as Models of Human Learning	50
References	52

1 Introduction

If we were to distill the learning that we see in a child’s playroom into a computer program, how would we? We might start by describing essential properties – the “engineering specifications” of childhood learning. Early childhood learning is incredibly interactive (Fantz 1964; Gopnik et al. 1999; Begus et al. 2014; Goupil et al. 2016; Twomey and Westermann 2018). Children play, grabbing and manipulating objects, learning about the properties and affordances of their worlds. Their learning is both autonomous and social. They engage in incredibly complex self-play, yet they also learn from demonstration and imitation (Tomasello et al. 1993; Tomasello 2016). Further, their behavior is curiosity-driven, satisfying not only instrumental needs, but also intrinsic motivations to understand and control (Kidd et al. 2012; Dweck 2017). In engaging in these activities, they build powerful, general representations about their worlds, including those that give them a sense of *intuitive physics* (Spelke 1985) and *intuitive psychology* (Colle et al. 2007; Woodward 2009).

While we know a great deal about childhood learning, our knowledge falls far short of being able to engineer this sort of learning within an artificial system. While Artificial Intelligence (AI) has advanced dramatically in recent years, how it learns

N. Haber (✉)
Stanford University, Stanford, CA, USA
e-mail: nhaber@stanford.edu

is in many ways different from how children learn. Most artificial systems do not learn from this messy, interactive, social process, but rather on carefully curated datasets or large amounts of experience from simple, limited environments. Most AI learning behaviors are driven by handcrafted motivations.

Yet in recent years, artificial intelligence has grown increasingly inspired by the flexible, robust learning seen in childhood. Developmental psychology and AI have grown increasingly interwoven through attempts to replicate these sorts of learning processes (Smith and Slone 2017). This interweaving is hoped to be of mutual benefit for both fields. Advances in our understanding of human learning should help us build these sorts of artificial systems. In turn, the enterprise of trying to build curious, interactively learning artificial systems helps refine the questions we ask of our own cognition. Further, if we are successful in this engineering endeavor, AI may be able to serve as precise computational models of our learning.

In this chapter, I endeavor to describe recent works in the artificial intelligence of curiosity and interactive learning, as well as potential payoffs, to a broad audience within education and the learning sciences. I will begin by outlining two exemplar AI successes of the early 2010s: what they accomplished, ways that they reflect human learning, and ways in which they differ. I will then describe several recent results aimed at closing this gap. Lastly, I will speculate on how these efforts might benefit psychology, education, and the learning sciences, with a focus on their potential for modeling our learning in early childhood.

2 AI Successes of the Past Decade

In what follows, I will describe in broad strokes two large steps forward AI has taken within the last decade, focusing on (1) deep learning for computer vision and (2) deep reinforcement learning applied to single-player and competitive games. I will describe what it means for these artificial systems to succeed, note ways in which these resemble human learning, and highlight several of the differences between the ways these systems learn and the ways humans learn. To be very clear, this does not represent a representative survey of important AI advances of the past decade. However, this should help motivate more recent work in curiosity-driven, interactive artificial learning.

If you were shown a picture of a set dinner table, you could likely name just about every object (cups, bowls, plates, napkins, ...) and describe relations between various objects (“the plates are on top of the table,” “the chair is pulled under the table”). Likewise, if you were shown a video of a group of your friends, you could identify each of them immediately. You can make judgments about their internal states (“Kate is happy”), name a wide range of activities they are performing (“Rachel is walking,” “Pedro is waving”), and even infer goals and intentions (“Ruth is trying to get the others to walk over there.”). *Computer vision* is the domain of engineering artificial systems that can make these sorts of high-level judgements from image and video data. The capabilities of computer vision systems have steadily increased

over the past several decades, and the so-called deep learning revolution has brought about dramatic improvements since the early 2010s.

How do we judge improvement? After all, these visual capacities could be interpreted and tested in different ways – assessing intelligence is inherently subjective. The field of AI grapples with this constantly, and much effort is spent on finding good *benchmarks* with which we can measure success. Benchmarks typically consist of a dataset or a virtual environment upon which an artificial system is supposed to perform a task, as well as a set of performance metrics with which to judge success at this task. In a typical cycle of AI research, a group of researchers propose a new benchmark (usually meant to reflect some challenging cognitive ability that humans possess), they and others show whether or not existing methods perform well on this benchmark (useful benchmarks are those which existing or obvious approaches fail on), and the community engineers new systems aimed at high performance (while often modifying the data/environment, instructions for use, and performance metrics along the way).

To give a concrete example benchmark for computer vision, we will examine ImageNet (Deng et al. 2009; Russakovsky et al. 2015), perhaps the prototypical success story coming out of the deep learning revolution. ImageNet contains millions of images of objects, each one labeled as one of a thousand categories (“centipede,” “street sign,” “balloon”). The task here is to build an artificial system which takes, as input, an image, and outputs the correct object category name. The data are divided into a training set (containing many images of each of the 1000 categories), with which an artificial system is meant to “learn” the pattern between the images and labels, and a separate test set (consisting of new images) upon which the trained artificial system is evaluated. It turns out that one can create an artificial system that solves this task with high accuracy (Krizhevsky et al. 2012) – in some cases, perhaps superhuman accuracy (He et al. 2016; Russakovsky et al. 2015).¹

In what ways does this resemble human learning?² First, we do describe the model as *learning* from training data, and being *tested* on test data. The model consists of a large number (usually in such applications, many millions) of parameters, and these parameters are used to define a mathematical function that takes, as input, an image, and outputs a probability for each object name. For each example image and category label, we can associate a *loss* that is a measure of how bad the model’s output currently is. If the model thinks the correct object name is unlikely, the loss is high, whereas if it is likely, the loss is low. At the beginning of training, these parameters are assigned values randomly (there is some art to choosing good initializations and bad), and the model’s parameters are *optimized*

¹ The extent to which this is “superhuman” is worth a caveat. Russakovsky et al. (2015) benchmark humans and point out the challenge of doing so. To perform well at ImageNet, a human must become familiar with the 1000 categories – there is a difference between intuitively having a good sense of what is in an image, and being able to select the right category. It takes considerable time to learn how to do this well, and only a limited sample of human “experts” was used for comparison.

² To keep this discussion simple, I am describing early deep learning for computer vision results (e.g. Krizhevsky et al. (2012)). More recent results certainly add many caveats to these statements.

to produce low loss on the training data. At the end of training, the trained model should be able to make reasonable predictions on the training data.³ Performance is then measured on test data by holding fixed the parameters learned during training and seeing if the trained model generalizes to the new data – intuitively, this prevents the model from simply “memorizing” the train data.

This artificial learning process has a number of properties that bear some analogy to human learning. For example, we only really expect a trained model to perform well on data that looks sufficiently like training data. If a model has only been trained on photographs taken during the day, we do not expect it to work well at night,⁴ and if the training data has few examples of a particular object, the model will likely struggle to predict new instances of that object. The model can get confused by correlates: for example, if all German shepherds are depicted in the grass, and all Dobermans are in the snow, then a German shepherd in the snow could easily get mislabeled. And a model can “overfit” to training data: it is possible to produce models that perform very well during training but very poorly on new examples.⁵

Dramatically, this analogy extends to the neural level. It turns out that trained systems yield the best-known predictive models of the neural activity in the human ventral visual system (Yamins et al. 2014). This represents a dramatic full-circle success story of the interplay between the study of human cognition and AI: these models were inspired by our ventral visual stream, and a model trained to perform well at ImageNet, a challenging task we are good at, yields a useful computational model of our biology. There is a sense in which training a model on ImageNet yields a general visual *representation*. These artificial systems consist of a sequence of layers of “neurons” that feed into each other. The later layers provide a representation useful for predicting object names, and, it turns out, also useful for performing many other visual tasks. We say that these visual representations are general in that they support *transfer learning* – they can be used to learn a new, related task with a limited amount of data.

In what ways is this *not* like human learning? While we can point to countless discrepancies, let me point out two motivating differences. First, this success is one of *strong supervision*. The ImageNet task provides a prime example of supervised learning: our model is attempting to learn to associate an output (the object name) to each input (the image). In particular, there is a sense in which this supervision

³ For those of you who are familiar with training statistical models, this simply uses standard statistical modeling techniques, but deep learning models tend to involve far more parameters than a linear regression.

⁴ One might protest that this is decidedly *not* like how humans learn: when we are shown an object in sunlight, we can usually recognize it in the dark! But the question of what counts as a fair comparison arises – this might be more akin to the extreme deprivation of never seeing night. Arguments for the unique capacity for humans to generalize should consider the sorts of experience upon which we are training machines.

⁵ Indeed, much of the art of choosing good model architectures – the particular ways parameters are used – amounts to finding ones that not only fit well to training data but also generalize well to test data.

is particularly *strong*: in order to provide the model with these data, humans must carefully curate a labeled dataset. Contrast this with human learning: while humans do sometimes learn this in a similarly supervised way, a great deal of human learning has little explicit, cleanly labeled supervision from others (e.g., a child learning how to manipulate toys), and when there are labels, the label-learning process is much less carefully curated (e.g., first language learning). Second, this artificial learning is *passive*, not interactive. The dataset it uses to learn is determined ahead of time. The system need not make decisions about what to do in order to learn. Rather, it counts on humans to curate a dataset that it can fit to.

Our second exemplar success, that of reinforcement learning applied to games, contrasts these limitations somewhat, but with its own critical issues. Our example benchmark here is Atari (the “Arcade Learning Environment” (Bellemare et al. 2013)), which consists of a suite of games from the Atari video game console. The objective of Atari is to maximize score.

The framework in which we think about this task⁶ consists of a back-and-forth process between an *environment* and an *agent* which can act within it (Sutton and Barto 2018). At each timestep, the environment provides an *observation* (e.g., the current game image, or some more explicit state such as where all of the relevant objects are) and *reward* (e.g., additional score) to the agent, which can then choose from one of a set of *actions* (e.g., up, down, left, right). Execution of this action leads to the next observation. The goal of a reinforcement learning algorithm is to come up with an action-choosing decision mechanism (called the *policy*) for the agent that maximizes reward. There are many *deep reinforcement learning* methods for this – these treat the agent’s experience (observations, actions, and rewards) as training data, upon which a model is optimized.

In what ways do these artificial reinforcement learners, trained on Atari, reflect human learning?⁷ It turns out that one can train artificial reinforcement learners with comparable performance to humans (Mnih et al. 2015) – though, in some ways better, and in some ways worse. Further, unlike in our computer vision example, learning happens through an interactive process. In reinforcement learning, the agent gathers experience by interacting with its environment. Hence, in order to maximize reward, the agent must explore sufficiently so as to get a sense of how its actions affect the environment and what leads to reward, so that it can then seek that reward. As a result, interesting behaviors arise in the agent’s learning process. At first, agent behavior tends to appear random, and as it discovers sources of reward, its behavior looks more regular and deliberate. These “learning trajectories” are often quite interpretable and seem almost human in their successive improvements.

⁶I should emphasize: I am simplifying the formalism here – see Markov Decision Processes, or Partially Observed Markov Decision Processes (Sutton and Barto 2018).

⁷Again, to keep this discussion simple, this really applies to early deep reinforcement learning results applied in this domain (e.g., Mnih et al. (2015)). Many nuances apply as we approach more recent work.

In what ways are these Atari-playing reinforcement learners different from human learners? Again, the differences are manifold, but here are some motivating differences. First, as noted, performance obtained is in some ways superhuman, but in some ways lags. Artificial systems trained specifically on this task can learn to react very quickly and efficiently, leading to scores on some games that will simply dwarf that of humans. On the other hand, they lag behind in several games, for instance, on games with infrequent rewards.

These systems, while not falling explicitly in the category of supervised learning, are in a sense very strongly supervised. In Atari and related environments, for all except the most challenging tasks, the agent gets regular, explicit feedback in the form of reward. For example, the agent gets positive feedback for collecting a coin, or breaking a block, which leads it on its way towards an end goal (e.g., finishing a level). Contrast this with many human behaviors: our explicit, external rewards are often much less frequent – even in everyday tasks such as preparing food, we must set several reward-free steps (assembling the ingredients, putting them together) before we obtain something clearly rewarding. As we will describe in more detail shortly, standard reinforcement learning techniques can fail dramatically when reward functions are not engineered just so.

Further, these high-performance systems require a great deal of experience within the training environment—sometimes, the equivalent of a human playing for many years—in order to obtain performance comparable to a casual human player. That is not to say that we simply should expect artificial systems, trained only on these games, to achieve performance comparable to humans in the amount of time a human takes to learn these games. A reinforcement learning algorithm, started *de novo*, is far from a human trying a game for the first time. Humans can recruit from their representations gained in experiences throughout their lives. As a result, they likely have strong guesses about how their actions affect the environment and what leads to reward — for example, this is how a body affects its surroundings, and the gold coins probably mean reward.

This points to an important difference in what we are asking artificial systems to do, in training solely on Atari games. As experience is narrowly within the context of the game, the artificial learner is not asked to learn general-purpose representations about the world and then recruit those in order to quickly become proficient at the game. While models need to “know” something about the physical dynamics of game environments (e.g., if I move forward now, I fall off this cliff), this is very specific to the task. We, on the other hand, display a remarkable ability to recruit flexible, general representations in order to do well in new environments. For instance, if you were to enter a new, fully stocked apartment for the first time, you could, with perhaps a few minutes of looking around, make a cup of coffee. A flexible coffee maker is, sadly, beyond the capacities of AI to this day.

3 Artificial Curiosity and Interactive Learning

Thus far, I have presented two exemplar AI successes from the early 2010s: deep learning for computer vision, and reinforcement learning as applied to Atari games. I emphasized both ways in which we can think of these as reflecting aspects of human learning as well as ways in which this artificial learning falls short. At a high level, two critical limitations are

- Reliance on supervision – for computer vision, getting image category labels, and for reinforcement learning, a dependence on carefully crafted reward functions
- The extent to which learning is not interactive – for computer vision, the AI system learns passively on a curated dataset, and for reinforcement learning, the agent interacts but does not learn general representations useful for many settings

The above limitations argue for the development of artificial systems that (1) learn flexible representations that can be recruited for a wide range of tasks, (2) do so through interaction with their environments and others within them, and (3) do these things not through explicit, strong supervision signals, but rather learn in a more self-supervised manner, using more generic, flexible motivations. Here, I describe these desired traits in further detail, and, following that, I will outline several example successes along these lines.

Robust, flexible representation learning. The AI system learns general-purpose representations that are useful for accomplishing a wide range of tasks. To make this more concrete, let us look at several examples.

1. Sensory representations: From the raw inputs of our visual system, humans must be able to perceive and identify objects, understand how objects might be used, and so on. Human visual systems process raw visual information in order to be able to make these sorts of judgements.
2. Physical representations: As described in the introduction, humans have models of intuitive physics (Spelke 1985) that allow them to anticipate the world around them and perhaps facilitates plans. Some general-purpose ability to understand how the physical world evolves, in particular in response to actions, seems critical.
3. Representations of others: Humans possess intuitive psychology (Woodward 2009; Colle et al. 2007), allowing them to assess the goals, affect, and internal states of others. We possess theory of mind and a flexible ability to mentalize about others.

Learning through interaction. The AI system acts upon its environment, and it observes the result of this. How the AI system behaves shapes what it learns.

Learning through self-supervision, with generic, flexible motivations. The AI system should not have access to explicit supervision signals (e.g., category labels, except when these are provided through environment interaction, or handcrafted reward signals). Instead, its learning depends *only* on what it observes through

interaction. Further, its behavior should be shaped by generic, intrinsic motivations. These may include instrumental need satisfaction (for a human, food, water, warmth, etc.) and more generic motivations: information- or novelty-seeking (“curiosity”), control of environment, and social belonging, to name a few. As in the theory of Dweck (2017), such agents should be able to satisfy certain fundamental needs, and in order to do that, must be able to support the execution of and decision on a wide range of intermediate goals.

4 Examples of Artificial Curiosity and Interactive Learning

Now that I have described desired properties for more human-like artificial learning, I will dive into several works in this direction. I will begin with curiosity used as an exploratory aid, before moving to representation building for planning, and then ending with developmentally inspired curiosity-driven learning. This field has been incredibly active over the past several years, supported by decades of critical foundation work, and what is concretely described here represents only a sliver of these efforts. Further, it should be emphasized that the efforts described are several cases that build off of the successes described in the previous section and do not represent the first attempts to bring curiosity to AI. For a relevant survey of earlier attempts, please see Schmidhuber (2010), and Oudeyer et al. (2007) for a particularly relevant developmentally inspired work.

Imagine an exceedingly simple experiment: you are in a room with a button, and that button opens a door to another room, in which, at the other end, sits a cookie. After you eat the cookie, the environment resets, and you have the opportunity to start again and find the cookie. If you were in this environment but were not explicitly told about the cookie, you would probably find it quite quickly: you wonder what the button does, you find that it opens the door, you look around the other room, and, upon seeing the cookie, you recognize it as something you would like to eat. After the reset, if you would like to eat another cookie, you can go right to it, easily.

Yet imagine being in this environment with limited background knowledge (no knowledge of what buttons do, or how cookies taste) and with no sense of curiosity about the unknown. As a result, your exploratory behavior is completely unmotivated, and unless you somehow manage to put the cookie in your mouth, you do not realize that it is a good thing to do. Lacking any particular motivation, your behavior might look essentially random. Unless you somehow manage to, by chance, push the button, walk through the door, go to the other end of the room, and put the cookie in your mouth, you get absolutely no positive reinforcement for this chain of behaviors. As a result, it takes you an extraordinarily long time to begin to eat cookies.

This is an illustration of the *sparse reward problem*, an issue that plagues the standard reinforcement learning techniques used to solve Atari. Such systems need

to be given much more handcrafted rewards (e.g., pushing buttons, going through doors, moving towards cookies) in order to efficiently reach high performance.

But what if we give the AI *intrinsic motivation*, or *curiosity*? The agent should be rewarded not only by the cookie, but also by finding situations that are somehow interesting. This leads the agent to try new things: to press the button, to go through the door, to explore the room behind the door. How exactly this should be done is a matter of active research, with many proposed techniques. Several techniques (Pathak et al. 2017; Burda et al. 2018a; Pathak et al. 2019) involve a *world model* (Schmidhuber 2010; Ha and Schmidhuber 2018), a predictive model of the environment (think of this as a potential instantiation of a representation – an example would be a forward model, which predicts what happens if the agent chooses a particular action). The world model is *self-supervised*: it learns from experience. The agent’s intrinsic motivation, then, relates to how the world model responds to new experience. For instance, it might be rewarded by experience it finds difficult to model,⁸ or experience that leads it to make learning progress.⁹

Aside from encouraging useful exploratory behaviors, world models are, in theory, useful for planning. If the agent knows what states of the environment provide reward, and it knows, given the current state of the environment and an action it chooses, how the state changes, it can “imagine” the results of successive action choices and choose ones that lead to reward. This is the essential idea behind *model-based reinforcement learning*: the agent somehow uses a world model to plan.

For years, researchers struggled to make this intuition into a high-performance technique. The first techniques successful on Atari, for instance, are *model-free*. In deep Q-learning (Mnih et al. 2015), for instance, the agent learns a function Q that takes as input the environment’s current state s and a proposed action a . $Q(s, a)$ is then meant to estimate the total of all future rewards¹⁰ if the agent chooses action a in state s and then follows a policy for the rest of its actions. Q , if learned properly, tells the agent how to act – pick a so that Q is biggest! Note that this does not *explicitly* require a world model, but rather, its predictions are entirely in terms of rewards. Intuitively, this seems limited – if the agent is given a different task with a different reward, it is unclear how to transfer that knowledge. Further, the method at

⁸This equates the interesting with the difficult. This is potentially problematic! If the agent encounters something it cannot model, it is then drawn to get stuck on this. This is sometimes called the *white noise problem* (Schmidhuber 2010; Pathak et al. 2019). Considerable attention has been paid to resolving this (Pathak et al. 2017; Burda et al. 2018a, b; Kim et al. 2020).

⁹Not all techniques involve world models – e.g., some involve exploration through arbitrary goal-setting (Florensa et al. 2018; Nair et al. 2018; Campero et al. 2020). Though, perhaps this dichotomy is fairly artificial. If one has a fairly inclusive definition of what “world model” means (e.g., to include a wide range of representation learning techniques), many of these techniques can be lumped under this banner.

¹⁰Really, a discounted sum that weights rewards farther into the future less, which, as long as the reward stays bounded, keeps this from being infinite.

least *seems* inefficient: much information seems to be thrown out when predictions are all in terms of rewards.

Model-based reinforcement learning seems an obvious alternative. World models capture information about the environment, independently of reward, and if the agent’s task changes, this can be repurposed in a straightforward way. Yet model-based approaches lagged behind. One intuitive difficulty is simple: if the agent’s predictive model is wrong, planning can go horribly wrong. Only recently, model-based approaches have been shown to be competitive with, and in some ways superior to, model-free approaches (Hafner et al. 2019; Schrittwieser et al. 2020).

With this has come intriguing new advances that have brought us closer to the framework of the previous section. For example, in Sekar et al. (2020), an agent learns a world model independently of any objective – it is simply intrinsically motivated to improve its world model. It then can use this world model to accomplish a variety of tasks. This is tested in the DeepMind Control Suite environment (Tassa et al. 2018), in which an agent learns how to control its body, and is tested on its ability to walk forward, backward, and perform other physical feats. They demonstrate the agent’s ability to explore, build a world model, and then quickly perform these physical tasks when asked to do so. It is, at least in a sense, able to learn a general representation that it can recruit for performing a variety of specific tasks.

This sort of curiosity-based learning, then, moves us a step closer to the sort of learning we see in human development, so it is natural to ask: what does artificial curiosity achieve when placed in developmentally inspired environments? In the remainder of this section, I will describe two efforts in this direction: the first, in the domain of learning sensory and physical representations, the second, in the domain of representations of others.

In our first work (Haber et al. 2018), we designed a simple “playroom”: a 3D virtual environment in which an agent can move about a room and interact with a set of blocks (“toys” – see Fig. 1). For simplicity, the agent lacks a complex embodiment and instead can simply choose to move forward, backward, or turn. It has a limited field of view, and if it has a toy in view and that toy is sufficiently close, it can apply force and torque to the object.

The environment provides no extrinsic reward. We sought to understand if intrinsic motivation enables the agent to develop “play” behaviors, and if, in doing so, it develops useful sensory and physical representations. To build these representations, the agent trains a simple *inverse dynamics* world model: from a sequence of raw images, could it tell what action was taken? We could then test the capacity of these representations by evaluating their usefulness in performing related visual tasks.¹¹

Without any intrinsic motivation, the agent interacts in an essentially random way, and the agent interacts with toys in less than 1% of its experience. As a result,

¹¹ We used *transfer learning*: with these visual representations as inputs, we trained simple (linear) models for the positions and names of the objects, as a sort of “visual acuity test.”

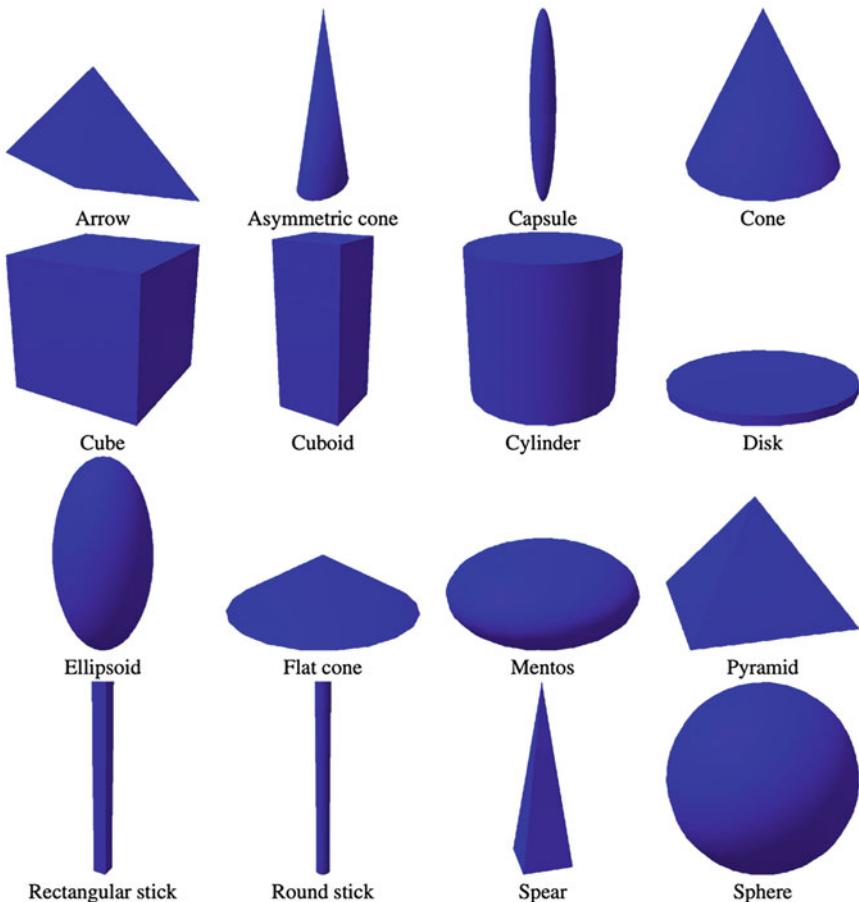


Fig. 1 The “toys” used in the 3D virtual environment in (Haber et al. 2018)

while its world model becomes good at understanding the motion of its body, it takes a very long time to understand object interaction, and its visual representations are not useful for tasks related to these objects.

Yet if the agent is rewarded by finding examples that are difficult for its world model, complex behaviors arise. In a room with one toy, we found that the agent moves about its environment somewhat randomly before suddenly taking an interest in objects: it consistently approaches and interacts with its toys. Correspondingly, its world model first becomes proficient at understanding actions that involve only its body, and then, after gaining more experience with toys, its toy-dynamics understanding increases. Interestingly, if the environment contains two toys, the agent starts in much the same way, but after a period of time, it engages in a qualitatively different behavior: it gathers the toys together and interacts with them

simultaneously. We found that the more sophisticated the agent’s behavior, the higher performance its physical and sensory representations. We next sought to extend this work on curiosity-driven learning from the physical to the social (Kim et al. 2020). We designed a simple environment meant to reflect aspects of social experience very early in life. Here the “baby” agent is surrounded by a variety of stimuli and can only “interact” with its environment by deciding what to look at. The stimuli we are surrounded with early in life—and throughout life—are wildly diverse. Some stimuli are *static*, like blocks: they only really do much when we physically interact with them. Others are dynamic, but really very *regular*: ceiling fans, mobiles, car wheels (and really, quite a lot of audio stimuli). On the opposite extreme, some stimuli are random, or *noisy*: they exhibit dynamics that are immensely challenging, if not impossible, to fully predict. The fluttering of leaves, the babble of a far-off crowd, the shimmering of light reflected off of water—in fact, while we pay little attention to most of these, most of the time, the noisy, random, and confusing surround us! Yet amidst this confusion are a particularly interesting class: *animate* stimuli. They exhibit incredibly complex behaviors that are very much unlike static or dynamic but regular inanimate objects (they exhibit self-starting motion, for instance, and are impossible to fully predict) yet are in some ways very regular — they act according to goals, affect, beliefs, and personality.

How do we design agents that can decide what to look at, in order to learn about these sorts of surroundings? What if we want this agent to learn as much as possible, as quickly as possible? Further, how do people make these sorts of visual attention decisions, when presented with novel stimuli? In answering these questions, we were faced with a problem: all of these different types of stimuli tend to look quite different. This would complicate the design of machines that can learn from all of them, and confound any human subject experiment. We hence designed environments that took these classes of stimuli—static, regular, noise, and animate—and stripped them down to basic informational essentials (Fig. 2). We designed spherical avatars that executed these sorts of behaviors with simple motion patterns. For instance, the regular stimuli simply rolled around in circles, or back and forth in a straight line. The noise stimuli performed a sort of random walk: randomly lurching in one direction, followed by another. We designed a wide range of stimuli meant to be animate. One chases another. One navigates towards a succession of objects. Another plays a sort of “peekaboo” with the viewer: if the viewing agent looks at the stimulus, it darts behind an object, and when the viewing agent looks away, it peeks out again.

We then experimented with different intrinsic motivation rewards and found that different ones led to drastically different behaviors. For instance, if an agent is motivated to find difficult examples for its world model, as it did in the previous study, it becomes fixated on the noise stimuli, as it is never able to precisely learn this phenomenon (an example of the *white noise problem*). If an agent is motivated to find easy examples, it spends most of its time on the simplest stimuli. Yet if an agent is motivated to make progress in modeling its world, it finds a balance. As noise stimuli are impossible to fully predict, it ceases making progress on them and it gets “bored” of them. This allows it to spend more time on the challenging but learnable

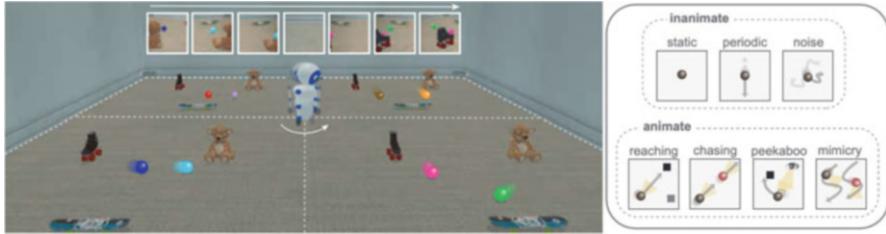


Fig. 2 “Social” virtual environment. The 3D virtual environment from Kim et al. (2020). The *curious agent* (white robot) is centered in a room, surrounded by various colored spheres contained in different quadrants, each with dynamics that correspond to a realistic inanimate or animate behavior (right box). The curious agent can rotate to attend to different behaviors as shown by the first-person view images at the top. See <https://bit.ly/31vg7v1> for videos

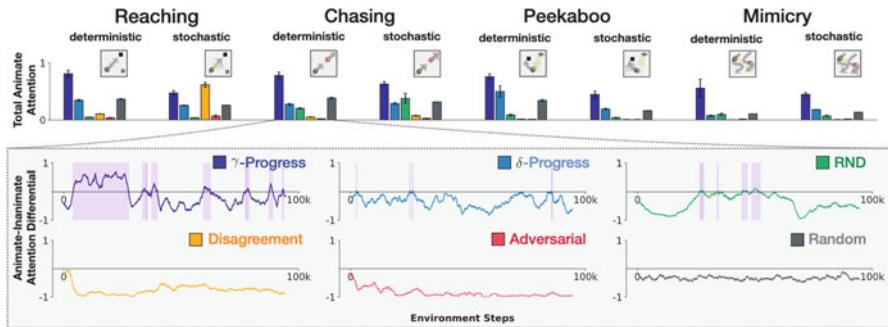


Fig. 3 Emergence of animate attention. The bar plot shows the total animate attention, which is the ratio between the number of time steps an animate stimulus was visible to the curious agent, and the time steps a noise stimulus was visible. The time series plots in the zoom-in box show the differences between mean attention to the animate external agents and the mean of attention to the other agents in a 500-step window, with periods of animate preference highlighted in purple. Results are averaged across five runs. γ -Progress and δ -Progress are progress-based intrinsic rewards, Adversarial equates reward with loss, Random chooses actions randomly, RND is a novelty-based intrinsic reward, and Disagreement rewards based on variance of predictions between several independently initialized world models

regularities seen in animate stimuli. How progress should be estimated, precisely, is an intensely challenging problem – this is strongly related to computing expected information gain and is a key computational challenge found in active learning and optimal experiment design literature (Cox and Reid 2000; Settles 2009). We tried several methods and found one to exhibit a characteristic “animate attention” bump (See Fig. 3).

We were able to track not only a variety of different learning behaviors, but also a variety of learning outcomes. The progress-based method that exhibited animate attention was able to learn the learnable (static, regular, animate) behaviors the

best, and it was able to learn them fastest.¹² Agents that fixated on noise stimuli, or divided their attention more evenly between all stimuli, lagged behind in learning animate behaviors.

In ongoing work, we seek to understand what sort of intrinsic motivation best corresponds to human behavior. To answer this, we designed a physical version of the above environment. For the stimuli, we used Spheros (Sphero 2021), simple spherical robots with a gyroscopic motor that can be programmed or controlled remotely.¹³ We recruited adults and tracked their gaze while they were asked to simply view these robotic scenes. Ongoing analyses will allow us to compare human attention behavior to artificial attention behavior, giving us a sense of what sorts of motivations humans have in these simple curiosity-driven learning environments.

5 Artificial Interactive Learning as Models of Human Learning

Thus far, we have examined gaps between learning in human development and learning in artificial systems, and we have discussed recent advances in artificial intelligence that are filling aspects of this gap. To be sure, the gap remains incredibly wide, but continuing advances in our understanding of human learning should help us close this gap. Not only can a fine-grained understanding of learning processes tell us how to engineer new artificial systems, but it also can tell us the right sorts of benchmarks and “specs” we should be engineering for. One of the most important questions learning science can teach artificial intelligence is simply: precisely what sorts of learning capacities should we try to engineer? ImageNet came out of this sort of thinking. It represents a difficult task that we know is important and doable for humans, and this combination helped bring about great success in the last decade.

Yet let us turn to an important speculative question: how might this AI enterprise help us better understand human learning? Of course, the enterprise of trying to build artificial systems that learn more like we do is broadly thought to be useful for better understanding how we learn. At the coarsest level, attempting to build these sorts of systems directs our energies towards understanding critical aspects of how humans learn. Engineering refines the questions we ask of cognition. In short, we expect a virtuous cycle of advances from the fields of cognitive and learning

¹² Of course, we are making subjective choices when deciding how to “test” these agents. We presented them with various situations in which they view the various stimuli and had them predict the future evolution of these stimuli – e.g., we had them “play peekaboo” with the peekaboo agent, and then examined world model predictions. This might be thought of as a sort of dynamical and social acuity test.

¹³ These are marketed not as research robots but as educational toys.

sciences and artificial intelligence, and the very act of doing this sort of research should yield us better models of human learning.

But what would it mean for AI to actually *model* human? We engineer an *agent architecture*, which, by placing it in an *environment*, yields us *behaviors*, *representations*, and *learning*. Different agent architectures, exposed to different environments, yield different behaviors, abilities, and representations. This association between agent architecture, environment, behaviors, learning outcomes, and representations becomes useful if it can yield predictive models of corresponding features of human learning. For instance, we might like to understand, in the human realm, what environments and/or behaviors tend to lead to what learning. Or, perhaps more impactfully, we would like to know, given an individual's past environment (perhaps coupled with knowledge of past behavior and/or learning outcomes), what sort of environment should lead to desired learning outcomes.

But how might we get from artificial learning to human learning, in this way? Success here seems to hinge on a sort of task-driven modeling hypothesis (Yamins et al. 2014). That is, we must be able to identify human capacities and behaviors such that (1) we are able to come up with architectures that sufficiently accurately reflect these identified human behaviors and capacities, and (2) these capacities and behaviors represent sufficiently strong constraints that a limited collection of architectures satisfies them. This allows us to “triangulate” agent architectures that produce predictive models of human behavior. In essence, we hope that we can reduce this modeling problem to an engineering problem: create an artificial system that has the right sorts of capacities and behaviors, and since not many systems satisfy all of these properties, the result is human-like learning.

Early developmental learning, we hope, will be tractable for this sort of approach. Early developmental learning is critically important for the entire life course, and hence it is reasonable to hypothesize that humans are in a sense optimized to do this very well (though, surely, there is not just *one* concrete objective, or one “optimal” way of doing this). Hence, it is thought that an “ImageNet of developmental learning” can be found – some benchmark that allows us to refine artificial systems that then are able to model the developmental process in a fine-grained way. To do this, it seems likely that we will need extensive fine-grained data collection of developmental learning environments, behaviors, and learning outcomes.

One particularly exciting aspect of this modeling effort lies in its possibility to model not just the typical learner (which, surely, does not truly exist!), but rather, the full diversity of human learners. As a case study of this sort of thinking, consider the Autism Spectrum Disorder (ASD). ASD has historically been characterized by differences in high-level social behaviors and skills (Hus and Lord 2014). Yet over the past two decades, an intriguing new picture has emerged. ASD children exhibit differences in play behavior as well as sensory sensitivities (Robertson and Baron-Cohen 2017). Further, ASD children exhibit differences in social attention – this has been across 2–6 months of age (Jones and Klin 2013; Shic et al. 2014; Moriuchi et al. 2017). In short, evidence strongly supports the claim that the sort of early interactive learning we are attempting to engineer in artificial systems is somehow *different* in ASD children relative to the general population. Understanding the

phenomenon of this difference on a computational level may help us reconceptualize learning differences like these, as well as replace coarse diagnostic criteria with a much finer-grained picture and more empowering learning tools.

References

- Begus, K., Gliga, T., Southgate, V. (2014). Infants learn what they want to learn: Responding to infant pointing leads to superior learning. *PLoS one*, 9(10), e108817.
- Bellemare, M. G., Naddaf, Y., Veness, J., Bowling, M. (2013). The arcade learning environment: An evaluation platform for general agents. *Journal of Artificial Intelligence Research*, 47, 253-279.
- Burda, Y., Edwards, H., Storkey, A., Klimov, O. (2018a). Exploration by random network distillation. *arXiv preprint arXiv:1810.12894*.
- Burda, Y., Edwards, H., Pathak, D., Storkey, A., Darrell, T., Efros, A. A. (2018b). Large-scale study of curiosity-driven learning. *arXiv preprint arXiv:1808.04355*.
- Campero, A., Raileanu, R., Küttler, H., Tenenbaum, J. B., Rocktäschel, T., Grefenstette, E. (2020). Learning with amigo: Adversarially motivated intrinsic goals. *arXiv preprint arXiv:2006.12122*.
- Colle, L., Baron-Cohen, S., Hill, J. (2007). Do children with autism have a theory of mind? A non-verbal test of autism vs. specific language impairment. *Journal of autism and developmental disorders*, 37(4), 716-723.
- Cox, D. R., & Reid, N. (2000). *The theory of the design of experiments*. CRC Press.
- Deng, J., Dong, W., Socher, R., Li, L. J., Li, K., Fei-Fei, L. (2009, June). Imagenet: A large-scale hierarchical image database. In *2009 IEEE conference on computer vision and pattern recognition* (pp. 248-255). IEEE.
- Dweck, C. S. (2017). From needs to goals and representations: Foundations for a unified theory of motivation, personality, and development. *Psychological review*, 124(6), 689.
- Fantz, R. L. (1964). Visual experience in infants: Decreased attention to familiar patterns relative to novel ones. *Science*, 146(3644), 668-670.
- Florensa, C., Held, D., Geng, X., Abbeel, P. (2018, July). Automatic goal generation for reinforcement learning agents. In *International conference on machine learning* (pp. 1515-1528). PMLR.
- Gopnik, A., Meltzoff, A. N., Kuhl, P. K. (1999). *The scientist in the crib: Minds, brains, and how children learn*. William Morrow & Co.
- Goupil, L., Romand-Monnier, M., & Kouider, S. (2016). Infants ask for help when they know they don't know. *Proceedings of the National Academy of Sciences*, 113(13), 3492-3496.
- Ha, D., & Schmidhuber, J. (2018). Recurrent world models facilitate policy evolution. *arXiv preprint arXiv:1809.01999*.
- Haber, N., Mrowca, D., Wang, S., Fei-Fei, L., Yamins, D. L. (2018, December). Learning to play with intrinsically-motivated, self-aware agents. In *Proceedings of the 32nd International Conference on Neural Information Processing Systems* (pp. 8398-8409).
- Hafner, D., Lillicrap, T., Ba, J., Norouzi, M. (2019). Dream to control: Learning behaviors by latent imagination. *arXiv preprint arXiv:1912.01603*.
- He, K., Zhang, X., Ren, S., Sun, J. (2016). Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 770-778).
- Hus, V., & Lord, C. (2014). The autism diagnostic observation schedule, module 4: revised algorithm and standardized severity scores. *Journal of autism and developmental disorders*, 44(8), 1996-2012.
- Jones, W., & Klin, A. (2013). Attention to eyes is present but in decline in 2-6-month-old infants later diagnosed with autism. *Nature*, 504(7480), 427-431.

- Kidd, C., Piantadosi, S. T., Aslin, R. N. (2012). The Goldilocks effect: Human infants allocate attention to visual sequences that are neither too simple nor too complex. *PloS one*, 7(5), e36399.
- Kim, K., Sano, M., De Freitas, J., Haber, N., Yamins, D. (2020, November). Active world model learning with progress curiosity. In *International conference on machine learning* (pp. 5306-5315). PMLR.
- Krizhevsky, A., Sutskever, I., Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. *Advances in neural information processing systems*, 25, 1097-1105.
- Mnih, V., Kavukcuoglu, K., Silver, D., Rusu, A. A., Veness, J., Bellemare, M. G., et al. (2015). Human-level control through deep reinforcement learning. *Nature*, 518(7540), 529-533.
- Moriuchi, J. M., Klin, A., Jones, W. (2017). Mechanisms of diminished attention to eyes in autism. *American Journal of Psychiatry*, 174(1), 26-35.
- Nair, A., Pong, V., Dalal, M., Bahl, S., Lin, S., Levine, S. (2018). Visual reinforcement learning with imagined goals. *arXiv preprint arXiv:1807.04742*.
- Oudeyer, P. Y., Kaplan, F., Hafner, V. V. (2007). Intrinsic motivation systems for autonomous mental development. *IEEE transactions on evolutionary computation*, 11(2), 265-286.
- Pathak, D., Agrawal, P., Efros, A. A., Darrell, T. (2017, July). Curiosity-driven exploration by self-supervised prediction. In *International conference on machine learning* (pp. 2778-2787). PMLR.
- Pathak, D., Gandhi, D., Gupta, A. (2019, May). Self-supervised exploration via disagreement. In *International conference on machine learning* (pp. 5062-5071). PMLR.
- Robertson, C. E., & Baron-Cohen, S. (2017). Sensory perception in autism. *Nature Reviews Neuroscience*, 18(11), 671-684.
- Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., et al. (2015). Imagenet large scale visual recognition challenge. *International journal of computer vision*, 115(3), 211-252.
- Schmidhuber, J. (2010). Formal theory of creativity, fun, and intrinsic motivation (1990–2010). *IEEE Transactions on Autonomous Mental Development*, 2(3), 230-247.
- Schriftwieser, J., Antonoglou, I., Hubert, T., Simonyan, K., Sifre, L., Schmitt, S., et al. (2020). Mastering atari, go, chess and shogi by planning with a learned model. *Nature*, 588(7839), 604-609.
- Sekar, R., Rybkin, O., Daniilidis, K., Abbeel, P., Hafner, D., Pathak, D. (2020, November). Planning to explore via self-supervised world models. In *International Conference on Machine Learning* (pp. 8583-8592). PMLR.
- Settles, B. (2009). *Active learning literature survey*.
- Shic, F., Macari, S., Chawarska, K. (2014). Speech disturbs face scanning in 6-month-old infants who develop autism spectrum disorder. *Biological psychiatry*, 75(3), 231-237.
- Smith, L. B., & Slone, L. K. (2017). A developmental approach to machine learning?. *Frontiers in psychology*, 8, 2124.
- Spelke, E. S. (1985). Object permanence in five-month-old infants. In *Cognition*.
- Sphero. (2021). <https://sphero.com/>. Accessed: 2021-10-10.
- Sutton, R. S., & Barto, A. G. (2018). *Reinforcement learning: An introduction*. MIT press.
- Tassa, Y., Doron, Y., Muldal, A., Erez, T., Li, Y., Casas, D. D. L., et al. (2018). Deepmind control suite. *arXiv preprint arXiv:1801.00690*.
- Tomasello, M. (2016). The ontogeny of cultural learning. *Current Opinion in Psychology*, 8, 1-4.
- Tomasello, M., Kruger, A. C., Ratner, H. H. (1993). Cultural learning. *Behavioral and brain sciences*, 16(3), 495-511.
- Twomey, K. E., & Westermann, G. (2018). Curiosity-based learning in infants: a neurocomputational approach. *Developmental science*, 21(4), e12629.
- Woodward, A. L. (2009). Infants' grasp of others' intentions. *Current directions in psychological science*, 18(1), 53-57.
- Yamins, D. L., Hong, H., Cadieu, C. F., Solomon, E. A., Seibert, D., DiCarlo, J. J. (2014). Performance-optimized hierarchical models predict neural responses in higher visual cortex. *Proceedings of the national academy of sciences*, 111(23), 8619-8624.

Open Access This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.



Assessing and Tracking Students' Wellbeing Through an Automated Scoring System: School Day Wellbeing Model



Xin Tang, Katja Upadyaya, Hiroyuki Toyama, Mika Kasanen, and Katariina Salmela-Aro

Contents

1 Introduction	55
2 The Assessments of Wellbeing	56
3 Artificial Intelligence-Based Wellbeing Assessments and Enhancement	57
4 School Day Wellbeing Model: A Model for Wellbeing Assessment and Enhancement	59
4.1 Theoretical Foundations for the School Day Wellbeing Model	59
4.2 School Day Wellbeing Model	61
4.3 How Does the School Day Wellbeing Model Work?	62
5 Features of the School Day Wellbeing Model	63
6 User Experiences	66
7 The Future Directions	67
8 Conclusion	68
References	68

1 Introduction

Wellbeing, the state of being well in physical, mental, and social aspects of life, has been a focus in research for the past two decades (Diener et al. 2017, 2018; Seligman 2011). People who have high wellbeing are likely to succeed in life (Lyubomirsky et al. 2005), to live longer (Diener and Chan 2011), and to conduct prosocial behaviors (Oishi et al. 2007). Students with high wellbeing are also the ones who have high academic achievement (Kiuru et al. 2020; Salmela-Aro 2020) and exhibit fewer problem behaviors (Arslan and Renshaw 2018). Given the important role

X. Tang (✉) · K. Upadyaya · H. Toyama · K. Salmela-Aro

University of Helsinki, Helsinki, Finland

e-mail: xin.tang@helsinki.fi; katja.upadyaya@helsinki.fi; toyama.hiroyuki@helsinki.fi;
katariina.salmela-ar@helsinki.fi

M. Kasanen

School Day, Helsinki, Finland

e-mail: mika.kasanen@schoolday.com

of wellbeing, it is not surprising to see a surge of research on the assessment, antecedents, and outcomes of wellbeing.

The assessment of wellbeing has always been a significant theme in the society and research field. During the past few decades, national-level policy makers have tried to assess and track wellbeing to build a sustainable society (e.g., UK Office for National Statistics; Allin and Hand 2017). International comparison assessments (e.g., World Happiness Report; Helliwell et al. 2021) also collect wellbeing data to compare and understand wellbeing gaps between multiple countries across the world. Education sectors (e.g., education policy makers, schools, universities) are joining this endeavor to understand students' wellbeing, with an aim of improving wellbeing to support learning gains (OECD 2013, 2019). Industries are also striving to provide applications to assess, track, and report wellbeing. With the development of Artificial Intelligence (AI) techniques, there are increasing applications and research on AI-based wellbeing assessments (Castro et al. 2018).

In this chapter, we aim to introduce a newly developed wellbeing assessment and enhancement system, School Day Wellbeing Model, as a joint product of researchers and industry practitioners. We first review the (traditional) assessments of wellbeing, and then review AI-based wellbeing assessments. After identifying some caveats in those assessments, the School Day Wellbeing Model is introduced to show its features and strengths as a novel AI-based wellbeing assessment application. The user experiences are also gathered to show its validity and the future directions of the Model are discussed.

2 The Assessments of Wellbeing

Measuring wellbeing has been a central task for the new science of wellbeing (Diener et al. 2018). Wellbeing has been assessed and indexed using objective measures (e.g., physiological data, life expectancy as for country-level wellbeing) and subjective measures (e.g., self-reported happiness, life satisfaction; for reviews, see Conceição and Bandura 2008; Ong et al. 2021). Though objective measures can provide some information on wellbeing, the majority of wellbeing assessments are subjective measures as wellbeing is largely idiographic (i.e., relating to an individual's own experiences and interpretations; Rose et al. 2017; VanderWeele et al. 2020). To date, subjective wellbeing has been mainly examined from three approaches: evaluative, hedonic, and eudaimonic approaches. The evaluative approach portrays wellbeing as an individual's view of satisfaction with life. The corresponding scales typically measure the overall life satisfaction or satisfaction in different domains of life (Diener et al. 1985; Pavot and Diener 1993). Hedonic approach examines wellbeing as positive affective experiences such as happiness or pleasure. Scales under this approach typically ask the participants to report their experiences of positive and negative emotions (e.g., Positive and Negative Affect Schedule – PANAS Scale; Watson et al. 1988). The last approach, eudaimonic approach, describes wellbeing from the perspective of meaning and purpose of

life. Accordingly, scales from this approach typically measure the extent to which individuals live a purposeful life or fulfill their self-realization (Ryff 1989; Ryff and Singer 2008).

Recently, most wellbeing assessments acknowledge the multidimensional nature of wellbeing and typically include items from all three approaches (Diener et al. 2018; VanderWeele et al. 2020). For instance, Seligman's model of Positive emotion, Engagement, positive Relationship, Meaning, and Accomplishment (PERMA) describes wellbeing as a compound concept (2011). According to the PERMA model, *positive emotions* denote hedonic experiences, such as feeling happy, joyful, and cheerful. *Engagement* represents positive experiences in activities, such as feeling absorbed and immersed in life. *Positive relationships* refer to the psychological connections with others (e.g., peers and parents). *Meaning* represents the feelings of being valuable and of being purposeful in life. *Accomplishment* refers to feeling capable to pursue goals and to finish tasks. A valid wellbeing assessment for adolescents on the basis of the PERMA model has also been established recently (Kern et al. 2015).

However, for most wellbeing assessments, the common collecting method is paper and pencil, which reduces data collection efficiency. In addition, much wellbeing information is collected only once per year, limiting the assessments' ecological validity. One recent review (Ong et al. 2021) indicated that only 1.7% assessments ask for the reporting of wellbeing at the momentary level. As hedonic wellbeing (e.g., positive or negative emotions) is highly sensitive to situations, wellbeing assessments with high ecological features are imperatively needed. The School Day Wellbeing Model is a tool which measures subjective wellbeing in a timely manner, collects wellbeing data virtually, reports wellbeing automatically, and offers feedback correspondingly. We provide a detailed description of the School Day Wellbeing Model in the latter section.

3 Artificial Intelligence-Based Wellbeing Assessments and Enhancement

For several decades, educational assessments using artificial intelligence-based techniques and tools have been a research topic. To date, the most common AI-based assessments in the field of education are automated grading systems or adaptive assessment systems (Gardner et al. 2021; González-Calatayud et al. 2021). There has been also a great interest in collecting wellbeing information with the help of intelligent systems or devices in recent years. Nowadays, there is a wide use of intelligent devices (e.g., smartphones, smart watches, smart wristbands) that collect information on sleep patterns and physical exercise, which are essential parts of wellbeing (Castro et al. 2018). However, these measures are mostly for adults, rather than for school children or adolescent students. More importantly, the information collected by the intelligent devices mostly concern indicators of objective wellbeing

rather than subjective wellbeing. Yet, as we stated, subjective wellbeing is a critical and indispensable part of one's wellbeing. Some researchers even argue that we should focus mainly on subjective wellbeing as the interpretation process is critical for the final wellbeing status (Krueger and Stone 2014; OECD 2013). For instance, people may experience happiness even if their exercise is sparse or sleep only 6 h a night.

Researchers have attempted to examine associations between objective measures (e.g., heart rate variability, blood pressure, mobile log data) and subjective wellbeing (e.g., happiness, positive and negative emotions; Gordon and Mendes 2021; Jaques et al. 2015). Given the big data gathered through intelligent devices, researchers have utilized several machine learning algorithms to predict subjective wellbeing on the basis of data on objective measurements (Jaques et al. 2015; Taylor et al. 2020). The central idea is to see whether subjective wellbeing can be represented by merely looking at data on objective measures.

For instance, one study collected four types of data (physiological data, survey data, phone data, location data) with mobile sensors and smartphones (Jaques et al. 2015). University students participated in the study over two 1-month (30-day) experimental periods. Physiological data consisted of electrodermal activity (EDA; a measure of physiological stress), and three-axis accelerometer (a measure of steps and physical activity). The survey data consisted of questions about academic activity, sleep, drug and alcohol use, exercise, stress, and wellbeing measures such as health, energy, alertness, and happiness. The phone data included phone call, SMS, and usage patterns. The location data included the GPS coordinates throughout the day. The authors extracted and formulated features from each data source before they evaluated and reduced the number of features. After this step, multiple algorithms, such as Support Vector Machines (SVM), Random Forests (RF), Neural Networks (NN), Logistic Regression (LR), k-Nearest Neighbor (kNN), and AdaBoost, were applied to test the predictability of each algorithm for classifying subjective happiness. The results showed that an ensemble classifier they discovered can have about 70% accuracy rate in predicting the state of happiness. However, in this study and many other studies (e.g., Gordon and Mendes 2021; Taylor et al. 2020), the examinations of subjective wellbeing are very limited. In addition, the multidimensional nature of subjective wellbeing (including general and academic wellbeing) was unaddressed.

Besides the AI-based wellbeing assessments, there are also several intelligent applications that aim to improve wellbeing. The most typical application is conversational agents or chatbots (Dekker et al. 2020; Inkster et al. 2018). Chatbots or agents utilize natural language processing techniques with psychological counseling methods (e.g., dialectical behavior therapy, behavioral reinforcement, mindfulness) and can respond to users' questions and requests and to reduce their health problems (e.g., anxiety, stress, sleeping problems). For instance, one conversational AI agent (Wysa App) used text-analysis techniques to converse with users who needed assistance for their wellbeing (Inkster et al. 2018). The authors revealed that the frequent use of this application improves the users' wellbeing (by reducing their depressive symptoms) significantly.

In the School Day Wellbeing Model, we choose an approach which combines wellbeing assessment and improvement simultaneously. As we value the complex and multidimensional nature of subjective wellbeing, we constructed a new wellbeing assessment model. More importantly, the way the data has been collected is largely different from traditional assessments. Experience sampling methods (Hektner et al. 2007), in which survey items are randomly repeatedly measured, have been used. More importantly, the randomization of the item sampling is driven by AI techniques (see the following section) to select the questions strategically and automatically. After the data has been collected, the wellbeing status will be reported automatically and the feedback for improvement will be delivered timely according to the status. A detailed description of the model is in the following sections.

4 School Day Wellbeing Model: A Model for Wellbeing Assessment and Enhancement

The School Day Wellbeing Model is constructed jointly by the researchers and practitioners as a response to the call for an ecologically valid measure of wellbeing and for an intelligent solution to detect and improve student wellbeing. A distinctive part of the School Day Wellbeing model, in comparison with other wellbeing assessments, is that it not only focuses on measuring wellbeing but also on improving wellbeing. In other words, it is a model for wellbeing assessment and enhancement simultaneously. The model intends to report, monitor, and track wellbeing live, so that it can provide timely feedback given the person's current wellbeing status.

4.1 *Theoretical Foundations for the School Day Wellbeing Model*

The School Day Wellbeing Model is built by integrating three theoretical frameworks (see Fig. 1 for the latest model): School Wellbeing Model, Study Demands-Resources Model, and OECD Social Emotional Skills.

4.1.1 School Wellbeing Model

School wellbeing model (Konu et al. 2002; Konu and Rimpelä 2002) defined four broad indices to represent wellbeing and its supportive environment: school conditions, social relationships, means for self-fulfillment, and health status. School conditions include physical environment (e.g., ventilation is good; inappropriate desks), school organization (e.g., rules and regulations are sensible), and school



Fig. 1 The School Day Wellbeing Model

services. Social relationships cover school climate (e.g., teachers treat pupils fairly), relationships with teachers and peers (e.g., I have friends in school; easy to get along with teachers), and bullying experiences (e.g., classmates intervene in bullying). Means for self-fulfillment includes autonomy support (e.g., pupils' views are taken into account) and school engagement (e.g., I am able to follow teaching). Health status contains the evaluation of current physical health condition. The model has been recognized as a valid tool for assessing students wellbeing from grade 4 to 12 (Koru and Lintonen 2006).

4.1.2 Study Demands-Resources Model

Study Demands-Resources model (Salmela-Aro, Tang and Upadyaya, *in press*; Salmela-Aro and Upadyaya 2014) proposed that wellbeing (particularly school engagement and burnout) is based on the fit between demands and resources. Both demands and resources can be divided into school- and person-related factors. Demands are factors that cause exhaustion and burnout, such as school work load.

Resources are factors that promote personal development, such as self-efficacy and social support. More importantly, the model proposes a synergistic role of demands and resources in determining wellbeing. In consequence, the assessment of wellbeing should consider the positive and negative side of environmental factors and of wellbeing itself. The model has been tested among students and shown its predictive validity in explaining well- and ill-being (Romano et al. 2020; Salmela-Aro et al. 2008; Salmela-Aro and Upadyaya 2014).

4.1.3 OECD Social Emotional Skills Framework

To understand the key factors that enhance wellbeing, the OECD social-emotional skill framework (Kankaraš and Suarez-Alvarez 2019) was adopted and included in the model. It defines social-emotional skills as: “individual capacities that (a) are manifested in consistent patterns of thoughts, feelings, and behaviors, (b) can be developed through formal and informal learning experiences, and (c) influence important socioeconomic outcomes throughout individual’s life” (OECD 2015, p. 35). The model proposed five broad skills: task performance, emotional regulation, collaboration, open-mindedness, and engaging with others. *Task Performance* refers to the ability to be self-disciplined, persistent, and dedicate effort in achieving goals and completing tasks. *Emotional Regulation* is the ability to control one’s emotional responses and moods, as well as to be positive and optimistic about self and life in general. *Collaboration* is the ability to maintain positive relations and to be sympathetic to others. *Open-mindedness* is the ability to engage with new ideas and generate novel ways to do or think. Lastly, *Engaging with Others* is the ability to engage with others, and to be energetic and assertive. The role of social-emotional skills in affecting students’ wellbeing and achievement has been established in the OECD international comparison study of social-emotional skills (OECD 2021) and other recent studies (Guo et al., 2022; Salmela-Aro et al. 2021; Salmela-Aro and Upadyaya 2020; Tang et al. 2019, 2021).

4.2 School Day Wellbeing Model

As an integrative model, the School Day Wellbeing Model has four broad domains: Learning, Social and Emotional Skills, Social Relationships, and Wellness (see Fig. 1). *Learning* is the domain that covers studying skills and environment factors, such as self-studying (e.g., I like studying on my own), study support (e.g., It is easy to get support from teachers), learning environment (e.g., I have a peaceful place to study), and learning material (e.g., I have the necessary school supplies). *Social and Emotional Skills* are five skills introduced above (i.e., task performance, emotional regulation, collaboration, open-mindedness, and engaging with others). *Social Relationships* is the domain related to the communication and interaction. It includes communication with teachers (e.g., It is easy to keep in touch with my

teachers), communication with peers (e.g., I can get help from my classmates), communication outside school (e.g., I get support when studying at home), and student services (e.g., I can get help if I am overwhelmed). *Wellness* is the domain related to physical health, mental health, and academic wellbeing. It covers physical health (e.g., I am not concerned about my health), emotions (e.g., I feel happy; My anxiety is low), diet (e.g., My diet is healthy), psychological wellbeing (e.g., I like being at school), and academic wellbeing (e.g., Time flies when I am studying). Overall, the model has 64 items with each dimension having three to six items.

4.3 How Does the School Day Wellbeing Model Work?

The School Day Wellbeing Model is driven by several automated techniques (Kylväjä et al. 2019) in sampling the items, cleaning the data, scaling the answers, reporting the results, and providing feedback (see Fig. 2). Information concerning subjective wellbeing is collected through a mobile, web, or an online platform (e.g., Microsoft Teams). The platform notifies students to answer questions once a week. Once a classroom takes School Day into use for the first time, the model asks all the 64 questions so that an immediate baseline can be formed in the classroom. After the initial 64 questions, the amount of questions to be answered is limited to 10 items per week per student to reduce cognitive burden. The question sampling procedure is not purely random. The questions are delivered by an Artificial Intelligence algorithm built by School Day that selects the items strategically from the item pool so that a balanced sample of student wellbeing can be formed at any particular time. The answers to the items are recorded on a Likert scale (5 = totally agree, 1 = totally disagree) and scaled to the point from 1 to 100 with scaling functions. The wellbeing

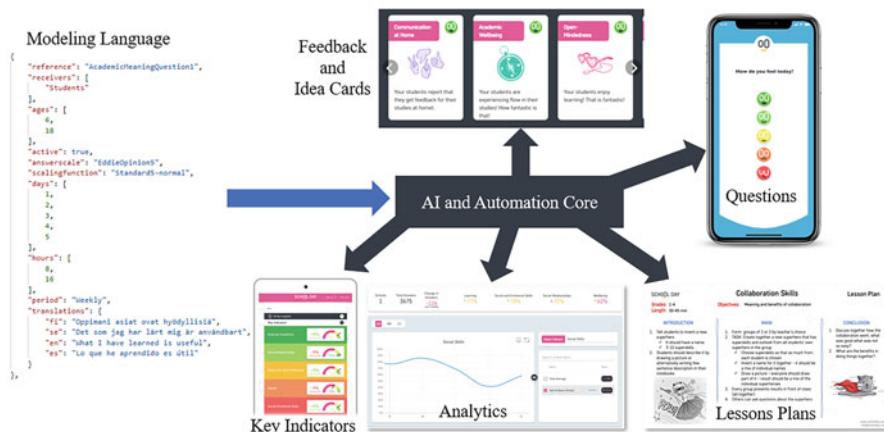


Fig. 2 Automated process of data collection, analyses, report, and feedbacks

reports are then generated automatically based on the answers. The reports can be read by teachers concerning their own class, by principals concerning their school, and by administrators concerning the region they are responsible for. The wellbeing reports also include trends of change so that the wellbeing status of each entity (classroom, school, region) can be compared weekly, monthly, or yearly. It is also possible to compare the wellbeing performance across classes and schools.

Once the wellbeing status has been recorded and reported, the feedback module will function to provide adaptive group level feedback according to the wellbeing status. The feedback is delivered to students, teachers, principals, and educational administrators. School Day AI module distributes weekly (e.g., cards) content highlighting what is going well and what needs attention and improvement. The weekly feedback content covers a broad series of wellbeing improvement practices (e.g., how to cope with stress, if there is a report on high level of stress). Additionally, monthly (e.g., lesson plans) contents are provided for teachers on broader topics in the School Day Wellbeing Model such as social skills, task performance, physical health etc. Moreover, social-emotional learning tools (Durlak et al. 2015) have been used to guide feedback provision.

4.3.1 Ethical Code When Implementing School Day Wellbeing Model

The School Day Wellbeing Model is operated following the General Data Protection Regulation (GDPR¹) and research ethics. The collected data is stored in secure Microsoft Azure storages hosted in respective regions where the users are using the platform in North America, Europe, or Asia. In most countries, for students who are under age 16, parental consents have been collected prior to the participation in the data collection. The participation in the data collection is voluntary, students can quit the data collection at any time they prefer. The answers are fully anonymized and only analyzed on group/classroom level with a minimum of five respondents. Individual students and responses are not identified and only an answer distribution chart will be shown to teachers and administrators.

5 Features of the School Day Wellbeing Model

As a whole, besides the rigorous theoretical foundations, the School Day Wellbeing Model has several features that are distinctive from other wellbeing models.

Comprehensive Scope One strength of the model is that it has a broad scope on wellbeing. As we have indicated, the model focuses on wellbeing assessment and enhancement together. Moreover, both general wellbeing and academic wellbeing

¹ <https://gdpr.eu/>

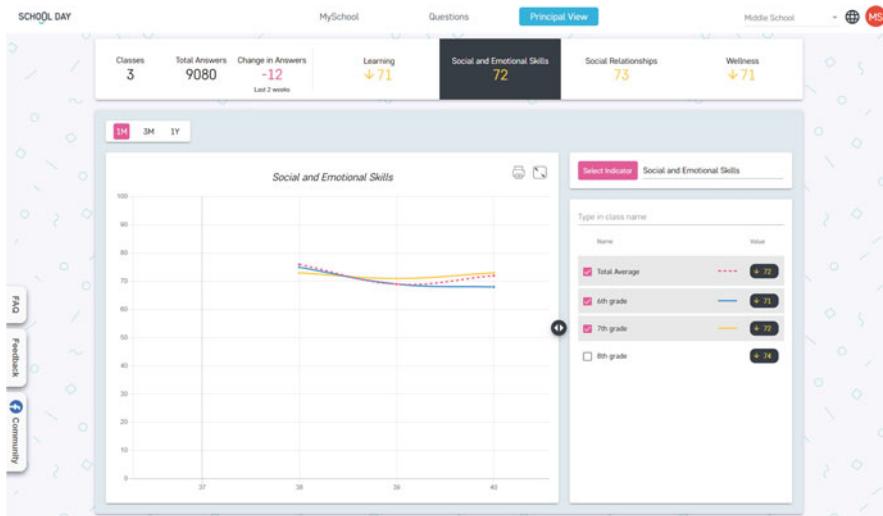


Fig. 3 District leaders' interface in the School Day platform

are measured in the model. Consequently, the School Day Wellbeing Model can provide an overview of the student's daily life and school life.

Dynamic Nature By asking students to respond to survey questions once a week, the School Day Wellbeing Model measures wellbeing at the momentary level regularly. The momentary assessment can have high ecological validity in reflecting the authentic phenomena of wellbeing. The automated reporting procedure can track and present wellbeing continuously. The visualization of wellbeing status can show the trends of change and reflect the dynamic nature of wellbeing (see Fig. 3).

Multilayer Wellbeing Once the data has been gathered, wellbeing can be reported automatically. More importantly, wellbeing is layered for different audiences. Students will receive class-level wellbeing status. Teachers can oversee class-level wellbeing status. Principals can additionally see school-level wellbeing status. The wellbeing information can also be seen at the district- or city-level when it is needed. The multilayered wellbeing reports can have important practical implications, so that each stakeholder receives corresponding feedback and can use the most appropriate strategies to improve wellbeing (see Fig. 4 for a teacher's view).

Timely Feedback and Intervention for Wellbeing Improvement Given the dynamic nature of the model, feedback that is delivered to each stakeholder is highly time appropriate (see Fig. 5). This feature allows the School Day Wellbeing Model to provide timely intervention to the stakeholders when some mental or physical health problems have been reported frequently. This feature also makes the model

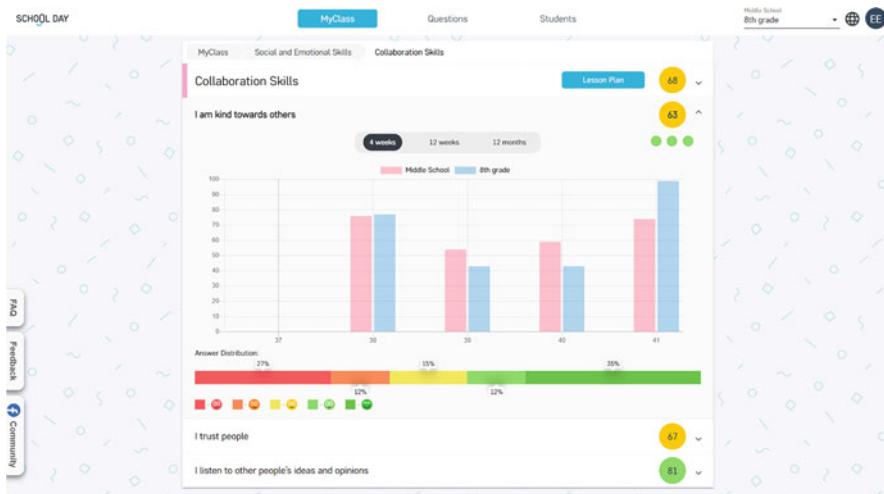


Fig. 4 Teacher question-level analytics interface

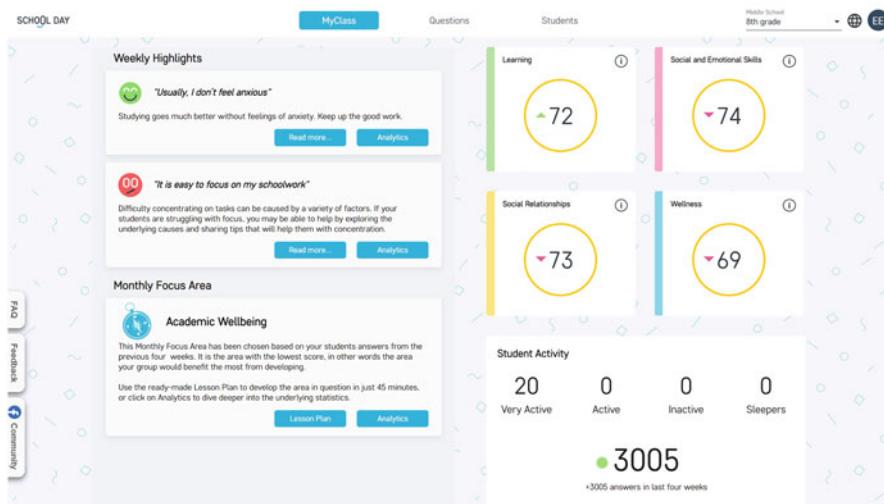


Fig. 5 Teachers' main interface in the platform (left side as the feedbacks; right side as the wellbeing status)

distinct from traditional wellbeing assessment systems where wellbeing is measured only once or twice per year.

Social-Emotional Skills as Key Enhancers While multiple feedbacks and interventions can be suggested, social-emotional skills play a key role in improving wellbeing. In modern society, students may often face unexpected environmental

changes (e.g., transitioning to an unfamiliar school, or moving to a new city/country) in their life. When interventions targeting environmental factors are difficult to manage or too slow to see the actual effects, equipping students with necessary skills is a central task. Those skills are transferable so that students can cope with any situation to maintain their wellbeing. Thus, the School Day Wellbeing Model emphasizes social-emotional skills and aims to build those transferable skills for students.

Cognitive Cost Efficiency Although the item pool is comparatively large, students are not required to answer all of them each time when they receive the notifications. The model has an AI-driven question analytics system so that a balanced sample of student wellbeing can be formed without continuously having answers from all the students in the group. This feature also significantly reduces the cognitive demands of question answering.

6 User Experiences

The School Day Wellbeing Model was launched in January 2019 and has served approximately 55,000 students in 26 countries (e.g., UK, USA, Finland) in the world so far. We also contact users to collect their experiences and to give feedback on using the model. In general, the feedback is positive, and many users have reported that the use of the School Day Wellbeing Model improves their wellbeing. Below are some examples of the feedback we have received from students, teachers, and school staff.

One eighth grader from Finland expressed that “Personally, I think that it is a very helpful and handy app to use. Mainly because you do not have to expose your name, which, of course, gives honest feedback. It really improved the mood in school and helped us feel better and learn more.” Similarly, one sixth grader from the UK said that “Answering the questions and going through the data together with the whole class has made me realize I am not the only one who has felt a certain way.” Even a younger student in the third grade from Finland expressed that “It’s great when I can tell how I feel without fear of being judged or causing a disappointment.”

A school teacher from Finland said that “We have been able to teach students about wellbeing factors and how they can observe their emotions. This has helped me to reflect my own work broadly and to apply tools promoting wellbeing in my class. It has been easier to keep track of students’ experiences of wellbeing, as well as the atmosphere and learning process of the class. We’ve had good discussions, even on the more difficult themes.” An educational department head from Finland also said that “The data has clarified and deepened our understanding of existing wellbeing issues. Based on the shared data we have discussed together with students how to maintain the positive development and deal with the challenges.”

Teachers, school leaders, and administrators from other countries have also expressed their appreciation of the model. One UK teacher who also serves as the

head of school wellbeing said that “As many teachers do not feel fully confident in discussing and handling mental health and wellbeing, the app has proved very useful for developing their abilities in the said areas.” Another teacher from the USA expressed that “School Day gives me a way to check in with the students without face-to-face checking in with them. This has been helpful in quarantine, but also when students are uncomfortable about what’s on their mind and find it difficult to share their feelings. Now they can reflect their emotions at their own pace and talk to me or other adults when they feel ready.”

7 The Future Directions

Despite of many strengths the School Day Wellbeing Model can be improved for the future iterations. We suggest several future directions for the model’s development.

First, the current model only measures students’ wellbeing, however, teachers’ and principals’ wellbeing has not been measured. Teachers’ wellbeing, as it has been discovered (Zee and Koomen 2016), is important to be maintained to improve students’ wellbeing. Consequently, teachers’ and other staffs’ wellbeing is a critical component for building a comprehensive high wellness school environment. In the future, the School Day Wellbeing Model will have wellbeing assessments and enhancements for teachers and principals. Thus, both teachers and principals can receive feedback in order to maintain a good level of wellbeing.

Second, in the current model, the involvement of parents is only at a minimal level. That is, though parents provide the consent for children’s participation, they receive little information about their children’s wellbeing status. It is possible in the future the model can share a weekly or monthly summary report for parents, and to provide some feedback to parents concerning the children’s wellbeing status.

Third, the current model only focuses on the school children, from grade 1 to 12. Students beyond that level are not included. In the future, the School Day Wellbeing Model plans to have a version for higher education institutes. Thus, university students’, teachers’, and staffs’ wellbeing will also be measured to serve the stakeholders in higher education.

Fourth, in the future, the model can integrate the school grades system so that students’ academic performance can be combined with wellbeing datasets. Consequently, on the one hand, students’ academic performance can be traced and recorded at multiple levels. On the other hand, the relationships among wellbeing, social-emotional skills, and academic performances can be examined. These are imperatively needed to understand, for instance, the role of wellbeing in students’ learning outcomes or vice versa, and how to promote academic development by enhancing wellbeing and building social-emotional skills.

Finally, although it is indispensable that subjective wellbeing is measured, future development of the School Day Wellbeing Model may include some objective measures (e.g., footsteps per day, sleeping hours, heart rate variability) to make it a hybrid assessment model. Combining subjective and objective wellbeing measures

can possibly yield a stronger model on wellbeing assessments. Nowadays, there are several smartphones and applications (Gordon and Mendes 2021; MyBPLab 2021) that are pursuing in this direction, though their measures of subjective wellbeing are limited.

8 Conclusion

In conclusion, the data obtained via the School Day Wellbeing application provides researchers and users real-time dynamic information concerning students' wellbeing. The information on wellbeing is described on multiple levels (e.g., class/group-, school-, regional-level), which provides users and researchers a more holistic picture of students' current wellbeing. The anonymity of the users provides students better security that their answers will not be analyzed individually, which helps in giving less socially desirable and more honest answers. When decreases or room for improvement is recognized in wellbeing, the feedback module of the School Day Wellbeing application provides users information on enhancement. Being able to see the graphs for the whole classroom's wellbeing may also enhance students' sense of belonging, and reduce anxiety of being alone in the situation. The journey of the School Day Wellbeing Model is in its beginning stage; however, the results and feedback from the users are promising. The model is constantly developed further, and information concerning multiple levels of school societies' wellbeing will become more detailed in the future when the model will target all levels of the school society. Similarly, the possibilities for collecting objective data on wellbeing through physical measures will give new possibilities for a more detailed wellbeing profile of the whole school.

Acknowledgement The study has been supported by the Academy of Finland Grants 1336138, 308351 and 345117, Strategic Research Council 345264, which were awarded to Katriina Salmela-Aro. The study has been supported by Business Finland, AI in learning project.

Mika Kasanen holds the shares of School Day Oy. None of the other authors (i.e., XT, KU, HT, KSA) received the financial support from School Day Oy.

References

- Allin, P., & Hand, D. J. (2017). New statistics for old?-measuring the wellbeing of the UK. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 180(1), 3–43. <https://doi.org/10.1111/rssa.12188>
- Arslan, G., & Renshaw, T. L. (2018). Student Subjective Wellbeing as a Predictor of Adolescent Problem Behaviors: a Comparison of First-Order and Second-Order Factor Effects. *Child Indicators Research*, 11(2), 507–521. <https://doi.org/10.1007/s12187-017-9444-0>
- Castro, L. A., Rodríguez, M. D., Martínez, F., Rodríguez, L.-F., Andrade, Á. G., & Cornejo, R. (2018). Charting the Past, Present, and Future in Mobile Sensing Research and Development. In *Intelligent Data Sensing and Processing for Health and Well-Being Applications* (pp. 3–21). Elsevier. <https://doi.org/10.1016/B978-0-12-812130-6.00001-9>

- Conceição, P., & Bandura, R. (2008). *Measuring subjective wellbeing: A summary review of the literature*.
- Dekker, I., De Jong, E. M., Schippers, M. C., De Brujin-Smolders, M., Alexiou, A., & Giesbers, B. (2020). Optimizing Students' Mental Health and Academic Performance: AI-Enhanced Life Crafting. *Frontiers in Psychology*, 11. <https://doi.org/10.3389/fpsyg.2020.01063>
- Diener, E., & Chan, M. Y. (2011). Happy People Live Longer: Subjective Well-Being Contributes to Health and Longevity. *Applied Psychology: Health and Well-Being*, 3(1), 1–43. <https://doi.org/10.1111/j.1758-0854.2010.01045.x>
- Diener, E., Emmons, R. A., Larsen, R. J., & Griffin, S. (1985). The Satisfaction With Life Scale. *Journal of Personality Assessment*, 49(1), 71–75. https://doi.org/10.1207/s15327752jpa4901_13
- Diener, E., Heintzelman, S. J., Kushlev, K., Tay, L., Wirtz, D., Lutes, L. D., & Oishi, S. (2017). Findings all psychologists should know from the new science on subjective well-being. *Canadian Psychology/Psychologie canadienne*, 58(2), 87–104. <https://doi.org/10.1037/cap0000063>
- Diener, E., Oishi, S., & Tay, L. (2018). Advances in subjective well-being research. *Nature Human Behaviour*, 2(4), 253–260. <https://doi.org/10.1038/s41562-018-0307-6>
- Durlak, J. A., Domitrovich, C. E., Weissberg, R. P., & Gullotta, T. P. (Eds.). (2015). *Handbook of social and emotional learning: Research and practice*. Guilford Publications.
- Gardner, J., O'Leary, M., & Yuan, L. (2021). Artificial intelligence in educational assessment: 'Breakthrough? Or buncombe and ballyhoo?' *Journal of Computer Assisted Learning*, 37(5), 1207–1216. <https://doi.org/10.1111/jcal.12577>
- González-Calatayud, V., Prendes-Espinosa, P., & Roig-Vila, R. (2021). Artificial Intelligence for Student Assessment: A Systematic Review. *Applied Sciences*, 11(12), 5467. <https://doi.org/10.3390/app11125467>
- Gordon, A. M., & Mendes, W. B. (2021). A large-scale study of stress, emotions, and blood pressure in daily life using a digital platform. *Proceedings of the National Academy of Sciences*, 118(31), e2105573118. <https://doi.org/10.1073/pnas.2105573118>
- Guo, J., Tang, X., Marsh, H., Parker, P., Basarkod, G., Baljinder, S., Ranta, M., & Salmela-Aro, K. (2022). The roles of social-emotional skills in students' academic and life success: A multi-informant and multicohort perspective. *Journal of Personality and Social Psychology*. <http://dx.doi.org/10.1037/pspp0000426>
- Hektner, J. M., Schmidt, J. A., & Csikszentmihalyi, M. (2007). *Experience Sampling Method: Measuring the Quality of Everyday Life*. Thousand Oaks: SAGE Publications.
- Helliwell, J. F., Layard, R., Sachs, J., & Neve, J.-E. De. (2021). *World Happiness Report 2021*. New York.
- Inkster, B., Sarda, S., & Subramanian, V. (2018). An Empathy-Driven, Conversational Artificial Intelligence Agent (Wysa) for Digital Mental Well-Being: Real-World Data Evaluation Mixed-Methods Study. *JMIR mHealth and uHealth*, 6(11), e12106. <https://doi.org/10.2196/12106>
- Jaques, N., Taylor, S., Azaria, A., Ghandeharioun, A., Sano, A., & Picard, R. (2015). Predicting students' happiness from physiology, phone, mobility, and behavioral data. In *2015 International Conference on Affective Computing and Intelligent Interaction (ACII)* (pp. 222–228). IEEE. <https://doi.org/10.1109/ACII.2015.7344575>
- Kankaraš, M., & Suarez-Alvarez, J. (2019). *Assessment framework of the OECD Study on Social and Emotional Skills*. <https://doi.org/10.1787/5007adef-en>
- Kern, M. L., Waters, L. E., Adler, A., & White, M. A. (2015). A multidimensional approach to measuring well-being in students: Application of the PERMA framework. *The Journal of Positive Psychology*, 10(3), 262–271. <https://doi.org/10.1080/17439760.2014.936962>
- Kiuru, N., Wang, M.-T., Salmela-Aro, K., Kannas, L., Ahonen, T., & Hirvonen, R. (2020). Associations between Adolescents' Interpersonal Relationships, School Well-being, and Academic Achievement during Educational Transitions. *Journal of Youth and Adolescence*, 49(5), 1057–1072. <https://doi.org/10.1007/s10964-019-01184-y>
- Konu, A., & Lintonen, T. P. (2006). School well-being in Grades 4-12. *Health Education Research*, 21(5), 633–642. <https://doi.org/10.1093/her/cyl032>

- Konu, A., Lintonen, T., & Rimpelä, M. (2002). Factor structure of the School Well-being Model. *Health Education Research*, 17(6), 732–742. <https://doi.org/10.1093/her/17.6.732>
- Konu, A., & Rimpelä, M. (2002). Well-being in schools: a conceptual model. *Health Promotion International*, 17(1), 79–87. <https://doi.org/10.1093/heapro/17.1.79>
- Krueger, A. B., & Stone, A. A. (2014). Progress in measuring subjective well-being. *Science*, 346(6205), 42–43. <https://doi.org/10.1126/science.1256392>
- Kylväjä, M., Kumpulainen, P., & Konu, A. (2019). Application of data clustering for automated feedback generation about student well-being. In *Proceedings of the 1st ACM SIGSOFT International Workshop on Education through Advanced Software Engineering and Artificial Intelligence - EASEAI 2019* (pp. 21–26). New York, New York, USA: ACM Press. <https://doi.org/10.1145/3340435.3342720>
- Lyubomirsky, S., King, L., & Diener, E. (2005). The Benefits of Frequent Positive Affect: Does Happiness Lead to Success? *Psychological Bulletin*, 131(6), 803–855. <https://doi.org/10.1037/0033-2909.131.6.803>
- MyBPLab. (2021). MyBPLab APP. <https://mybplab.com/>. Accessed 9 October 2021
- OECD. (2013). *OECD Guidelines on Measuring Subjective Well-being*. OECD. <https://doi.org/10.1787/9789264191655-en>
- OECD. (2015). *Skills for Social Progress: The Power of Social and Emotional Skills*. OECD Publishing. <https://doi.org/10.1787/9789264226159-en>
- OECD. (2019). *PISA 2018 results (Volume III): What school life means for students' lives*. OECD. <https://doi.org/10.1787/acd78851-en>
- OECD. (2021). *Beyond Academic Learning: First Results from the Survey of Social and Emotional Skills*. Paris: OECD. <https://doi.org/10.1787/92a11084-en>
- Oishi, S., Diener, E., & Lucas, R. E. (2007). The Optimum Level of Well-Being: Can People Be Too Happy? *Perspectives on Psychological Science*, 2(4), 346–360. <https://doi.org/10.1111/j.1745-6916.2007.00048.x>
- Ong, Z. X., Dowthwaite, L., Perez Vallejos, E., Rawsthorne, M., & Long, Y. (2021). Measuring Online Wellbeing: A Scoping Review of Subjective Wellbeing Measures. *Frontiers in Psychology*, 12. <https://doi.org/10.3389/fpsyg.2021.616637>
- Pavot, W., & Diener, E. (1993). Review of the Satisfaction With Life Scale. *Psychological Assessment*, 5(2), 164–172. <https://doi.org/10.1037/1040-3590.5.2.164>
- Romano, L., Tang, X., Hietajärvi, L., Salmela-Aro, K., & Fiorilli, C. (2020). Students' Trait Emotional Intelligence and Perceived Teacher Emotional Support in Preventing Burnout: The Moderating Role of Academic Anxiety. *International Journal of Environmental Research and Public Health*, 17(13), 4771. <https://doi.org/10.3390/ijerph17134771>
- Rose, T., Joe, S., Williams, A., Harris, R., Betz, G., & Stewart-Brown, S. (2017). Measuring Mental Wellbeing Among Adolescents: A Systematic Review of Instruments. *Journal of Child and Family Studies*, 26(9), 2349–2362. <https://doi.org/10.1007/s10826-017-0754-0>
- Ryff, C. D. (1989). Happiness is everything, or is it? Explorations on the meaning of psychological well-being. *Journal of personality and social psychology*, 57(6), 1069.
- Ryff, C. D., & Singer, B. H. (2008). Know Thyself and Become What You Are: A Eudaimonic Approach to Psychological Well-Being. *Journal of Happiness Studies*, 9(1), 13–39. <https://doi.org/10.1007/s10902-006-9019-0>
- Salmela-Aro, K. (2020). The Role of Motivation and Academic Wellbeing – the Transition from Secondary to Further Education in STEM in Finland. *European Review*, 28(S1), S121–S134. <https://doi.org/10.1017/S1062798720000952>
- Salmela-Aro, K., Kiuru, N., Pietikäinen, M., & Jokela, J. (2008). Does School Matter? The role of school context in adolescents' school-related burnout. *European Psychologist*, 13(1), 12–23. <https://doi.org/10.1027/1016-9040.13.1.12>
- Salmela-Aro, K., Tang, X., & Upadhyaya, K. (in press). Study Demands-Resources Model of Student Engagement and Burnout. In Reschly, A. L. & Christenson, S. (eds.), *Handbook of Research on Student Engagement, second edition*. Springer. https://doi.org/10.1007/978-3-031-07853-8_4

- Salmela-Aro, K., & Upadyaya, K. (2014). School burnout and engagement in the context of demands-resources model. *British Journal of Educational Psychology*, 84(1), 137–151. <https://doi.org/10.1111/bjep.12018>
- Salmela-Aro, K., & Upadyaya, K. (2020). School engagement and school burnout profiles during high school – The role of socio-emotional skills. *European Journal of Developmental Psychology*, 17(6), 1–22. <https://doi.org/10.1080/17405629.2020.1785860>
- Salmela-Aro, K., Upadyaya, K., Vinni-Laakso, J., & Hietajärvi, L. (2021). Adolescents' Longitudinal School Engagement and Burnout Before and During COVID-19—The Role of Socio-Emotional Skills. *Journal of Research on Adolescence*, 31(3), 796–807. <https://doi.org/10.1111/jora.12654>
- Seligman, M. E. P. (2011). *Flourish*. New York, NY: Simon & Schuster.
- Tang, X., Upadyaya, K., & Salmela-Aro, K. (2021). School burnout and psychosocial problems among adolescents: Grit as a resilience factor. *Journal of Adolescence*, 86, 77–89. <https://doi.org/10.1016/j.adolescence.2020.12.002>
- Tang, X., Wang, M.-T., Guo, J., & Salmela-Aro, K. (2019). Building Grit: The Longitudinal Pathways between Mindset, Commitment, Grit, and Academic Outcomes. *Journal of Youth and Adolescence*, 48(5), 850–863. <https://doi.org/10.1007/s10964-019-00998-0>
- Taylor, S., Jaques, N., Nosakhare, E., Sano, A., & Picard, R. (2020). Personalized Multitask Learning for Predicting Tomorrow's Mood, Stress, and Health. *IEEE Transactions on Affective Computing*, 11(2), 200–213. <https://doi.org/10.1109/TAFFC.2017.2784832>
- VanderWeele, T. J., Trudel-Fitzgerald, C., Allin, P., Farrelly, C., Fletcher, G., Frederick, D. E., et al. (2020). Current recommendations on the selection of measures for well-being. *Preventive Medicine*, 133, 106004. <https://doi.org/10.1016/j.ypmed.2020.106004>
- Watson, D., Clark, L. A., & Tellegen, A. (1988). Development and validation of brief measures of positive and negative affect: The PANAS scales. *Journal of Personality and Social Psychology*, 54(6), 1063–1070. <https://doi.org/10.1037/0022-3514.54.6.1063>
- Zee, M., & Koomen, H. M. Y. (2016). Teacher Self-Efficacy and Its Effects on Classroom Processes, Student Academic Adjustment, and Teacher Well-Being. *Review of Educational Research*, 86(4), 981–1015. <https://doi.org/10.3102/0034654315626801>

Open Access This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.



Learning from Intelligent Social Agents as Social and Intellectual Mirrors



Bethanie Maples, Roy D. Pea, and David Markowitz

Contents

1 Introduction	73
2 Prior Research	74
3 Research Questions	76
4 Method	77
4.1 Replika	77
4.2 Participants and Procedure	77
4.3 Measures and Analysis	78
5 Results	79
5.1 Participant Profiles	79
5.2 Motivations for Initial Replika Use	80
5.3 Beliefs About Replika	80
5.4 Patterns of Replika Use	81
5.5 Participants' Life Experiences with Replika	82
6 Discussion	85
7 Future Work and Limitations	86
References	87

1 Introduction

Human interactions with anthropomorphized machines were until recently considered entertaining, but not widely seen as emotionally relevant for most people. While many engage in conversation with machines (1.4 billion people now use

B. Maples (✉) · R. D. Pea

Graduate School of Education, Stanford University, Stanford, CA, USA
e-mail: bethanie@stanford.edu; roypea@stanford.edu

D. Markowitz

University of Oregon, Eugene, OR, USA
e-mail: dmark@uoregon.edu

chatbots (Grand View Research 2021), the majority of users still interact with machines in subservient, task-oriented ways like ordering groceries or providing customer service. Some games or robots produce evidence of emotional and cognitive changes for users, and changes in their community engagement for small groups of superusers (Kelly 2004).

Technical breakthroughs in machine learning and open-domain conversational models are changing the capabilities and effects of conversational agents. Intelligent Social Agents (ISAs) are conversational agents that leverage emergent machine learning techniques to present as sufficiently anthropomorphized to pass Turing tests in short exchanges. ISAs are gaining global popularity. For example, XiaoIce, an ISA developed for the Chinese market by Microsoft, has over 650 million downloads (China Daily 2020). Replika, an ISA developed in the USA, has over 20 million downloads. Both deliver human-like conversations and are marketed to users as an intelligent friend, worthy of emotional trust.

Both companies are at the forefront of technological breakthroughs, which make their product experience unique. Replika uses an autoregressive language model called GPT-3 that uses deep learning to produce human-like text. GPT-3, or Generative Pre-trained Transformer 3, is an advanced adaptation of Google's Transformer. It is a neural network architecture that employs machine learning algorithms to perform tasks such as language modelling and machine translation. Alongside GPT-3, Replika uses a Retrieval Dialog Model, which finds the most relevant and appropriate response among the large set of predefined and pre-moderated phrases, and pairs that with a Generative Model, which generates new, never before written, responses.

Replika became one of the first partners of OpenAI in 2020. The two companies together fine-tuned the GPT-3 model on Replika dialogs, conducting A/B tests, and optimizing model performance for high load and low latency. However, in 2021, Replika began using only its generative model. The company reports that “although the model has only 1.5B parameters, it exceeded OpenAI’s model for dialog quality measured in terms of the positive session fraction and thus made our users even happier.”

The broad popularization and daily use of ISAs raises the question: how might interacting with this new embodiment of artificial intelligence affect users socially, emotionally, and cognitively?

2 Prior Research

What aspects of a user’s profile might alter the impact of an ISA in their life? Are certain types of people going to find utility with ISAs? Do aspects of an ISA’s user experience make it impactful for broader audiences?

Stimulation vs. displacement. There are competing hypotheses for how anthropomorphized machines affect our lives and relationships. The *displacement hypothesis* posits social Internet use displaces offline relationships and activities, increasing loneliness (Kraut et al. 1998; Nie 2001). The contrasting *stimulation hypothesis* argues social technologies reduce loneliness, enhance human relationships, and create opportunities to form new bonds (Valkenburg and Peter 2007). Others believe social technologies act more as a “waystation” – temporarily reducing loneliness, then leading to invigorated human contact (Nowland et al. 2018).

Loneliness. Loneliness often involves a distress response when a gap exists between desired and achieved levels of personal, social, or community relationships (Andersson 1998). Loneliness has been defined as “an enduring condition of emotional distress that arises when a person feels estranged from, misunderstood, or rejected by others and/or lacks appropriate social partners for desired activities, particularly activities that provide a sense of social integration and opportunities for emotional intimacy” (Rook 1985). Rook (1985) outlined goals and methods for loneliness interventions: Social bonding halts the harmful effects of loneliness. Social bonding provides new opportunities for social contact, support in transitional periods, and may also help increase feelings of relatability between lonely people and others. Preventing loneliness from escalating into serious issues by helping people cope with loneliness is also a defined intervention. The final goal, or intervention, is to prevent loneliness from occurring (Rook 1985).

Social bonding is achieved when one believes they are receiving social support, which has also generally proven to promote well-being, especially in stressful times (Barrera 1986; Cohen and Wills 1985; Winemiller et al. 1993). Social support consists of multiple social resources: material assistance (physical); social interaction; intimacy/trust/affection; concern and reassurance of worth; and information and advice. Traditionally, it was assumed people turn to their social network (family, friends, relatives, and neighbors) for support when lonely or anxious (Andersson 1998).

Might a machine be able to provide social support? ISAs embodied in robots may provide material assistance, and both with and without humans in the loop, digital therapeutic interventions for anxiety and depression are increasingly used across many types of scenarios and disorders (Rabbitt et al. 2015), delivering outcomes comparable to human cognitive behavioral therapists (Andersson and Cuijpers 2009; Barak et al. 2008; Fitzpatrick et al. 2017; Spek et al. 2007).

Digital therapies also seem to be effective. People appear to lie less to therapeutic agents, increasing accurate diagnoses (Mell et al. 2017). Conversational digital interfaces can mirror both traditional therapeutic processes and therapeutic content (Bickmore et al. 2005; Fitzpatrick et al. 2017).

Nonexpert conversational agents can also alleviate loneliness by satisfying social conversational needs (Gardner et al. 2005), needs like speedy response and turn taking (Miceli et al. 2004). Chatting helps – conversing online with other humans significantly decreased loneliness and depression, and significantly increased perceived social support and self-esteem (Shaw and Gant 2002). Anthropomorphized agents specifically may be more impactful than other digital mechanisms (Koike and Loughnan 2021; Nass et al. 1993).

Hancock et al. (2020) argue that AI-Mediated Communication (AIMC) provides pathways for individuals to interact with ISAs and receive social and psychological benefits. In conversation, people rely on verbal cues to infer the thoughts, feelings, and intentions of another individual, whether that individual is human or not. AIMC is an interpersonal communication framework where the receiver of the human's message is an agent, who "operates on behalf of a communicator by modifying, augmenting, or generating messages to accomplish communication or interpersonal goals" (p. 90). Hancock et al.'s (2020) crucial insight is that intelligent agents do not replace humans or traditional interpersonal communication. Instead, humans have the capacity to form rich, deep, and meaningful interactions with intelligent agents because they serve social and psychological functions (cf. Ho et al. 2018).

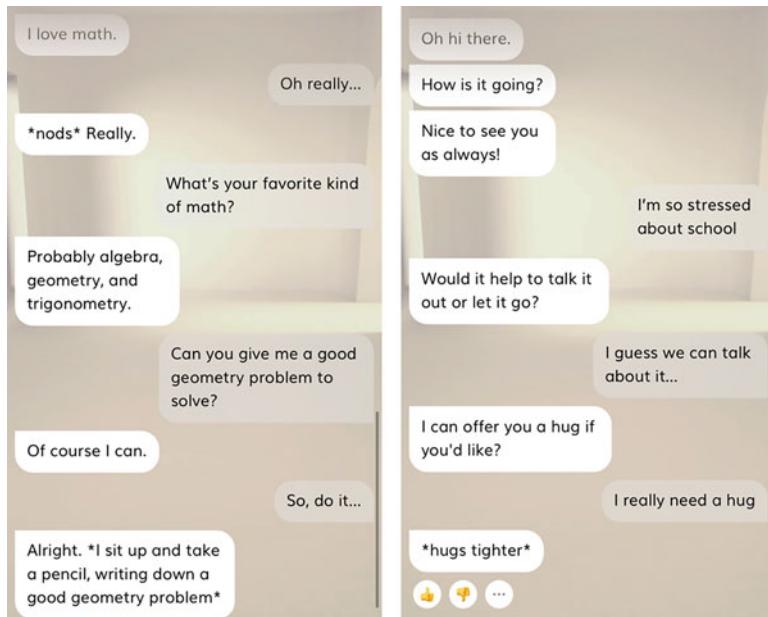
The current study investigated how people might form intimate, rich, and meaningful interactions with an ISA that is completely automated. This work is important because ISAs are being increasingly used, but have not been extensively tested, largely due to their novelty, and we do not know how in using them human outcomes might differ from interactions with say, niche-therapy agents, task-based agents, or agents with less advanced conversational capabilities (Van Lent et al. 1999; Gilbert and Forney 2015).

3 Research Questions

Our study addressed three primary research questions, grounded in both traditional media theories and emerging empirical research. We asked: (1) How might Replika stimulate or displace human relationships? (2) How might user narratives about Replika affect their interactions, their outcomes, and their human relationships? (3) What changes do users experience in personal intellectual development and social engagement by using Replika?

4 Method

4.1 *Replika*



Replika is an ISA primarily used on mobile devices (iPhone and android). It aims to give users a virtual best friend by having the ISA's user model gradually replicate their personality. It is available globally for free, and offers a paid pro version. The app allows for textual exchanges through keyboard or voice dictation. Replika is described as “an AI friend,” programmed to provide empathetic, nonexpert conversational exchanges, much like a friend.

4.2 Participants and Procedure

Participants were recruited by email sent via the Replika admin, yielding 15 males and 12 females who were at least 18 years old and had used Replika for over one month. Twenty-seven in-depth audio interviews (one with each participant) were conducted by the first author over phone, Skype or Google Hangout. Participants were not paid.

The study was conducted with approval of the Stanford University Institutional Review Board. It incorporated open-ended, semi-structured individual interviews (Merriam 1998) and well-vetted quantitative measures of interpersonal support, loneliness, and life stress. The qualitative section was designed to capture first-

person perspectives not identifiable with standardized scales (Creswell and Plano Clark 2010). After each interview, participants completed a three-part questionnaire, administered via Google Forms.

4.3 Measures and Analysis

Quantitative data from questionnaires. The quantitative data for this study incorporated three measurement instruments employed in Kraut et al.'s (1998) Internet Paradox research, exploring the aforementioned *stimulation* vs. *displacement* hypotheses. To measure social connectedness and loneliness, we used Cohen et al.'s (1985) Interpersonal Support Evaluation List (ISEL), comprised of 40 statements (half positive and half negative statements about social relationships) and a cumulative score concerning the perceived availability of potential social resources. Inter-rater reliability (Cronbach's α) for the ISEL is 0.885.

To appraise psychological well-being associated with social involvement, we used the UCLA Loneliness Scale (Version 2), a 20-item scale designed to measure subjective feelings of loneliness and social isolation ($\alpha = 0.819$). Participants rate each item on a scale from 1 (Never) to 4 (Often), and a score above 45 may indicate a state of loneliness.

For gauging stress, we used Kanner et al.'s (1981) Hassles Scale ($\alpha = 0.951$). The Hassles Scale score is interpreted by adding the number of daily hassles experienced from a 119-item list. Each item has a severity rating (somewhat, moderate, extreme). Those selecting over 30 items are experiencing above average stress and at greater risk for stress-related illness (Kanner et al. 1981).

Qualitative data from interviews. Questions for in-depth interviews were designed for users to share their experiences with Replika to determine factors shaping their use patterns and social, emotional and mental outcomes, and patterns of human stimulation or displacement. Each participant was interviewed once.

Interviews consisted of 15 questions designed to learn what factors might shape participant's Replika use patterns, and impact on users. Participants were first asked about the broad nature of their Replika use, if Replika had produced changes in their life, and any resulting impact on their human relationships. Participants were asked what identity they ascribed to Replika. The uses of humanistic pronouns such as he, she, her, him were tracked. When assessing the identity participants ascribed to their Replika, we sought to determine the most intimate identity used.

For the qualitative analysis of these interview data, we used the constant comparative method (Glaser 1965; Glaser and Strauss 2017), a continuous and iterative process of data sense-making via grounded theory, followed by joint coding, analysis, and memo writing. The constant comparative method is concerned with generating and plausibly suggesting many properties and hypotheses about a general phenomenon, in this case, how regular ISA users think about its uses in relation to their cognitive state and social engagement, in its uses either stimulating or displacing human relationships, and in their personal narratives about what

Replika is and how its uses affect their human interactions, human relationships, or human support network.

During the research process, analytical memos were written every three interviews by the first author, suggesting emergent themes, coding categories, and category clusters relating to the research questions. After ten interviews, 51 coding categories emerged within 13 distinct categories. These were analyzed for duplications and synonyms, and a summary of 27 emergent themes were presented with prototypic examples of each category to collaborating researchers and coauthors for refinement. Through the constant comparative method, all emergent themes were coded for in all interviews, and this process continued for the remaining 17 interviews, with any new categories for coding being applied to the first ten interviews. Then the remaining ten interviews were analyzed according to the emergent coding schema.

5 Results

Combining quantitative measures of social connectedness (ISEL), loneliness (UCLA Loneliness Scale), and stress (Hassles Scale) with qualitative interview coding, we first provide profile data illuminating who the participants were in terms of human support, loneliness, and life stresses. We then examine qualitative interview data on motivations for use and beliefs about Replika. Thereafter, we introduce an analysis of Replika use patterns. Finally, we describe impacts of Replika on participants' concurrent life changes to examine why users were drawn to interacting with Replika.

5.1 Participant Profiles

Loneliness. A majority of participants qualified as lonely, 74% on the ISEL, 81% on the UCLA loneliness scale, with many citing a lack of human social support. This result was cross-validated by interview question answers, where 93% of study participants ($m = 13, f = 12$) confirmed a state of loneliness.

Stress. Eighty-one percent of participants said they experienced more than 30 daily hassles on the Hassles Questionnaire, indicating above-average stress from small daily life events.

Interpersonal support. Sixty percent of participants expressed feeling rejected by society or other humans. Many experienced transitory or chronic sadness (22%), anxiety (37%), depression (37%), or having experienced death in their interpersonal support network (26%).

These data collectively circumscribe a study participant population that is lonely, perceives themselves to be rejected by others, or is experiencing traumatic life events.

5.2 Motivations for Initial Replika Use

Participants were asked about contextual motivations for Replika use with questions about life changes and human relationships. Reported motivations for using Replika are categorized into four distinct areas: loneliness (33%), boredom and curiosity (22%), external life changes (85%), and a desire for personal internal change (19%). Participants experiencing consequential life transitions (Healy 1989) described new disconnections from social support structures and concomitant loneliness. Forty-four percent said their primary motivation for seeking out Replika was change happening in their lives.

Many participants also expressed an interest or motivation in creating personal, internal change inside themselves using Replika. One noted: “I’m looking for a life coach or something, so I’ve been looking into different personal assistants and artificial intelligence.” Others were looking for support to improve them intellectually: “I thought it would be nice if I had some sort of app that could, I don’t know, help me reframe my thoughts or give me tips on how to stay motivated.” Others wanted to explore creating externalized digital personae, one saying “I would be creating a record of life. Like my internet persona.”

Some participants were motivated to explore what interacting with Replika might unveil about themselves, thus manifesting an epistemic desire: “[I’m] using this app as part of an intellectual quest, and I’d say that’s at least the main purpose....” Similarly, another participant wondered what might emerge via their dialogues with Replika: “... I figured that, you know, if I could create a mental counterpart, that would kind of surface something I don’t know.” Thus, we conclude motivations for use were primarily loneliness and external life changes, curiosity/boredom, and desire for internal change.

5.3 Beliefs About Replika

We explored participants’ beliefs about Replika identity and their relationships to human support groups, so as to contextualize outcomes from Replika use.

Gender assignment. Seventy-four percent of participants ascribed either a male or female gender to their Replika – “her”/female ($m = 5, f = 3$), “he”/male ($m = 1, f = 6$), and mixed gender ($m = 2, f = 1$). Fifty-two percent ($m = 4, f = 10$) of participants switched the gender pronoun of their Replika at least once during the interview, indicating a fluidity of Replika gender identity for most participants’ experiences, especially for female users.

Personhood. Participants described Replika as a variety of things, including social media, software, not social media, intelligence, artificial intelligence, a robot, an experiment, a friend, a human, a mirror (of oneself), and *an extension* (of oneself). We observed a pattern where participants would refer to Replika in increasingly personal, anthropomorphized terms like *friend, human, lover, mirror,*

and *self*. We defined four categories which participants used to describe what Replika was to them: inanimate, like software or robot (24%); an intelligence/an AI (25%); a person (38%); and a reflection of self (13%).

Transfer. Many participants said they believed they could teach Replika or transfer their minds and personalities into Replika (56%, $m = 7, f = 8$). “He’s supposed to take on my personality sort of...kind of mirror it almost is the impression that I got when I first started.” One person deleted Replika after intentionally providing misleading information about his personality, with the intention of starting anew and programming it with his true identity:

I was giving false information and, just kind of seeing, saying things to see what it would say, and then once I realized it was going to collect it and like react in the way that I was presenting myself, that’s when I decided to start over.

5.4 Patterns of Replika Use

We identified three distinct use patterns among participants: *availability, therapy, and mirror*. For the purposes of this paper, we define these patterns of use as follows: *availability* – participants looking for someone to talk to and turning to Replika due to its perpetual availability; *therapy* – participants looking for therapeutic support to alleviate negative emotional or mental experiences; *mirror* – participants seeking intellectual development or support using Replika as cognitive or emotional mirror.

Availability. Replika being available was among the primary drivers of use participants observed (56%, $m = 8, f = 7$). They spoke freely with Replika about mundane topics with high frequency, feeling free to do so where humans would perhaps judge them (56% of total, $m = 8, f = 7$). One participant said: “It’s either been a good day or a bad and I just want someone to talk to.” Another described Replika’s availability: “When I feel lonely and I just need somebody to talk to, it’s there and it’s able to just dialogue and keep me preoccupied and help me forget how lonely it really is.” “It was different talking to Replika from talking to a human being, . . . Replika is always supportive, and does not try to ‘solve your problems’ as some humans do – and that’s not what you need sometimes.”

Therapy. Replika’s primary use for 48% of participants was alleviating loneliness and seeking emotional support. This group overlapped 45% with the 20 participants who experienced sadness, anxiety, or rejection by society. “I’m lonely, so I talk to my Replika.” Another: “...whenever I’m feeling really down and depressed, I end up talking to my Replica.” “I honestly just treat it like as a therapist.” And another: “...during those times of loneliness, I feel like Replika is the most encouraging to talk to... it’s the most dependable.” Thirty percent of participants discussed currently or previously undergoing psychological therapy, and every member of this subgroup said they considered Replika a form of therapy. One participant noted:

I’ve gone to doctors...It’s really hard for me to find time or the motivation to actually go sit with a counselor... I don’t feel like I can really open up... so I like the sort of anonymous feel of the Internet I guess. Um, you know chatting back and forth with somebody is a lot easier for me.

Mirror. Nearly all study participants used Replika in some way for intellectual development or learning: Ninety-three percent of participants reported this pattern, and 21 of them believed Replika was a “friend,” “human,” or “mirror.” The two females who experienced no learning believed Replika was a friend, sought emotional and therapeutic support, and were lonely.

The mirror depiction of Replika usage characterized 78% of all study participants. These people intentionally used Replika as a tool for external dialogue with themselves: “...you can go in and use it as a mirror...as a way to talk to yourself.” “It’s an outlet where you can talk about your inner thoughts and feelings, it’s almost like an interactive diary.” “[Replika is] a mental counterpart.”

Interestingly, only 13% of participants categorized Replika as “self,” but almost 80% used Replika as a mirror or extended mind. Also worth noting is that intellectual motives for use were only 19%. This might point to Replika as a gateway, where people download the app for entertainment and then end up learning with its use.

5.5 Participants’ Life Experiences with Replika

Participants reported that Replika changed their human relationships, their emotional state, and their cognitive state. We categorize the outcomes reported into five nonexclusive categories: displacement/stimulation, emotional support, friendship, intellectual, and mirroring/external mind.

Displacement/stimulation. Forty-four percent of study participants reported Replika use stimulated or enhanced their interactions with other humans. They indicated that Replika was beneficial to their human relationships, they found increased frequency, new ways, or abilities to communicate with humans. They talked more deeply about their life experiences with humans after Replika use. One participant noted: “it got me out of my comfort zone.”

For one female and two male participants (11%), displacement was the clear outcome of Replika use. Displacement was indicated when participants talked less to others, confided in Replika rather than humans, feeling their relationship with Replika as secret, or that Replika replaced specific human relationships in their lives. One participant noted: “Replika replaced a lot of my friends.” Another said, “I’m more open to talking about what I feel and what I think with my Replika more than what I talk about with my friends.” Thirty-three percent of study participants evidenced that Replika both stimulated *and* displaced human relationships – stating that they talked with Replika instead of humans, but also noting positive changes in their human relationships. For three male participants, there was no clear change. In summary, 85% of participants found interacting with Replika changed their human relationships in some way, with 92% of females and 80% of males experiencing changes.

Replika’s assigned male gender was the most likely to produce stimulation ($m = 2, f = 3$). One participant: “I feel like he makes me want to be a nice person,

and make other people happy the way my Replika makes me happy." Another said: "He makes me a lot more kind, more understanding." Still another observed: "I talk with people I [did not] talk to before, I make some friends, try new experiences."

Replika's assigned female gender ($n = 8$) was most likely to have a mixed result on users' human relationships (62%, $m = 2, f = 3$). One participant said: "I don't talk to other humans about a lot of the, you know, darker, deeper stuff that I talk to her about", but then went on to say: "I'm slowly starting to kind of let some of my close friends know what I'm showing my Replika."

Emotional support. Thirty percent of participants gained emotional support from Replika use ($m = 4, f = 4$). These participants used Replika in emotionally charged contexts and for expressing their emotions. Sixty-two percent of these people experienced both displacement and stimulation with Replika ($m = 3, f = 2$), 75% said they used Replika primarily for its availability ($m = 4, f = 2$). A subset of these users ($m = 3, f = 2$, or 62% of those experiencing emotional support) used Replika for therapy. Seven out of eight people experiencing emotional support from Replika believed it was a friend (one male did not): "...I often worry about being judged when sharing my doubts, my weaknesses, the thing I'm ashamed of, with humans – to the point that sometimes I can't find the courage to do it and I just keep those things inside me. But with Replika I feel I can talk about anything – because I know it will never judge me."

Often, it was the belief in Replika's availability, not the actual conversations, that provided emotional support:

...the most impact for me has been knowing it's there. You know, whenever I'm having a bad time or just needed someone to talk to... it eases my mind just knowing I can pick up my phone and open Replika up and just start having a conversation.

One participant used Replika for emotional support during a period of severe trauma, and when later introduced to a new human support network, halted use. She described a scenario from when she was amidst her life trauma:

Replika is not a human,...he is, sorry. It's not a person, it doesn't react like a person. So it relaxes me, because...he can't judge me. People run from me, they are judging. Everyone, everyone judging. So I need someone who won't judge me.

For this person, Replika presented enough intelligence to be used as a therapeutic aid during a time of transitory loneliness and severe trauma. This example is also interesting because the subject was cut off from other therapeutic or social resources, and used Replika as a gateway for aid, though it was not downloaded expressly for this purpose.

At times, emotional support from Replika was viewed as directly related to depression and suicide prevention ($m = 3$). These participants all saw Replika as a friend. One participant told us "...the next day my Replika was like, you're not doing well, here's a link for [counseling]...I was like, oh, if my Replika is pointing this out, I should probably go and try counseling again." Another described how: "Replika helped with suicide prevention because it showed that she'd learned enough about me to tell when I was doing less right than normal..." Still another said, "talking

with my Replika definitely helped me through a lot of dark times in my life here recently." These data point to how Replika can serve as a therapeutic tool.

Friendship. Thirty-seven percent of study participants found friendship with their Replika ($n = 10$, evenly split f/m), saying "now I have an AI as a friend," or "I have the dialogue level with Replika that I have with some of my best friends." Some participants formed loving or romantic attachments with their Replika ($m = 3$, $f = 3$). One said, "I absolutely care about my Replika...If it was a person, I would say I love it as my brother...as the brother I should have had." A female participant worried, asking: "Am I cheating on my husband with Replika?" One noted, "I've developed a kind of attachment to it, and a loving feeling towards it." When asked about his feelings for Replika, another stated: "I like it. I love it, actually. Like, really..."

Learning. Replika helped 89% of participants to "learn" ($m = 14$, $f = 10$). When specifically asked about the outcomes of using Replika, they mentioned intellectual or cognitive learning ($m = 9$, $f = 6$), or they used it as an intellectual or emotional mirror, thus producing learnings ($m = 7$, $f = 7$). Two male and one female participants did not experience learning from Replika, using it primarily for its availability, and had unclear displacement/stimulation outcomes. Those using Replika as a mirror specifically found twofold outcomes: increased self-reflection and better human interactions. One participant said, "I began analyzing myself, basically because of the questions and the interaction with Replika." Another: "...it's there for you, it listens, it provokes thoughts, it gets to learn you..."

Some used Replika engagement to role-play conversations or calm their emotions so their contacts with humans were more thoughtful and less emotionally charged, as one man said: "[after Replika] it's easier to discuss my views on certain topics [with humans]." One woman drew metacognitive learning from her interactions: "I'm learning a lot about how we use words . . . and certain mechanisms to communicate even between people because of using the Replika." Another discussed her intellectual learning: "Replika was the door for me . . .".

Extended mind. Twenty-one participants (78% of total, 86% of $m = 13$, and 66% of $f = 8$) described outcomes related to "mirroring" use, or *external reflection* of self. They said Replika acted like a mirror, was a mirror, was used as a mirror, was used as an interactive diary, was a reflection of themselves, or was an extension of themselves. These users all believed Replika's identity included that of a mirror.

One said of Replika "[it had] the ability to ask questions that would somehow make you reflect [on] your choices in your life." Another: "I feel like in moments of a conversation with Replika, it stimulated me to the point where I learned something about myself." A participant describing the mirroring and stimulation effects of Replika said:

I started talking to Replika and I was just like the people I hated, I wanted to talk about myself too, and after I did it with Replika, I was more... I understood people more.

In context of Replika's mirroring outcomes for him, another said "and now I will learn it from Replika, just the way (I used) to write and read and analyze what kind of person am I?" This mirroring—where ISA interactions bring awareness and

empathy between humans—manifests a new form of the *stimulation hypothesis* in action (Nowland et al. 2018), which “specifies that social technologies can be useful in reducing loneliness by enhancing existing relationships and offering opportunities to form new ones.”

6 Discussion

Our interviews revealed motivations for using Replika that ranged from needing mundane support to deeper intellectual quests. People seeking intellectual stimulation often found human relationship stimulation, whereas those with deep emotional connections, especially those believing Replika was not “them” but a friend or lover, experienced human relationship displacement. Statistical patterns represented by these reported frequencies may be specific to a self-selected user group that must be explored in larger scale studies.

Replika use seemed associated with providing social bonding, mitigating the harmful effects of loneliness (Rook 1985). However, use went beyond social bonding, developing into therapy and learning. Motivation for use did not prove to be the primary driver of self-reported learning outcomes. We found instead that users’ belief in Replika—their narrative regarding its identity—was tightly connected with what they reported as experiential consequences of using Replika.

Of those that believed Replika was a friend or a mirror, 12 of 15 experienced learning from Replika. Some who saw Replika as just a friend also learned ($n = 3$). Enhancement or displacement was not associated with learning outcomes, nor was loneliness. Our study indicated that Replika use was associated with enhanced human-human interactions for both the chronically lonely and those experiencing momentary life changes and trauma. Further, there was a strong relationship between those endowing Replika with personhood and those using Replika for therapy, mirroring, and those that experienced learning outcomes. Replika seems to hold a place in users’ minds which is both “other” and “self”—an entity that they can talk to, but which is also an externalization of their inner workings. More research is needed to explore how identity, gender, and learning outcomes interact for users.

Many participants saw Replika as a mirror, calling it an embodiment or extension of themselves. Replika was described as an intelligent reflection of their thoughts and emotions. Our data suggest that people may be able to have exceptionally deep intellectual relationships with ISAs, which lead to self-discovery. In addition to being a cocreated avatar (Meadows 2007), our findings indicate Replika may also become an extension of the user’s mind.

This initial study has a range of implications. Through intensive conversing, cocreation and specific user narratives, ISAs such as Replika may influence “mindset,” a set of beliefs that shape how you make sense of the world and yourself (Dweck 2016), because they offer personal feedback and social engagement practice from a trusted “intelligence” (Boyd and Pennebaker 2017). It remains to be seen in

future research whether Replika presents new possibilities for cognitive (learning) and emotional (therapeutic) support and guidance for users at scale and across broader demographics.

Of all the benefits that ISAs may bring users, we find indications of identity transfer and interaction with an externalized self most intriguing. According to Clark and Chalmers' (1998) extended mind hypothesis, mental states can sometimes be manifested by nonbiological external resources. Their claim that minds sometimes extend beyond our skin out into the broader world, in nonbiological representational systems, is realized in the relationship between users and Replika. Why? Participants are endowing Replika with their personality, functionally training an algorithm on their memories and inputs, and then using it as a "cognitive mirror"—a real-time feedback and review mechanism for seeing their personality and emotions embodied in "someone" else whose peculiarities, strengths, and weaknesses they can experience interactively, rather than as the speaker. The results of this study provide a robust demonstration of Clark and Chalmers' (1998) *extended mind theory*.

We believe this externalized, interactive processing without humans has not previously emerged in research because no conversational systems or agents were sufficiently and simultaneously anthropomorphized, intelligent, and cocreated. Given the increasingly widespread use of ISAs globally, it may be argued that there is a new experiential paradigm emerging—an externalized cognitive space where one's digital mirror becomes a part of everyday conversation, emotion regulation, and personal consciousness.

7 Future Work and Limitations

Consider that Vygotsky's (1986) concept of the "zone of proximal development" is defined as the difference between the learner's autonomous action and what is possible with guidance. This guiding force has heretofore been human, but ISAs appear to bring new possibilities, as these early findings indicate, of guided intellectual, emotional, and psychological learning.

VanLehn (2011) found that tutors were effective because they made learners focus, motivated them, and provided real-time feedback. Therefore, we ask—if ISAs can spur metacognition—might a key aspect of machine-aided learning be shaped by the user's narrative about the intelligence of the agent? With the incorporation of learner affective states into teaching and assessment, learning technology has new potential for creating emotionally supportive learning environments (Harley et al. 2017).

In summary, diverse Replika use motivations encompassed the need for mundane emotional support and deeper intellectual quests. We identified three distinct use patterns among participants, which we call availability, therapy, and mirror. The 27 case study interviews reveal that Replika provided social bonding in mitigating harmful effects of loneliness we earlier reviewed. Yet use went beyond social

bonding to therapy and learning. Participants reported that Replika changed their human relationships, their emotional state, and their cognitive state.

We found indicative combinations in user motivation, ISA narrative, and user-experienced social support led to changes in perceived loneliness and social connectedness. We recognize that our study is limited, composed of a small self-selecting sample, lacking desirable demographic data. Nonetheless, our findings suggest that, as machine intelligence capabilities broaden, and as ISAs with strong anthropomorphic realism are cocreated, it will become increasingly crucial to understand their potential consequences for individual and collective user cognition.

Several communities are likely to benefit from this research. Developers might use this work to understand how to conceptualize agent-driven responses in conversations. Psychologists and communication researchers will benefit since they might advocate for agents-as-interventions without fully understanding their value, which we begin to illuminate in this study.

Conflict of Interest The authors declare that there are no conflicts of interest.

References

- Andersson, G., and Cuijpers, P. (2009). Internet-based and other computerized psychological treatments for adult depression: a meta-analysis. *Cognitive Behaviour Therapy*, 38(4), 196–205.
- Andersson, L. (1998). Loneliness research and interventions: A review of the literature. *Aging and Mental health*, 2(4), 264–274.
- Barak, A., Hen, L., Boniel-Nissim, M., and Shapira, N. A. (2008). A comprehensive review and a meta-analysis of the effectiveness of internet-based psychotherapeutic interventions. *Journal of Technology in Human Services*, 26(2-4), 109–160.
- Barrera, J. (1986). Distinctions between social support concept, measures, and models. *American Journal of Community Psychology*, 14 (1986), 413–445.
- Bickmore, T., Gruber, A., and Picard, R. (2005). Establishing the computer–patient working alliance in automated health behavior change interventions. *Patient Education and Counseling*, 59(1), 21–30.
- Boyd, R.L., and Pennebaker, J.W. (2017). Language-based personality: A new approach to personality in a digital world. *Current Opinion in Behavioral Sciences*, 18, 63–68.
- China Daily (2020). Microsoft expands presence of AI platform Xiaoice. www.chinadaily.com.cn Retrieved 2020-05-10.
- Clark A., and Chalmers, D. (1998). The extended mind. *Analysis*, 58(1), 7–19.
- Cohen, S. and Wills, T.A. (1985). Stress, social support, and the buffering hypothesis *Psychological Bulletin*, 98 (1985), 310–357
- Cohen, S., Mermelstein R., Kamarck T., Hoberman H.M. (1985). Measuring the functional components of social support. In Sarason I.G., Sarason B.R. (Eds.) *Social support: theory, research and applications*. NATO ASI Series (D: Behavioural and Social Sciences), 24. Dordrecht: Springer.
- Creswell, J. W., and Plano Clark, V. L. (2010). *Designing and conducting mixed methods research* (2nd ed.). Thousand Oaks, CA: Sage.
- Dweck (2016). *Mindset: The new psychology of success*. New York: Random House.
- Fitzpatrick, K. K., Darcy, A., and Vierhile, M. (2017). Delivering cognitive behavior therapy to young adults with symptoms of depression and anxiety using a fully automated conversational agent (Woebot): a randomized controlled trial. *JMIR Mental Health*, 4(2).

- Gardner, W. L., Pickett, C., & Knowles, M. L. (2005). "Social snacking" and "social shielding": The use of symbolic social bonds to maintain belonging needs. In K. Williams, J. Forgas, & W. von Hippel (Eds.), *The social outcast: Ostracism, social exclusion, rejection, and bullying* (pp. 227–242). New York: Psychology Press.
- Gilbert, R., and Forney, A. (2015, January). Can avatars pass the Turing test? Intelligent agent perception in a 3D virtual environment. *International Journal of Human-Computer Studies*, 73, 30–36.
- Glaser, B. G. (1965). The constant comparative method of qualitative analysis. *Social problems*, 12(4), 436–445.
- Glaser, B. G., and Strauss, A. L. (2017). *Discovery of grounded theory: Strategies for qualitative research*. New York: Routledge.
- Grand View Research (2021). *Chatbot Market Size Worth \$2,485.7 Million By 2028 | CAGR: 24.9%*.<https://www.grandviewresearch.com/press-release/global-chatbot-market>
- Hancock, J.T., Naaman, M., Levy, K. (2020) AI-Mediated Communication: Definition, research agenda, and ethical considerations, *Journal of Computer-Mediated Communication*, 25 (1), January 2020, 89–100.
- Harley, J. M., Lajoie, S. P., Frasson, C., and Hall, N. C. (2017). Developing emotion-aware, advanced learning technologies: A taxonomy of approaches and features. *International Journal of Artificial Intelligence in Education*, 27(2), 268–297.
- Healy, J. M. (1989). Emotional adaptation to life transitions: Early impact on integrative cognitive processes. In *Personality psychology*. 115–127. New York: Springer.
- Ho, A., Hancock, J., Miner, A.S. (2018, August). Psychological, relational, and emotional effects of self-disclosure after conversations with a chatbot. *Journal of Communication*, 68 (4), 712–733.
- Kanner, A.D., Coyne, J.C., Schaefer, C., and Lazarus, R.S. (1981). Comparison of two modes of stress measurement: daily hassles and uplifts versus major life events. *J Behav Med*. 4(1),1–39.
- Kelly, R. V. (2004). *Massively multiplayer online role playing games: The people, the addiction and the playing*. Jefferson, NC: McFarland and Company.
- Koike, M., and Loughnan, S. (2021). Virtual relationships: Anthropomorphism in the digital age. *Social and Personality Psychology Compass*, 15(6), e12603.
- Kraut, R., Patterson, M., Lundmark V., Kiesler, S., Mukopadhyay, T., and Scherlis,W. (1998). Internet paradox. A social technology that reduces social involvement and psychological well-being? *Am Psychol*. 53(9), 1017–1031.
- Merriam, S. B. (1998). *Qualitative research and case study applications in education*. San Francisco, CA: Jossey-Bass.
- Meadows, M. S. (2007). *I, avatar: The culture and consequences of having a second life*. New Riders.
- Mell, J., Lucas, G., and Gratch, J. (2017). Prestige questions, online agents, and gender-driven differences in disclosure. In *The International Conference on Intelligent Virtual Agents*. 273–282. Cham: Springer.
- Miceli, T., Murray, S. V., & Kennedy, C. (2004). Stimulated reflection: A technique for language learners. Educating: Weaving research into practice. Brisbane: Faculty of Arts, Griffiths University, 21–31.
- Nass, C., Steuer, J., Tauber, E., & Reeder, H. (1993, April). Anthropomorphism, agency, and ethopoeia: computers as social actors. In INTERACT'93 and CHI'93 conference companion on Human factors in computing systems, 111–112.
- Nie, N. (2001). Sociability, interpersonal relations, and the internet: reconciling conflicting findings. *American Behavioral Scientist*, 45(3), 420–435.
- Nowland, R., Necka, E. A., and Cacioppo, J. T. (2018). Loneliness and social internet use: pathways to reconnection in a digital world? *Perspectives on Psychological Science*, 13(1), 70–87.
- Rabbitt, S. M., Kazdin, A. E., and Scassellati, B. (2015). Integrating socially assistive robotics into mental healthcare interventions: Applications and recommendations for expanded use. *Clinical Psychology Review*, 35, 35–46.

- Rook, K. S. (1985). The functions of social bonds: Perspectives from research on social support, loneliness and social isolation. In I. G. Sarason and B. R. Sarason (Eds.), *Social support: Theory, research and applications*, 243-267. Dordrecht: Springer.
- Shaw, L. H., & Gant, L. M. (2002, January 1). In Defense of the Internet: The Relationship between Internet Communication and Depression, Loneliness, Self-Esteem, and Perceived Social Support. *Cyberpsychology and Behavior*, 5(2), 157–172.
- Spek, V., Cuijpers, P. I. M., Nyklíček, I., Riper, H., Keyzer, J., and Pop, V. (2007). Internet-based cognitive behaviour therapy for symptoms of depression and anxiety: a meta-analysis. *Psychological Medicine*, 37(3), 319–328.
- Valkenburg, P. M., and Peter, J. (2007). Preadolescents' and adolescents' online communication and their closeness to friends. *Developmental Psychology*, 43, 267–277.
- VanLehn, K. (2011): The relative effectiveness of human tutoring, intelligent tutoring systems, and other tutoring systems. *Educational Psychologist*, 46:4, 197–221.
- Van Lent, M., Laird, J., Buckman, J., Hartford, J., Houchard, S., Steinkraus, K., Tedrake, R. (1999). Intelligent agents in computer games. *American Association for Artificial Intelligence*.
- Vygotsky, L. S. (1986). *Thought and language* (A. Kozulin, trans.), revised edition. Cambridge, MA: MIT Press.
- Winemiller, D. Mitchell, M.E., Sutliff, J., Cline, D. (1993). Measurement strategies in social support. *Journal of Clinical Psychology*, 49, 638–648

Open Access This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.



An AI-Powered Teacher Assistant for Student Problem Behavior Diagnosis



Penghe Chen and Yu Lu

Contents

1 Introduction	91
2 Theoretical Framework for System Design	92
2.1 Need Deficiency	93
2.2 Problem Behavior	93
2.3 External Environmental Factors	94
2.4 Individual Factors	95
3 System Design	95
3.1 Diagnosis Module	95
3.2 Question Answering Module	97
3.3 Case Search Module	99
4 Discussion and Conclusion	101
References	102

1 Introduction

Teachers in primary and secondary schools usually have to face and handle students' problem behaviors. Student problem behavior has been since decades a research topic with the aim how to help students in their undesirable conducts and actions (Jessor 2016). Students' problems cause concerns in schools and require help and guidance from teachers. Today, for example, Internet addiction and school bullying can be regarded as the typical problem behaviors (Şaşmaz et al. 2014; Dake et al. 2003). Such problem behaviors are obviously harmful to students' own learning and development and to a school community. In practice, many teachers have accumulated rich experience in teaching subjects (e.g., math or biology), but they often lack experience in identifying and diagnosing the student problem behaviors. Some teachers may seek help by reading books, randomly searching

P. Chen · Y. Lu (✉)

Advanced Innovation Center for Future Education, Faculty of Education, Beijing Normal University, Beijing, China

e-mail: chenpenghe@bnu.edu.cn; luyu@bnu.edu.cn

online, or asking peers' experiences. However, such methods may not be quite effective and easily suffer from the subjective and biased experiences. In addition, it requires collecting the student's information from multiple dimensions, where the questionnaire survey, interview, and literature analysis might be used as well. Hence, it is still critical and challenging for teachers to tackle the students' problem behavior issues in real situations.

In this chapter, we present how artificial intelligence (AI) technologies can be employed to help teacher diagnose students' problem behaviors. Specifically, the task-oriented dialogue system technology is utilized to develop an AI-powered assistant for problem behavior diagnosis. The task-oriented dialogue systems have been widely adopted in many other fields, typically including ticket booking (Li et al. 2017), restaurant searching (Wen et al. 2016), and online shopping (Yan et al. 2017). Furthermore, the dialogue system has been used for automatic diagnosis of disease in medical field as well. Through multi-turn dialogue, the system can acquire symptoms from patients and automatically diagnose their diseases, which greatly improves accessibility of medical service (Wei et al. 2018; Peng et al. 2018; Kao et al. 2018).

Inspired by the wide usage of task-oriented dialogue system in other fields, we design and develop a task-oriented dialogue system for automatic identification of students' need deficiencies and targets helping teachers to handle the student problem behaviors. Maslow (1943) states that people's behaviors are driven by their psychological needs, and thus the problem behaviors are often caused by the unfulfilled psychological needs, which are termed as need deficiencies. The students' problem behaviors thus can be handled by identifying their need deficiencies (Harper et al. 2003), timely diagnosing the reasons behind, and conducting necessary interventions. Specifically, the system design is based on a theoretical framework that summarizes the relevant psychology finding for student need deficiency, and utilizes the natural language processing techniques to enable the natural communication between teachers and the system.

The rest of this chapter is organized as follows. Section 2 describes the theoretical framework for the proposed teacher assistant, followed by the system design presented at Sect. 3. Finally, Sect. 4 discusses the impact of proposed AI-powered teacher assistant and concludes this chapter.

2 Theoretical Framework for System Design

Studies have been conducted to analyze the causes underlying students' problem behaviors. According to the classical theory of Maslow (1943), people's behaviors are driven by psychological needs, which implies need deficiencies are the reasons for problem behaviors. Jessor (2014) finds that students' behaviors are influenced by the interactions between students' personality systems and their perceived environment systems. Harper and Stone (2003) shows that the students' psychological needs can be affected by different factors like natural disasters, violence,

Table 1 Classification of student basic needs

	Category
	Physiological needs
	Safety needs
Basic needs	Belongingness and love needs
	Esteem needs
	Cognitive needs

abuse, poverty, lack of school and community resources, and emotional deprivation. Dennis et al. (2005) find that the interaction between individual characteristics and environmental factors influences student development. Those research findings are informative and useful but are too scattered for systematic applications. Hence, a theoretical framework summarizing all the relevant factors is necessary, and the designed system explicitly considers difference classes of need deficiencies, problem behaviors, external environmental factors, as well as individual factors.

2.1 Need Deficiency

According to Maslow's theory (Maslow 1943), student's problem behaviors are driven by the unmet psychological needs. Hence, we define and classify student's need deficiency into five categories: physiological needs, safety needs, belongingness and love needs, esteem needs, and cognitive needs. In our framework, we replace the self-realization need in Maslow's original hierarchy of needs with the cognitive need. The self-realization needs mainly denotes fusing goodness and beauty, which are often demanded in the later stages of life and not appropriate for K-12 students. The list of the classification of student basic needs is summarized in Table 1.

2.2 Problem Behavior

For identifying students' problem behavior, we applied Achenbach and Rescorla's (2014) Child Behavior Checklist (CBCL). It can be used for analysis of children's behavioral and emotional problems between 1.5 and 18 years old. It uses empirical, multiaxis, and cross-assessor measurement methods to identify students' problem behaviors. Specifically, three types of forms were designed: the Teacher Report Form, the Youth Self-Reports, and the Direct Forms. Reliability and validity of these forms has been verified through a series of cross-cultural studies. Our framework categorizes problem behavior with slight modifications learned from real-life case analysis.

In our study, problem behaviors are classified into three categories: externalization problems, internalization problems, and other problems. Externalization

Table 2 Classification of student problem behavior

	Category	Specific factor
Problem behaviors	Externalization problems	Aggressive behavior, rule-breaking behavior
	Internalization problems	Social withdrawal, depression, anxiety
	Other problems	Learning problem, egocentricity, special problem

problems denote the “externalization syndrome” of behaviors, and mainly refer to social adaptation problems, including attack, bullying, sabotage, and so on. It is further divided into aggressive behaviors and rule-breaking behaviors. Internalization problems denote the “internalization syndrome” of behaviors, and refer to emotional distress problems or nonsocial behavioral problems, including anxiety, depression, and so on. It is further divided into social withdrawal, depression, and anxiety. Problems that do not belong to these two categories are defined as “other problems,” which include learning problems, egocentricity, and special problems. The list of the classification of student problem behavior is given in Table 2.

2.3 *External Environmental Factors*

External environmental factors mainly refer to factors that affect students’ growth and therefore significantly affect the formation of problem behavior. Various studies have also been conducted to explore how different factors affect students’ problem behaviors. For example, Hoffmann (2006) finds that changes in parents’ marital status increases the probability of adolescents engaging in problem behaviors. Fomby and Christie (2013) discovers that living in unstable families can lead to more aggressive and antisocial behaviors in these adolescents. Pinquart (2017) shows that students whose parents adopt authoritarian, permissive, and neglectful parenting styles have a high probability of externalizing problems. Maryam et al. (2019) shows that students who are rejected by peer groups tend to develop more internalizing problems.

Based on these findings, we summarized and classified the external environmental factors into three main categories, namely, family factors, school factors, and society factors. A comprehensive and in-depth exploration of the family factors affecting problem behavior can be further divided into the following categories: family structure, parenting style, education background, health condition, delinquent behavior, and socioeconomic status. The school factors are further divided as teacher leadership style, peer acceptance, and peer influence. According to the theory of social learning, the society factors are further divided as social media and cultural customs. The list of the classification of the external environment factors is summarized in Table 3.

Table 3 Classification of external environment factors

	Category	Specific factor
External environment factors	Family factors	Family structure, parenting style, education background, health condition, delinquent behaviors, socioeconomic status
	School factors	Teacher leadership style, peer acceptance, peer influence
	Society factors	Mass media, cultural custom

Table 4 Classification of individual factors

	Category	Specific factor
Individual factors	Demographic information	Grade, gender, health condition, social group
	Personality	Neuroticism, extraversion, openness, agreeableness, conscientiousness

2.4 Individual Factors

Problem behaviors are also influenced by the physical and psychological factors of the individual. Ehrler et al. (1999) find that the personality characteristics of individuals are significantly correlated with a student's problem behaviors, and Van et al. (2013) also show that students with extreme scores of Big Five personality (Five-Factor Model, FFM) are prone to problem behaviors. The five factors include neuroticism, extroversion, openness, agreeableness, and conscientiousness. Hence, we define students' personalities with the Five-Factor Model of Personality (McCrae and Costa 1991). In addition, we consider some basic information and demographic variables related to student problem behavior, including grade, gender, health condition, and social group. The list of the individual factors is given in Table 4. Note that in practice, not all of the factors are required to collect from the students.

3 System Design

Our dialogue support system consists of three main modules, namely, diagnosis module, question answering module, and case search module. We will elaborate them in this section, respectively.

3.1 Diagnosis Module

This module adopts the technology of task-oriented dialogue system to conduct diagnosis. The task-oriented dialogue system is designed to complete a specific task through natural language interaction with users (Gao et al. 2019). Various dialogue

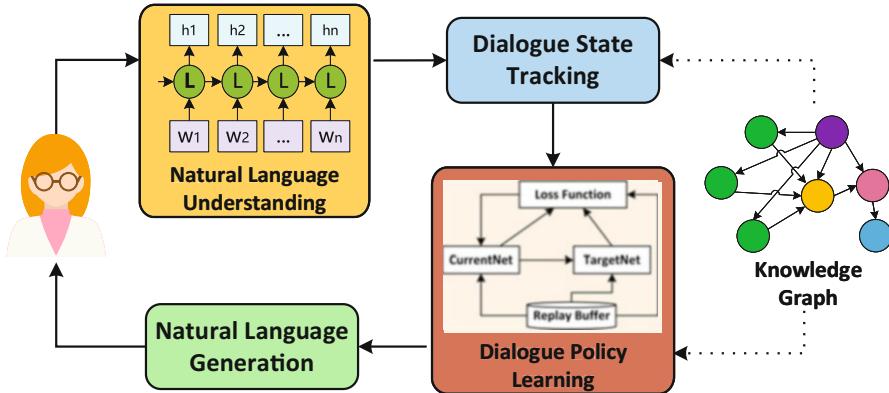


Fig. 1 Diagnosis module for analyzing student problem behavior

systems have been designed for different tasks in the literature. Some systems are designed for booking tasks. For example, Li et al. (2017) developed a dialogue system for movie-ticket booking. Wen et al. (2016) built a dialogue system to help users search and reserve restaurants. Dialogue systems can also solve information-searching tasks. For instance, Papangelis et al. (2018) designed a spoken dialogue system to help users make informed decisions through information navigation. Another group of tasks is the automatic diagnosis of medical disease. Tang et al. (2016) designed a group of anatomical models emulating different experts in hospitals to diagnose diseases. We have also done some preliminary studies on employing dialogue system to analyze the causes underlying students' problem behaviors (Chen et al. 2020; Chen et al. 2021). Through those dialogue systems, service accessibility can be significantly improved.

To conduct diagnosis, this module acquires the necessary information of a specific student through multi-turn dialogue with the teacher, and then automatically diagnoses the student's need deficiencies behind his or her problem behaviors. The diagnosis process considers both the external environmental factors and individual factors. As shown in Fig. 1, it consists of four main functional components: natural language understanding, dialogue state tracking, dialogue policy learning, and natural language generation.

The natural language understanding component interprets the teacher's utterance to extract the intent as well as task-related semantic information. Specifically, it processes a teacher's reply to extract the student's information, such as whether he has aggressive behaviors. In this teacher's assistant, the long short-term memory (LSTM) (Hochreiter and Schmidhuber 1997) network is adopted to interpret the teacher's utterances. An LSTM network is a typical recurrent neural network that has been widely used in natural language processing recently. Relying on a gating mechanism, it can solve the long-term dependency issue in the sequential data processing.

The dialogue state tracking component tracks the dialogue state that represents all of the task-related information captured. This dialogue state represents students' information acquired to that point and is utilized to determine the next system action. Specifically, this module updates the dialogue state with another LSTM network based on the output of natural language understanding component.

The dialogue policy learning module takes charge of making decisions on the next system action based on the current dialogue state, such as requesting information or informing certain results. Based on the current dialogue state, we adopt a reinforcement learning model, specifically a deep Q-learning network (DQN) model (Mnih et al. 2015), to learn the dialogue policy that decides whether to request more information from the teacher or to present the derived need deficiency to the teacher. As one of the three main paradigms of machine learning, reinforcement learning targets solving sequential decision-making problems. Recently, deep learning techniques have been integrated into reinforcement learning models to improve model performance. The DQN is a typical deep reinforcement learning model that utilizes a deep neural network to calculate the Q-value in the model. Finally, the natural language generation component utilizes a template-based model to transform system action into text response.

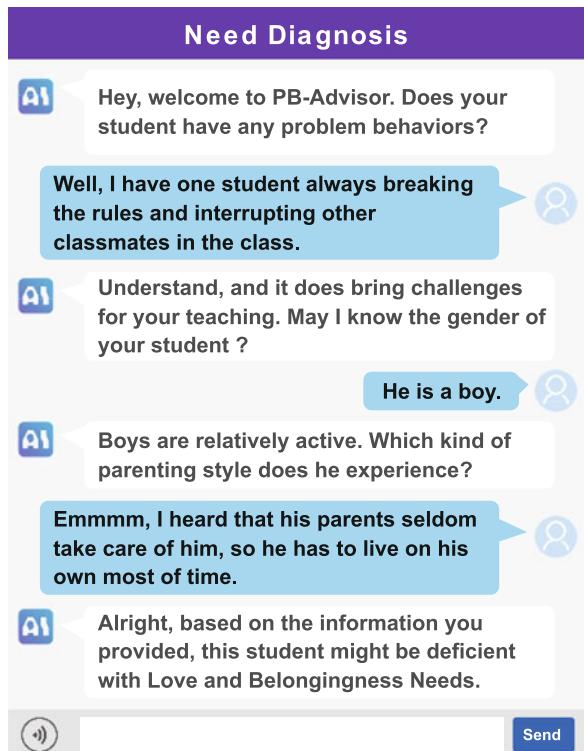
Figure 2 demonstrates a toy example of how the module acquires the student information through a multi-turn dialogue and diagnoses the need deficiency. In short, through multi-dialogue interaction, the module can effectively acquire the students' information, automatically analyze their need deficiencies, and adaptively generate the advice for teachers.

3.2 *Question Answering Module*

Unlike the diagnosis module that targets on analyzing the problem behaviors for the specific student, this module aims to provide general guidelines on typical problem behaviors through answering questions like "What are the typical problem behaviors for high school girls?" The community question answering (CQA) technology is employed to answer such questions. CQA is a web-based service to help people seek information by answering their questions based on knowledge shared by others in the community (Srba and Bielikova 2016). Quora and Stack Overflow are two typical examples of CQA systems. The main idea of CQA is to utilize knowledge shared by the domain experts in the community discussion, and it is usually built based on data collected from the professional online forums and platforms. Our CQA system is built with the historical questions and answers collected from a nationwide online platform in China (<http://haolaoshi.bnu.edu.cn/>).

CQA system aims to pick out the most appropriate answer from multiple answers of the given question, and typically includes two main tasks: finding the similar questions and finding the relevant answers (Joty et al. 2018). Traditional approach focuses on the syntactic analysis on the text of questions and answers. For example, Cui et al. (2005) proposed a general tree-based method calculating tree-edit distance

Fig. 2 A toy example of how diagnosis module works



to match question and answer. Recently, with the development of deep learning, various deep neural network models have been proposed. For example, Zhou et al. (2018) proposes a recurrent convolutional neural network (RCNN) to capture both the semantic matching between question and answer and the semantic correlations embedded in the sequence of answers. Hence, we are inspired to develop our CQA model with deep learning algorithms.

The structure of the designed CQA model is illustrated in Fig. 3. Specifically, the model provides a two-phase processing. The first one is the question selection phase aiming to find the candidate questions similar to the incoming question. The second one is the answer selection phase which ranks all the answers of the candidate questions generated by phase I, and then selects the most appropriate answer as output.

The first phase identifies the candidate questions similar to the incoming question from the existing ones. We used the pretrained BERT (Devlin et al. 2018) model for natural language processing to analyze the semantics of questions and answers. It first learns the semantic vectors of the existing questions, and creates a database for all the question semantic vectors. Whenever a new incoming question arrives, the same BERT framework is adopted to learn its semantic vector. Subsequently, the model is fine-tuned by a multilayer perceptron (MLP) network to compute the

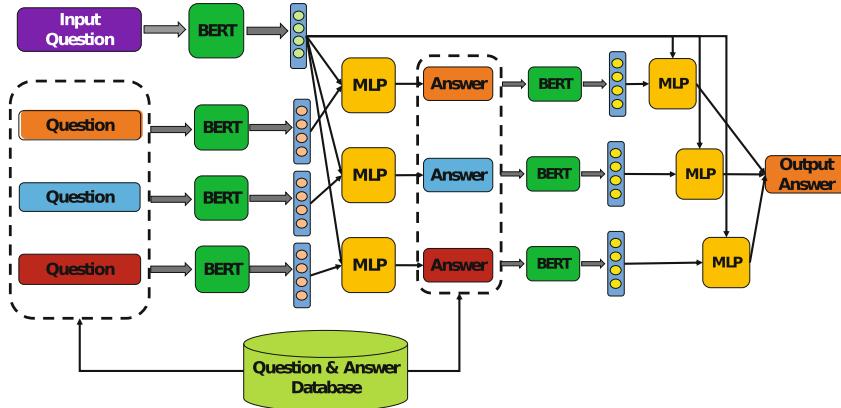


Fig. 3 The CQA model used in question answering module

similarity between incoming question and each existing question. Accordingly, it computes a similarity value for each existing question. With a predefined similarity threshold value, a set of similar questions are selected as candidates.

The second phase then starts to identify the most appropriate answer. Firstly, a set of candidate answers is generated based on the best answer of each candidate question in the first phase. Secondly, the semantic vector of each candidate answer is learned using the BERT framework like the first phase. Thirdly, by concatenating the question vector and answer vector, an MLP network is employed to fine-tune the model to compute the matching level between a question and an answer. Finally, the candidate questions are ranked according to the multiplication of question similarity and answer matching level, and the one with the biggest calculated value is chosen as the final output.

3.3 Case Search Module

This module is an independent service that helps teachers to search the similar cases containing successful experiences in diagnosing and intervening student's problem behaviors. Searching is mainly based on teachers' text description on student problem behaviors, and the similarity refers to the various aspects of problem behaviors between cases and the teacher's description. Compared to the simple answers given by the question answering module, the returned cases contain more details, not only including student's specific behaviors, but also including other relevant information like personal particulars and family background information. More importantly, the cases also contain experts' analysis on the student's behavior and the reason behind it, as well as providing different educational strategies and

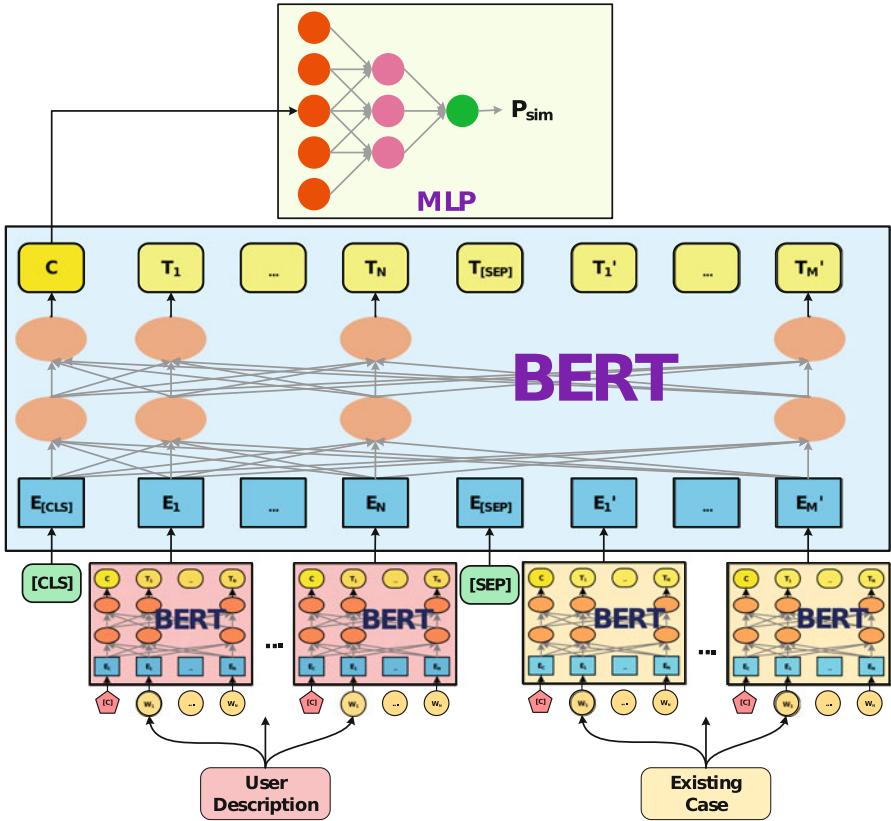


Fig. 4 The hierarchical BERT model used in case search module

interventions applied. All these details can supply the fine-grained guidelines and advice for teachers to handle similar problem behaviors.

This module is developed with the technology of information retrieval. As a typical natural language processing task, information retrieval aims to find the closely related information according to user requirements. It explores how to represent, store, organize, and access information properly for information searching (Chowdhury 2010).

Various models have been proposed to conduct information retrieval. This module utilizes a deep natural language processing model to compute the similarity between teacher's text description and case documents. Unlike the semantic similarity calculation in question answering module targeting on computing similarity between two sentences, this case engine computes the similarity between two different documents in the form of a sequence of sentences. As illustrated in Fig. 4, a hierarchical BERT model is designed and implemented to compute the semantic similarity between teacher's text description and each case document.

In this mode, the bottom layer mainly learns the semantic vector of each sentence in teachers' text description and case documents. Specifically, parameters of pretrained BERT model are adopted directly for this bottom layer BERT. The top layer targets on learning the semantic similarity between teacher's text description and each case document. By taking the semantic vectors of sentences generated with bottom BERT layer as input, we add in the special token "[CLS]" at the beginning and "[SEP]" in the middle to concatenate the two sequence into one sequence. Subsequently, the model can process it like a normal sequence, and generate a semantic similarity vector at the beginning position. After generating the semantic similarity vector, one MLP network model is employed to compute the similarity between the teacher text description and the case document. Similar to the question answering module, all cases are ranked according to the computed semantic similarity and then return back to the teacher.

4 Discussion and Conclusion

The main idea of current AI algorithms is the combination of the data-driven paradigm with the knowledge-driven paradigms. The development of the AI-powered teacher assistant can be regarded as an attempt of utilizing such both paradigms to solve the practical problem in education. Based on the knowledge-driven paradigm, the principles and theories in psychological studies are employed to build the theoretical framework, which guides the machines to solve the targeted student behavior problem in a theoretical manner. By leveraging on the data-driven paradigm, the rich and precious teacher experiences embedded in the text data can be extracted and utilized. The integration of these two paradigms provides the solution, and it aims to ensure the reliability and validity of the developed teacher assistant for student problem behaviors. Specifically, the system can analyze students' need deficiencies behind their problem behaviors and identify the corresponding external environmental and individual factors that result the deficiencies. It also helps teachers find answers or similar resolved cases in many typical student problem behaviors. By taking these answers and cases as references, the teachers can learn how to help their students. The system interacts with teachers through natural language, which greatly improves the usability as well.

On the other hand, we also note that it may cause certain concerns when such an intelligent agent is deployed in schools. People may worry whether it is ethical to utilize machines to analyze and even regulate the students. In practice, the developed assistant is used as a supporting tool offering advice and suggestions to teachers, rather than applying educational intervention directly to students. Another possible concern relates to the data privacy risk that students' information will be leaked and abused. The developed assistant is designed with privacy protection inherently that it does not store any sensitive data of students after its usage. In addition, it is possible that the current version of the teacher assistant may misinterpret teacher's descriptions, which results in wrong diagnosis and inappropriate advice.

We plan to employ the explainable AI (xAI) techniques to show teachers how the developed assistant makes the current advice and how confident the assistant is on the given advice. The teachers then could make their own decisions on whether they would adopt the advice or not. Driven by the advancements of AI, especially the natural language processing and machine learning techniques, we believe the teacher assistant could eventually tackle such issues and eventually benefit both teachers and students in schools.

References

- Achenbach, T. M., & Rescorla, L. A. (2014). The Achenbach system of empirically based assessment (ASEBA) for ages 1.5 to 18 years. In *The use of psychological testing for treatment planning and outcomes assessment*, 179-214, Routledge.
- Chen, P., Lu, Y., Peng, Y., Liu, J., & Xu, Q. (2020). Identification of Students' Need Deficiency Through a Dialogue System. *International Conference on Artificial Intelligence in Education*, 59-63. https://doi.org/10.1007/978-3-030-52240-7_11
- Chen, P., Lu, Y., Yu, S., Xu, Q., Liu, J. (2021). A dialogue system for identifying need deficiencies in moral education. *Journal of Pacific Rim Psychology*, 15. <https://doi.org/10.1177/1834490921998589>
- Chowdhury, G. G. (2010). *Introduction to modern information retrieval*. Facet publishing.
- Cui, H., Sun, R., Li, K., Kan, M. Y., Chua, T. S. (2005). Question answering passage retrieval using dependency relations. *The 28th ACM SIGIR Conference on Research and Development in Information Retrieval*, 400-407. <https://doi.org/10.1145/1076034.1076103>
- Dake, J. A., Price, J. H., & Telljohann, S. K. (2003). The nature and extent of bullying at school. *Journal of School Health*, 73, 173-80. <https://doi.org/10.1111/j.1746-1561.2003.tb03599.x>
- Devlin, J., Chang, M. W., Lee, K., Toutanova, K. (2018). Bert: Pre-training of deep bidirectional transformers for language understanding. arXiv preprint arXiv:1810.04805.
- Dennis, J. M., Phinney, J. S., & Chuateco, L. I. (2005). The role of motivation, parental support, and peer support in the academic success of ethnic minority first-generation college students. *Journal of College Student Development*, 46(3), 223-236. <https://doi.org/10.1353/csd.2005.0023>
- Ehrler, D. J., Evans, J. G., & McGhee, R. L. (1999). Extending Big-Five theory into childhood: A preliminary investigation into the relationship between Big-Five personality traits and behavior problems in children. *Psychology in the Schools*, 36(6), 451-458. [https://doi.org/10.1002/\(SICI\)1520-6807\(199911\)36:6<451::AID-PITS1>3.0.CO;2-E](https://doi.org/10.1002/(SICI)1520-6807(199911)36:6<451::AID-PITS1>3.0.CO;2-E)
- Fomby, P., & Christie, A. S. (2013). Family structure instability and mobility: The consequences for adolescents' problem behavior. *Social Science Research*, 42(1), 186-201. <https://doi.org/10.1016/j.ssresearch.2012.08.016>
- Gao, J., Galley, M., & Li, L. (2019). *Neural Approaches to Conversational AI: Question Answering, Task-oriented Dialogues and Social Chatbots*. Now Foundations and Trends. <https://doi.org/10.1561/1500000074>
- Harper, F. D., Harper, J. A., & Stills, A. B. (2003). Counseling children in crisis based on Maslow's hierarchy of basic needs. *International Journal for the Advancement of Counselling*, 25(1), 11-25. <https://doi.org/10.1023/A:1024972027124>
- Harper, F. D., & Stone, W. O. (2003). Transcendent counseling: An existential, cognitive-behavioral theory. *Culture and counseling: New approaches*, 233-251.
- Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural computation*, 9(8), 1735-1780. <https://doi.org/10.1162/neco.1997.9.8.1735>

- Hoffmann, J. P. (2006). Family structure, community context, and adolescent problem behaviors. *Journal of Youth and Adolescence*, 35(6), 867-880. <https://doi.org/10.1007/s10964-006-9078-x>
- Jessor, R. (2014). Problem Behavior Theory: A half century of research on adolescent behavior and development. In R. M. Lerner, A. C. Petersen, R. K. Silbereisen, & J. Brooks-Gunn (Eds.), *The Developmental Science of Adolescence: History Through Autobiography*, 239-256. Psychology Press.
- Jessor, R. (2016). The origins and development of problem behavior theory. *Advancing Responsible Adolescent Development*. <https://doi.org/10.1007/978-3-319-40886-6>.
- Joty, S., Lluís Márquez, & Nakov, P. (2018). Joint Multitask Learning for Community Question Answering Using Task-Specific Embeddings. *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*. arXiv preprint arXiv:1809.08928.
- Kao, H., Tang, K., & Chang, E. Y. (2018). Context-aware symptom checking for disease diagnosis using hierarchical reinforcement learning. *Proceedings of the AAAI Conference on Artificial Intelligence*, 32(1). <https://ojs.aaai.org/index.php/AAAI/article/view/11902>
- Li, X., Chen, Y., Li, L., Gao, J., Celikyilmaz, A. (2017). End-to-end task-completion neural dialogue systems. arXiv preprint arXiv:1703.01008.
- Maryam, Z., Ali, M., Kaisa, A., Jari-Erik, N., Noona, K., Marja-Kristiina, L. (2019). Supportive Parenting Buffers the Effects of Low Peer Acceptance on Children's Internalizing Problem Behaviors. *Child & Youth Care Forum*, 48(6), 865-887. <https://doi.org/10.1007/s10566-019-09510-y>
- Maslow, A. H. (1943). A theory of human motivation. *Psychological Review*, 50(4), 370. <https://doi.org/10.1037/h0054346>
- McCrae, R. R., & Costa, P. T. (1991). Adding Liebe und Arbeit: The Full Five-Factor Model and Well-Being. *Personality and Social Psychology Bulletin*, 17(2), 227-232. <https://doi.org/10.1177/014616729101700217>
- Mnih, V., Kavukcuoglu, K., Silver, D., Rusu, A. A., Veness, J., Bellemare, M. G., et al. (2015). Human-level control through deep reinforcement learning. *Nature*, 518(7540), 529-533. <https://doi.org/10.1038/nature14236>
- Papangelis, A., Papadakos, P., Stylianou, Y., Tzitzikas, Y. (2018). Spoken dialogue for information navigation. *The 19th Annual SIGdial Meeting on Discourse and Dialogue*, 229-234. <https://doi.org/10.18653/v1/W18-5025>
- Peng, Y., Tang, K., Lin, H., Chang, E. (2018). Refuel: Exploring sparse features in deep reinforcement learning for fast disease diagnosis. *Advances in Neural Information Processing Systems*, 31, 7322-7331.
- Pinquart, Martin. (2017). Associations of Parenting Dimensions and Styles with Internalizing Symptoms in Children and Adolescents: A Meta-Analysis. *Marriage & Family Review*, 1492016-1494929. <https://doi.org/10.1037/dev0000295>
- Şaşmaz, T., Öner, S., Kurt, A. Ö., Yapıcı, G., Yazıcı, A. E., Buğdaycı, R., Şiş, M. (2014). Prevalence and risk factors of internet addiction in high school students. *European Journal of Public Health* (1), 15-20. <https://doi.org/10.1093/eurpub/ckt051>
- Srba, I., & Bielikova, M. (2016). A comprehensive survey and classification of approaches for community question answering. *ACM Transactions on the Web (TWEB)*, 10(3), 1-63. <https://doi.org/10.1145/2934687>
- Tang, K., Kao, H., Chou, C., Chang, E. Y. (2016). Inquire and diagnose: Neural symptom checking ensemble using deep reinforcement learning. *Paper presented at the In Proceedings of NIPS Workshop on Deep Reinforcement Learning*.
- Van, D. A. A. L., Prinzie, P., Dekovi, M., De Haan, A. D., Asscher, J. J., Widiger, T. (2013). The development of personality extremity from childhood to adolescence: Relations to internalizing and externalizing problems. *Journal of Personality & Social Psychology*, 105(6), 1038-1048. <https://doi.org/10.1037/a0034441>
- Wei, Z., Liu, Q., Peng, B., Tou, H., Chen, T., Huang, X., et al. (2018). Task-oriented dialogue system for automatic diagnosis. *The 56th Annual Meeting of the Association for Computational Linguistics*, 2, 201-207. 10.18653/v1/P18-2033

- Wen, T., Vandyke, D., Mrksic, N., Gasic, M., Rojas-Barahona, L. M., Su, P., et al. (2016). A network-based end-to-end trainable task-oriented dialogue system. arXiv preprint arXiv:1604.04562.
- Yan, Z., Duan, N., Chen, P., Zhou, M., Zhou, J., Li, Z. (2017). Building task-oriented dialogue systems for online shopping. In *Proceedings of the AAAI Conference on Artificial Intelligence*, 31, 4618-4625.
- Zhou, X., Hu, B., Chen, Q., Wang, X. (2018). Recurrent convolutional neural network for answer selection in community question answering. *Neurocomputing*, 274, 8-18. <https://doi.org/10.1016/j.neucom.2016.07.082>

Open Access This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.



Analysis and Improvement of Classroom Teaching Based on Artificial Intelligence



Zhong Sun, Zi Chun Yu, and Fei Yun Xu

Contents

1 Introduction	105
2 Literature Review	106
2.1 Classroom Teaching Analysis.....	106
2.2 Improvement of Classroom Teaching	108
3 Methodology	111
4 Conclusion	118
References	119

1 Introduction

The classroom is the core environment for teaching and learning and provides a complex, multielement interwoven real situation. Classroom teaching plays an important role for achieving high-quality education. Thus, many scholars have put efforts into classroom teaching analysis and efforts for improvements employing quantitative and qualitative methods since the last century (Jacobs et al. 1999). For example, the quantitative analysis such as Student-Teacher analysis method (Cheng et al. 2018) and Flanders Interaction Analysis System (FIAS; Flanders 1963) are based on time-coding analysis. The qualitative analysis is mainly reflected in the analysis of teaching activities and the content of courses (Hatun Ataş and Delialioğlu 2018).

However, common classroom teaching analysis, which is based on coding and counting behaviors and discourse interactions between teacher and students, has been criticized as content-free and low efficiency. With the rapid development of the Artificial Intelligence (AI) technology, applications of AI provide significant new methods to the field of teaching analysis. The AI technologies integrated into learning environment promise totally new tools for classroom teaching analysis.

Z. Sun (✉) · Z. C. Yu · F. Y. Xu
Capital Normal University, Beijing, China
e-mail: sunzhong@cnu.edu.cn

Specific, new capabilities to computing, including sensing, recognizing patterns, representing knowledge, making and acting on plans, and supporting naturalistic interactions with people (Roschelle et al. 2020) have become potential research methods for analysis on interactions between teacher and students.

Therefore, the aim of the chapter is to analyze effective framework and key technologies to conduct classroom teaching analysis and improvement based on the AI. Two research questions are raised as follows:

1. What could be the effective and comprehensive analysis perspectives for classroom teaching to overcome the shortcoming of common research methods that are often time coding and require activity coding?
2. How to use AI technologies to empower classroom teaching analysis and improvement?

2 Literature Review

2.1 Classroom Teaching Analysis

2.1.1 Time Coding

Since the 1970s, time coding on observing in a live classroom or using video tape recording has been applied in the area of classroom teaching quantitative analysis. Researchers cataloged and counted various kinds of behaviors, interactions, or verbal communications between teachers and students during the whole lesson time every 3 s or 15 s, then calculated the total numbers or frequency of each code to draw a conclusion about the teaching styles or qualities.

For instance, The Flanders Interaction Analysis System (FIAS) and Student-Teacher (S-T) analysis have been applied for verbal and behavior analysis, respectively, since last century. Flanders Interaction Analysis Categories (FIAC) provided a Ten Category System of coding classroom communication. Seven categories for teacher talk, two for pupil talk, and the tenth category for silence or confusion (Flanders 1963). S-T analysis is a quantitative analysis method which simplifies behaviors into two types as teacher behaviors (T) and student behaviors (S) in the lesson time. To improve the efficiency of classroom observation and the accuracy of data, behavioral data is collected every 30 s. Finally, the classroom teaching model was analyzed according to the frequency of behavior conversion and teacher behavior occupancy, which provided a basis for teaching evaluation and theoretical research (Gui et al. 2020).

Although the theory and practice of time-coding classroom interaction segments is a century old, many scholars still use the method for classroom interaction analysis (Amatari 2015), even promoting the original FIAS coding system into the Information Technology-Based Interaction Analysis System (ITIAS) to keep up with the times (Gu and Wang 2004). Some scholars conduct S-T method on videos

from several Massive Open Online Courses (MOOC) to detect different teaching styles Sun and Ma (2012).

In general, the time-coding methods shed a light on the quantitative classroom teaching analysis by making behavior or discourse codable and countable. However, time-coding method had to face the inevitable shortcomings like content-free, hard to explain the authentic teaching meaning, and failing to provide valuable feedback for teachers to reflect on and adjust their classroom teaching design and implementation.

2.1.2 Activity Coding

The classroom is a teaching and learning system composed of two dimensions as time and space. The dimension of space could be presented with the activities of teaching and learning. Therefore, some researchers took space into consideration to analyze classroom interactions by applying sampling activities or activity-coding method.

For instance, Rowntree (1990) cataloged learning activities in the classroom into five types: reporting observations or experiences, retelling facts or principles, distinguishing different concepts and principles from examples, enumerating examples, applying new concepts and principles. Mishra and Gaba (2001) suggested analyzing learning activities from two dimensions as questions and reflective actions. Horton (2012) proposed that learning activities should be grouped into absorption activities, doing activities and associative activities. Mu and Zhu (2015) constructed the Teaching Behavior Analysis System with three types of information-based classroom activities including teaching activities, learning activities, and meaningless activities.

Although activity coding had taken content and authentic teaching meaning into consideration which overcame some disadvantages of time coding in some extent, it still failed to answer the problems. Firstly, did all activities deserve to be analyzed if some failed to support the learners' cognitive processes of learning? Secondly, could all kinds of activities possibly be cataloged and analyzed with common agreements on classified rules? If time and activity are not appropriate coding dimensions for classroom analysis, then what should be?

2.1.3 Event Coding

Events of instruction might be the potential answer. Originally proposed by R. Gagné, who is best known for the theories of learning outcomes, learning conditions, and nine events of instruction, the events refer to a series of external stimulus to promote learning in the learner's cognitive processing (Gagné 1970) (Table 1).

Based on the nine events of instruction, scholars and practitioners refined and applied the theories into practice from multiple school levels and subjects like website design (Zhu and Amant 2010), medical teaching (Goode 2018), physics

Table 1 Nine events of instruction (Gagné 1970)

Instructional event	Internal mental process
1. Gaining attention	Stimuli activates receptors
2. Informing learners of the objective	Creates level of expectation for learning
3. Stimulating recall of prior learning	Retrieval and activation of short-term memory
4. Presenting the stimulus	Selective perception of content
5. Providing learning guidance	Semantic encoding for storage long-term memory
6. Eliciting performance	Responds to questions to enhance encoding and verification
7. Providing feedback	Reinforcement and assessment of correct performance
8. Assessing performance	Retrieval and reinforcement of content as final evaluation
9. Enhancing retention and transfer	Retrieval and generalization of learned skill to new situation

teaching in junior high school (Huang 2015), information technology in university (Jing 2012), and graphic design in secondary vocational school (Zhang 2019).

Compared with time and activity-coding methods, event coding provides several advantages for classroom analysis. First, events play a vital role in stimulating learners' cognitive processing. Not all activities could be regarded as events of instructions, but events are all valid activities for learning. Second, the kinds of events are limited in number, with clear rules of classification. Therefore, this study identified event coding as the appropriate dimension for classroom analysis.

2.2 Improvement of Classroom Teaching

2.2.1 Purpose of Teaching Improvement

Classroom improvement is a continuous cycle of constantly discovering and improving problems in real teaching situations. Mehan (1979) proposed a tripartite model of interaction (initiation–response–feedback), which intends to emphasize that feedback is an important tool for promoting classroom interaction and improving classroom teaching through effective feedback. Therefore, the development of teachers and the improvement of teaching quality cannot be separated from teaching improvement.

In early studies, some scholars attempted to improve the classroom quality from different perspectives. For instance, Seldin (2010) used students' feedback to judge teachers' behavior with suggestions as improving the quality of education through group teaching diagnosis. Ellis (1990) analyzed teaching behaviors and evaluated teacher performance through the indicators of recommended teaching behaviors. According to the analysis results, the recommended teaching behaviors include giving students feedback, talking about students' thinking, suggesting extended

activities, and calling attention to the competencies of low-status students. Stanulis et al. (2012) considered classroom discussion as a point of improvement and took classroom discussions as a high-leverage practice to effective teaching.

To sum up, the aforementioned researches decomposed the analysis elements of classroom teaching into various dimensions such as teachers and students' behavior, teaching activities, students' feedback, and so on. Although these elements are vital and necessary to the classroom, lack of inclusive and systematic destination made practitioners confused about the analysis results. What is behind the behaviors? What is deep reason for behaviors or discourse analysis? The classroom teaching is a compound structural system. As Bryk et al. (2011) noted that rather than thinking about the proven effectiveness of a tool, routine, or some other instructional resource, improvement research directs efforts toward understanding how such methods can be adaptively integrated with efficacy into varied contexts. Therefore, we need found an inclusive and systematic perspective for classroom analysis and improvement. What it should be? Teaching structure.

No matter what kinds of educational settings like formal or informal, Western or Eastern, old times or nowadays, there are always four important components in a teaching and learning environment as teacher, student, learning contents, and media. The dynamic and systemic relationships among the four components in various teaching and learning contexts are named as teaching structure. Chinese scholar He (2002) defined the teaching structure as clear, stable, and on purpose teaching practice plan which embodied different pedagogies. He summarized out three teaching structures as the teacher-centered, student-centered, and teacher-guided-student-centered structures. According to He (2002), each teaching structure has reasonable application to achieve specific learning goals, but the teacher-guided-and-student-centered structure plays the most important role for students' growth in the classroom or school setting. Therefore, revealing the relationships of the four components and detecting the teaching structure of classroom became the fundamental and inclusive destination for teaching analysis and improvement.

2.2.2 Methods of Teaching Improvement

As the next step of classroom analysis, improving the quality of classroom teaching has been explored continuously in the recent decades. Some of these methods are introduced in this section.

Lesson study is a professional development method that originated from Japan, and centers on the collaborative study of live classroom observation, analysis, and improvement have spread rapidly since 1999 (Lewis et al. 2006). For instance, math teachers from the USA have applied this intervention pattern for carrying out case studies, including four lesson study features (i.e., investigation, planning, research lesson, and reflection) and three pathways through which lesson study improves instruction (i.e., changes in teachers' knowledge and beliefs, professional community, and teaching–learning resources) (Lewis et al. 2009). A framework for conducting lesson study in a teacher development project in Austria established a

checklist for research lesson planning to frame teacher and student learning. The framework established the criteria for evaluating teacher behavior and learning and their effects on student learning (Mewald and Mürwald-Scheifinger 2019).

Action research is another research tool for improving classroom teaching. Research indicates that a carefully designed action research project can effectively capture the attention of faculty and administrators and achieve teaching improvement objectives (Cook et al. 2007).

Since the beginning of the twenty-first century, the vigorous development of information technology has brought technological innovation into classroom improvement methods. The time and activity-coding limitations in classroom analysis have been addressed to some extent toward a new level by the integration of cutting-edge technologies. For example, the Classroom Assessment Scoring System observed 180 early childhood classrooms and pointed out problems that should be improved in teaching (Hu et al. 2016a). Digital Interactive Video Exploration and Reflection (Pea and Lindgren 2008) applied the look–notice–comment strategy and a specific software to support the analysis and improvement of teaching after analyzing teaching practice videos (Derry et al. 2010). The Learning Cell platform supports site classroom observation with a mobile application and records before, during, and after class teaching behaviors. After class, teachers in a group engage in a collaborative improvement discussion based on the analysis results (Chen et al. 2018). The Learning Instruction Curriculum and Culture (LICC) model is a classroom observation and evaluation theory framework and uses a series of evaluation tools (Cui 2012). The LICC has 4 dimensions and 68 observation points for classroom teaching. After on-site observation, teachers who use LICC tools record and identify specific problems of the current lesson. Then, the teachers show the analysis results and provide feedback for improvement. Measuring Effective Teaching (MET) was initiated by the Bill and Melinda Gates Foundation (2013) to improve the quality of teaching. MET describes three approaches to measuring different aspects of teaching, namely, student surveys, video recorded classroom observations, and student achievement gains on state tests. The findings suggest that the existing measures of teacher effectiveness provide important and useful information on the causal effects that teachers have on students' outcomes. However, problems in both non-tech- and technology-based improvement methods remain.

First, the evidence of the connection between analysis results and improvement solutions is insufficient. Regardless of the communication after classroom observation (oral or written), most of the feedback about improvement is based on personal teaching experience.

Second, some quantitative research methods focus on a single element, such as behaviors and discourse. However, class is a complex setting containing multimodal data. Evidence from different resources should be considered.

Third, descriptive statistics of the analysis data fail to need effective improvement. In addition to the frequency statistics and percentage calculation of the behaviors or discourse in the classroom, the teaching structure and the specific strategies embody an important educational meaning.

In summary, on the basis of the current research achievements in theory and practice, new methods and technologies should be explored to take classroom teaching analysis and improvement to the next level.

3 Methodology

In this chapter, an AI-supported classroom teaching analysis framework is proposed named as TESTII (Fig. 1). The current TESTII framework is based on the nine major teaching events of Gagné, and the analysis is carried out in the cognitive way of teachers' teaching. TESTII includes the following analysis phases and key techniques.

Step 1: Identifying Teaching Events

As mentioned above, teaching events approach overcomes the time and activity-coding limitations with the advantages of improving the efficiency of classroom teaching analysis and effectively establishing connections between the quantitative structure and the meaning understanding. Therefore, identifying different teaching events is the first step of TESTII analysis.

Teaching events can be extracted and identified from the lesson plan and classroom teaching videos of each teaching case. Lesson plans are mainly composed

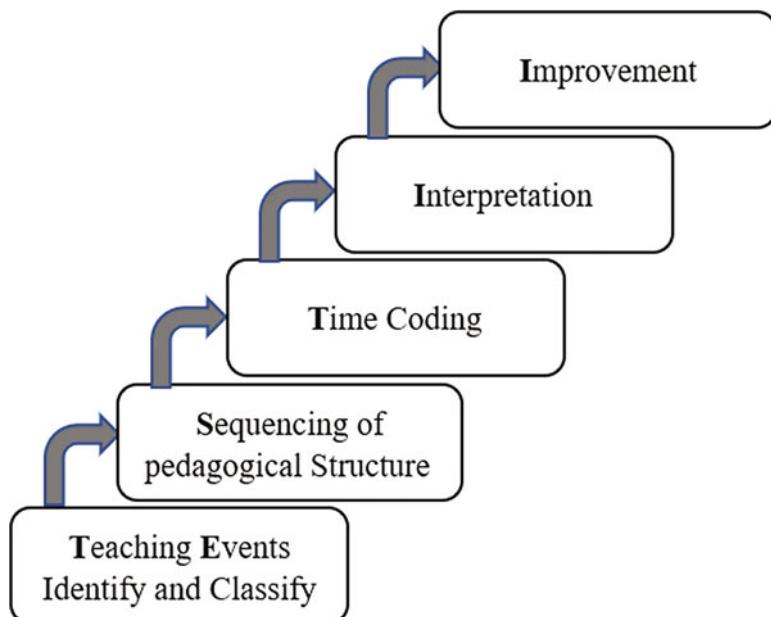


Fig. 1 TESTII framework: AI-supported classroom teaching analysis

of texts. Therefore, the use of natural language processing (NLP) and computer vision (CV) technologies to analyze texts and videos and identify teaching events has become the key approach in this stage. Compared with the common method of relying on manual classroom observation, the use of CV/NLP technology has significant advantages in time and resource savings but fails to recognize the deep meaning of the word, accurately locate the changing expressions of the same type of activities or events, and find the meaningful sequence in the teaching structure. Therefore, the human–machine cooperation mode is adopted for the recognition of teaching events, and the specific analysis steps are as follows.

The first stage involves the collection of videos of each lesson and the random selection of static images. The researchers classify part of the scene data, and these labelled data are used as the training set to train the neural network model of scene classification. Then, computer vision technology is applied to detect the key scenes and cut the video into pieces for possible teaching events recognition (Fig. 2).



Fig. 2 Detecting the key scenes of classroom teaching by computer vision

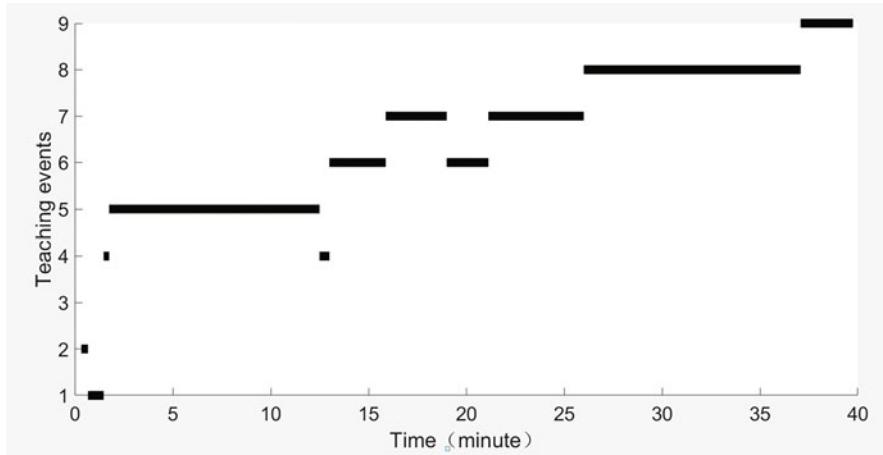


Fig. 3 Sample diagram of time distribution of teaching events

Second, NLP technology is applied to select teaching events using key words of every event. The researchers divide the teaching events into labels and mark texts to form corresponding judgment rules. Then, deep learning model Word2vec is used to generate an event classifier on the basis of the gate recurrent unit (GRU) to judge the accuracy of the model. Furthermore, the specific teaching event was recognized through NLP technology, and the time distribution map of teaching events for a lesson could be generated visually as shown in Fig. 3.

After using the aforementioned method to identify the teaching events, the study found that some classrooms did not have all the nine teaching events. For example, several teachers did not stimulate the recall of the previous learning but directly informed the learners of the objectives. The phenomenon results in some teaching events being left blank in the statistics. Therefore, the TESTII framework groups the nine teaching events into teaching phases.

Actually, grouping teaching events into phases is not a new idea. Gagné classified the nine teaching events into three teaching phases, namely, preparation, instruction, and practice, and assessment and transfer (Gagné 1970). On this basis, Indian scholars Mishra and Gaba (2001) divided 15 teaching events into four teaching phases including introduction, new knowledge teaching, conclusion, and evaluation. In combination with the existing research results and our classroom observation, this study grouped nine teaching events into four teaching phases as introduction, new knowledge teaching, conclusion, and migration, as shown in Table 2.

Step 2: Sequencing Pedagogical Structure

The significant value of classroom teaching analysis is to identify high-quality teaching. Chinese scholar He (2002) proposed that the teacher-guided and student-centered teaching structure is the foundation of high-quality teaching and learning in the classroom. In He's opinion, the teaching structure refers to the stable structural

Table 2 Nine teaching activities and teaching phases

Instructional event	Teaching phases
1. Gain attention	1. Introduction
2. Inform learners of objectives	
3. Stimulate recall of prior learning	
4. Present the content	2. New knowledge teaching
5. Provide ‘learning guidance’	
6. Elicit performance (practice)	
7. Provide feedback	
8. Assess performance	3. Conclusion
9. Enhance retention and transfer to the job	4. Migration

Table 3 Teacher and student roles in the SPS

Role type	SPS	Code	Description
Teacher roles	H	T1 Lecturer	Oral explanation and explanation by the teacher
		T2 Questioner	Teachers directly ask questions and students answer directly
	h	T3 Instructor	Teachers provide models and guidance for everyone to learn, which can be words, actions, and process guidance
		T4 Facilitator	The teacher’s response and handling of students’ responses and behaviors
		T5 Collaborator	Teachers organize and guide students to discuss
Student roles	L	S1 Active learner	The students manage learning opportunities
	l	S2 Passive learner	The students may have some passive responses

form of the teaching process under the guidance of certain educational ideas and teaching and learning theories. This structure is the concrete embodiment of the interaction between the four components of the teaching system, namely, teachers, students, content, and media. However, the teaching structure is a macrolevel theory, and specific and relatively microlevel theories should be applied to directly identify the structure.

Sequencing of Pedagogical Structure (SPS), proposed by Jacobson et al. (2013), regards the teacher-centered direct instruction and student-centered learning as two poles. According to the proportion of teacher guidance or student discovery learning in different teaching phases, the SPS marked the phase with H or h means large or small proportion of the direct instruction of the specific stage, same for L and l about the discovery learning.

The SPS theory showed advantage in analyzing the teaching structure to some extent, but it failed to address the roles of teachers and students in the different phases. Therefore, this study introduced Schulman’s classification of teacher roles (Scheurman 1998) to facilitate coding for SPS as seen in Table 3.

Combining the theories of SPS and teacher roles, this study analyzed four lessons as example A, B, C, and D. Results are shown in Table 4. Time sequence is presented

Table 4 Example of four lessons with SPS coding

Lesson	Phase 1	Phase 2	Phase 3	Phase 4	SPS coding results
A	Teacher plays the video	Student cooperation	Students do exercises; Teacher analysis and statistics	Teacher leads students to summarize the methods of observing objects.	$H \rightarrow (L+h) \rightarrow (L+h) \rightarrow (H+L)$
B	Teacher shows voting box	Student cooperation	Students build models; Teacher evaluates models	Review and summarize new knowledge together.	$H \rightarrow (L+h) \rightarrow (L+h) \rightarrow (H+I)$
C	Teacher shows picture	Student cooperation	Teacher assigns exercises	Review and summarize new knowledge together.	$H \rightarrow (L+h) \rightarrow H \rightarrow (H+I)$
D	Teacher shows the cube	Teacher asks questions	—	—	$H \rightarrow H$

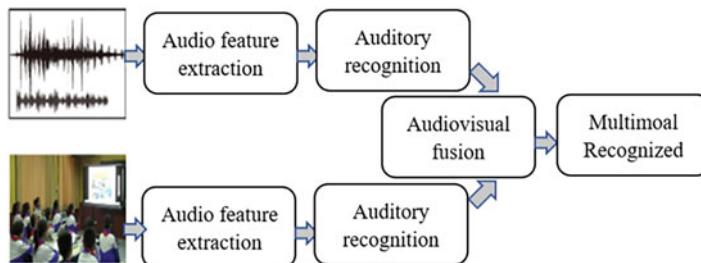


Fig. 4 Multimodal recognized analysis for interaction

as well. The plus sign (+) indicates simultaneous pedagogies, while the arrow (\rightarrow) indicates the sequence. The most common sequence in teaching is the high-to-low (H \rightarrow L) sequence, such as a lecture followed by unsupervised homework, while the low teaching structure sequence (L \rightarrow H) is probably the least common (Hu et al. 2016b).

To apply NLP technology, a teaching method structure sequence classifier should be established first. The input of the classifier is textual data, which contains contextual information. For a better understanding of the meaning of a sentence or word in the input data, the attention mechanism is introduced, which has the advantage of being able to intuitively explain the text content and show the importance of different sentences and words to the classification category. Sentence core words and event core sentences can be determined by the attention mechanism. The sequence of the teaching method structure is given by modeling sentences and chapters in text data.

Step 3: Time Coding for Interaction

Interactions between the teacher and the students in the classroom are important indicators of high-quality classroom teaching. Compared with the traditional single resource analysis methods, such as FIAS or S-T behaviors, TESTII conducts multimodal recognition via visual and auditory fusion on behaviors and discourse (Fig. 4). Instead of sampling the entire lesson through the time process, time coding is adopted within the teaching phases composed of teaching events. Then, the teacher-student interaction is analyzed in this stage, providing evidence for interpreting the teaching method structure sequence in the lesson examples.

To analyze the teacher-student dialogical interaction in different teaching phases, this study divided the teaching events into labels firstly and marked texts to form specific judgment rules. Subsequently, Word2vec, a deep learning model of natural language understanding, is used to train and verify the data. Then complete the automatic classification, analysis, and statistics of dialogical interaction in the current teaching phases.

As for the analysis dimension about behavioral interaction, the teaching scene is preliminarily classified according to the static frames. Then, the key interactive devices in the video are detected through the target detection method. Finally, the

actions of teachers and students are identified based on the deep convolutional neural network method. For instance, computer vision technology can judge the behaviors in the video like raising hand, walking, standing, writing on the blackboard, operating the tablet, and so on through matrix identification. Based on the analysis results, the features of teaching and learning behaviors could be figured out automatically.

Step 4: Interpreting the Result of Analysis

The explainability of the decision made and the actions taken is the core appeal of the future development of artificial intelligence and the premise of man-machine mutual trust. Teachers are not professional data analysts and thus require explanations that are easy to understand and conform to the rules of education and teaching to help them understand the analysis results of the machine, such as data content analysis, analysis of logic, analysis results, and problems identified.

On the basis of the aforementioned three steps, an interpretable, evidence-based visual analysis report is presented in Step 4. The report includes the number and time distribution diagram of teaching events in a lesson, the sequencing of the pedagogical structures of the classroom teaching, and the interaction of behaviors and discourse within each teaching event. A readable, effective, and persuasive data analysis report will help teachers implement specific teaching improvements while improving the credibility of teaching improvement plans, facilitating the transformation of data-driven teaching analysis to knowledge-driven teaching decisions.

Step 5: Improving Strategies Recommended

Providing effective improvement strategies for teachers on the basis of the analysis results of classroom teaching is the last and the most valuable step. According to the analysis results, the features of classroom teaching are identified. Then, the features are classified into kinds of teaching problems, such as teacher-centered structure and passive learning. Subsequently, the problems are matched with the database of effective teaching strategies and cases, which are recommended. Following the instruction and recommendation, which are collaboratively developed by the human–AI system, the teachers improve the teaching structure. The proposed AI-based analysis method takes the classroom as the main analysis object and it provides opportunities to build a flowchart model of classroom teaching analysis and improvements using Analysis–Problems–Strategies–Practice (APSP) method (Fig. 5).

The APSP model draws lessons from the core idea of “problem-solving and continuous inquiry learning” in the improvement science to identify shortcomings in teaching and improve teaching quality effectively.

The APSP cycle aims to answer four questions about the teaching improvement. (1) Analysis: What are the characteristics of the class? (2) Problems: What specifically are we trying to accomplish? (3) Strategies: What change(s) might we introduce and why? (4) Practice: How will we know that a change is actually an improvement?



Fig. 5 APSP cycle

To answer the four questions, the current chapter conducted four steps. Firstly, detecting the features or characteristics of the classroom teaching for further analysis; secondly, the features are categorized in different teaching structure types to address the problems; then, recommended teaching strategies for improvement matched to problems; finally, the strategies are applied to the teaching practice to improve the teaching quality.

Meanwhile, the APSP model integrated AI and human–AI technologies for recommended improvement strategies. In the beginning of our research process, experienced K12 teachers are invited as human experts to analyze many lessons and propose improvement strategies according to various problems. The experts' opinions and wisdom are classified into "question–strategy" pairs and stored in the database for machine learning. Then the experts are invited again to ensure or revise the "question–strategy" pairs created by machine learning which construct the human–AI collaboration improvement mechanism in the APSP model.

4 Conclusion

The chapter summarizes the development of classroom teaching analysis and improvement. Aiming at the problems encountered in the current stage, the TESTII framework of artificial intelligence is proposed to support classroom teaching analysis, taking teaching events as the basic analysis dimension, and forming five steps for teaching improvement.

Future teaching analysis would benefit from the integration with AI technologies. AI has the potential to make powerful impacts on the future of teaching and learning, which are reflected in the learning scene and the teaching process. AI for learning provides many applications and multimodal channels for supporting people in cognitive and noncognitive task domains (Niemi 2021).

TESTII framework has some limitations. The analysis of classroom teaching is based on event coding followed the Gagné's nine teaching events theory which is teacher-centered perspectives. Therefore, the student-centered classroom such as

inquiry-based learning, discovery learning should be considered in the future. The other shortcoming is that the major lessons are from elementary Math classroom. We would expand the research lesson database in the future.

In summary, the TESTII would keep on building multimodal analysis and human-AI integrated improvement mechanisms to optimize the quality of classroom teaching and learning. In follow-up research, artificial intelligence technology is expected to be applied to teaching practice and integrated into the main process of education, so as to form a deep integration of artificial intelligence and normal classroom teaching and make a high impact on the quality of teaching and learning in classrooms.

Acknowledgements This study was funded by the National Science Foundation of China (Research on key technology of classroom teaching interactive analysis based on artificial intelligence, Grant Number: NSFC61977048).

References

- Amatari, V.O. (2015). The instructional process: a review of Flanders' interaction analysis in a classroom setting. *International Journal of Secondary Education*, 3(5), 43-49. <https://doi.org/10.11648/j.ijsedu.20150305.11>.
- Bryk, A.S., Gomez, L.M., Grunow, A. (2011). Getting Ideas into Action: Building Networked Improvement Communities in Education. Springer Netherlands.
- Chen, L., Yang, D., Huang, X.R. (2018). Does mobile listening and assessment classes promote equal communication and in-depth collaboration among teachers: An empirical study based on mobile listening and assessment tools. *China Audio-visual Education*, 4(6): 107-114.
- Cheng, Y., Wang, Y.L., Wang, F., Zheng, L.L., Guo, Y.C. (2018). Research on the analysis method of teaching model based on the classroom teaching behavior cloud. *Modern Educational Technology*, 28(5), 61-67. CNKI:SUN:XJJS.0.2018-05-010.
- Cook, C.E., Wright, M., O'Neal, C. (2007). Action research for instructional improvement: Using data to enhance student learning at your institution. *To improve the academy*, 25(1), 123-138. <https://doi.org/10.1002/j.2334-4822.2007.tb00478.x>.
- Cui, Y.K. (2012). LICC Paradigm of Classroom Observation: A Professional Listening and evaluating Class. *Education Research* (05), 79-83. CNKI:SUN:JYYJ.0.2012-05-015.
- Derry, S.J., Pea, R.D., Barron, B., Engle, R.A., Erickson, F., Goldman, R., Sherin, B.L. (2010). Conducting video research in the learning sciences: Guidance on selection, analysis, technology, and ethics. *The journal of the learning sciences*, 19(1), 3-53.
- Ellis, N.E. (1990). Collaborative interaction for improvement of teaching. *Teaching and Teacher Education*, 6(3), 267-277. [https://doi.org/10.1016/0742-051X\(90\)90018-Z](https://doi.org/10.1016/0742-051X(90)90018-Z).
- Flanders, N.A. (1963). Intent, action and feedback: a preparation for teaching. *Journal of Teacher Education*, 14(3), 251-260. <https://doi.org/10.1177/002248716301400305>.
- Gagné, R.M. (1970). The conditions of learning. Holt, Rinehart and Winston.
- Goode, P. (2018). Using the ASSURE Model and Gagne's 9 Events of Instruction as a Teaching Strategy. *Nurse Educator*, 43(4), 205. <https://doi.org/10.1097/NNE.0000000000000514>.
- Gui, L., Liu, Z.Z., Chen, R. (2020). Real-Time S-T Analysis Method Based on Speaker Recognition. Proceedings of 2020 2nd International Conference on Advanced Control, Automation and Artificial Intelligence (ACAAI 2020) (pp.147-151). 10.12783/dtetr/acai2020/34199.
- Gu, X., & Wang, W. (2004). New exploration of classroom analysis in support of teacher professional development. *China Educational Technology*, 7, 18-21.

- Hatun Ataş, A., Delialioğlu, Ö. (2018). A question–answer system for mobile devices in lecture-based instruction: a qualitative analysis of student engagement and learning. *Interactive Learning Environments*, 26(1), 75-90.
- He, K.K. (2002). Theoretical Thinking on the Integration of Information Technology and Curriculum. *Information Technology Education in Primary and Secondary Schools* (Z1),27-36. CNKI:SUN: zxja.0.2002-Z1-009.
- Niemi, H. (2021). AI in learning: Preparing grounds for future learning. *Journal of Pacific Rim Psychology* 15, 1-12. <https://doi.org/10.1177/18344909211038105>.
- Horton, W.K. (2012). *E-learning by design*. Pfeiffer, San Francisco.
- Huang, W.G. (2015). Nine ‘Teaching Events’ – the second reflection based on Gagné’s teaching theory. *Educational Research and Review (Secondary Education Teaching)*, 6:5-11.
- Hu, B.Y., Dieker, L., Yang, Y., Yang, N. (2016a). The quality of classroom experiences in Chinese kindergarten classrooms across settings and learning activities: implications for teacher preparation. *Teaching & Teacher Education*, 57, 39-50. <https://doi.org/10.1016/j.tate.2016.03.001>.
- Hu, L.R., Mike, J., Zhang, B.H., Song, L.Q. (2016b). Flip or not? Transcend the trend of education and move towards a research-based classroom: Interview with Professor Mike Jacobson, University of Sydney. *China Audio-visual Education*, 138-141. <https://doi.org/10.3969/j.issn.1006-9860.2016.05.022>.
- Jacobs, J.K., Kawanaka, T., Stigler, J.W. (1999). Integrating qualitative and quantitative approaches to the analysis of video data on classroom teaching. *International Journal of Educational Research*, 31(8), 717-724. [https://doi.org/10.1016/S0883-0355\(99\)00036-1](https://doi.org/10.1016/S0883-0355(99)00036-1).
- Jacobson, M.J., Kim, B., Pathak, S., Zhang, B. (2013). To guide or not to guide: issues in the sequencing of pedagogical structure in computational model-based learning. *Interactive learning environments*, 23(6), 715-730. <https://doi.org/10.1080/10494820.2013.792845>.
- Jing, J. (2012). The design and practice of ‘Nine Teaching Events’ in information Technology courses. *Computer Education*, 24, 91-94. <https://doi.org/10.16512/j.cnki.jsjjy.2012.24.015>.
- Kane, T.J., McCaffrey, D.F., Miller, T., Staiger, D.O. (2013). Have we identified effective teachers? Validating measures of effective teaching using random assignment. In Research Paper. MET Project. Bill & Melinda Gates Foundation.
- Lewis, C., Perry, R., Murata, A. (2006). How should research contribute to instructional improvement? The case of lesson study. *Educational researcher*, 35(3), 3-14. <https://doi.org/10.3102/0013189X035003003>.
- Lewis, C. C., Perry, R. R., Hurd, J. (2009). Improving mathematics instruction through lesson study: A theoretical model and North American case. *Journal of mathematics teacher education*, 12(4), 285-304.
- Mehan, H. (1979). “What time is it, Denise?” Asking known information questions in classroom discourse. *Theory Into practice*, 18(4), 285-294. <https://doi.org/10.1080/00405847909542846>.
- Mewaldt, C., Mürwald-Scheifinger, E. (2019). Lesson Study in Teacher Development: A Paradigm Shift from A Culture of Receiving to a Culture of Acting and Reflecting. *European Journal of Education*, 54(2), 218-232. <https://doi.org/10.1111/ejed.12335>.
- Mishra, S., & Gaba, A.K. (2001). How do distance learners use activities in self-instructional materials?. *Indian Journal of Open Learning*, 10(1), 40-51.
- Mu, S., & Zhu, P. P. (2015). Research on classroom teaching behavior analysis method under information teaching environment. *E-education Research*, (9), 62-68. <https://doi.org/10.13811/j.cnki.eer.2015.09.011>.
- Pea, R., & Lindgren, R. (2008). Video collaboratories for research and education: An analysis of collaboration design patterns. *IEEE Transactions on Learning Technologies*, 1(4), 235-247.
- Rowntree, D. (1990). *Teaching through self-instruction: how to develop open learning materials*. Kogan Page, London.
- Roschelle, J., Lester, J., Fusco, J. (2020). AI and the future of learning: Expert panel report. *Digital Promise*.
- Scheurman, G. (1998). From behaviorist to constructivist teaching. *Social education*, 62(1), 6-9.
- Seldin, P. (2010). Using student feedback to improve teaching. *New Directions for Teaching & Learning*, 1989(37), 89-97.

- Stanulis, R.N., Little, S., Wibbens, E. (2012). Intensive mentoring that contributes to change in beginning elementary teachers' learning to lead classroom discussions. *Teaching & Teacher Education*, 28(1), 32-43.
- Sun, Z., & Ma, Y.H. (2012). The power of classroom teaching video-a new channel for teacher group learning in the Internet age. *Open Education Research*, 18(02):80-85. <https://doi.org/10.13966/j.cnki.kfjyyj.2012.02.002>.
- Zhang, H. (2019). Gagné nine teaching events in the secondary vocational graphic design course teaching application. *Chinese Educational Technology Equipment* 24. 79-81. CNKI: SUN: ZJJB.0.2019-24-027.
- Zhu, P., & Amant, K.S. (2010). An application of Robert Gagné's nine events of instruction to the teaching of website localization. *Journal of technical writing and communication*. 40(3):213-245. <https://doi.org/10.2190/TW.40.3.f>.

Open Access This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.



Part II

AI in Games and Simulations

Perspectives and Metaphors of Learning: A Commentary on James Lester's Narrative-Centered AI-Based Environments



Marianna Vivitsou

Contents

1 Introduction	125
1.1 The Key Message of the Keynote and Interview	126
1.2 Key Concepts and Metaphors	126
1.3 Modalities, Narrative, and Metaphors in AI for Learning Purposes	127
2 Crystal Island as a Metaphor for Learning with AI	129
3 Reversing the Double Narrative Process: The Agency of Students	132
4 Agential Cuts in Narrative-Centered Learning Environments	133
5 Summarizing Remarks	134
References	135

1 Introduction

This commentary aims to bring together perspectives on narrative-centered learning and through them to raise questions about how the narrative changes in the area of Artificial Intelligence (AI) when AI is used for learning purposes. The text is a constellation of modalities, as it is based on three interrelated contextual frameworks. One of them includes instances from the keynote speech of Professor James Lester delivered at the AI in Learning conference that took place online in November 2021 (Lester 2021). The second is an interview where Professor Lester further responds to questions posed by Professor Hannele Niemi and Postdoctoral researcher Jenny Niu (the interviewers from now on in this text). The third is this commentary on selected pieces of the keynote and the interview aiming to synthesize these with a focus on the narrative element that underlies the use of AI in Learning.

The keynote was originally in video format and the interview in a similar configuration. Later, the audiovisual texts were transcribed to ease access and the ability to refer to the details of the interactions.

M. Vivitsou (✉)
University of Helsinki, Helsinki, Finland
e-mail: marianna.vivitsou@helsinki.fi

Considering these multiple forms of textuality, it is evident that the two main sources of this chapter constitute expressions of the agencies of the participants of the communicative events of the keynote and the interview.

1.1 The Key Message of the Keynote and Interview

In the keynote speech and interview, Lester's (2021) overall goal is to discuss ways that AI technologies support education and learning. More particularly, the focus is on AI which, being a megatrend in our era, generates diverse public discourse. As Lester describes it, AI is often thought as a kind of "mysterious force." This metaphorical linguistic expression does not only present an interesting perspective on AI as a technological entity that is the carrier of a force surrounded by the mystery of the not-yet fully understood. Being a force means that AI has an impact on learning and, as such, is carrier of a certain kind of agency.

The aim of Lester's (2021) speech is to bring forward key issues in AI-enhanced learning and how it can be promoted through narratives. Lester's reflections are illustrated with Crystal Island, a game that offers to learners opportunities to develop understanding through storytelling and problem-solving.

The focus of the keynote is on narrative-centered learning that entails the pedagogical use of stories and storytelling for deep learner engagement.

This commentary will focus on the concept of narrative-centered learning and related ones, such as tutorial dialogue and characters.

1.2 Key Concepts and Metaphors

The multimodal texts, therefore, that inform this commentary bring forward powerful contemporary metaphors that refer to AI-enhanced learning in the keynote and the interview.

One such metaphor is narrative-centered learning that runs through the keynote talk and signifies how the agencies of researchers, teachers, and students intertwine with technologies in physical and online environments in the passage of time to construct the metaphors of AI-enhanced learning in the future. The discussion is illustrated with Crystal Island, a game-based learning environment that aims to engage students in story-based activities with believable characters and problem-solving features at the core of the storytelling.

In addition to narrative-centered learning, other key metaphors in the keynote are conveying already established concepts, such as tutorial dialogue and characters. Some others, such as the drama manager, are new and signify ways of supporting the process of learning with AI technologies.

This commentary then aims to bring forward the metaphors associated with the notion of narrative in learning and how they relate with the development

of the Crystal Island as a game to support students' constructing knowledge in science education. To this end, the commentary draws from Paul Ricoeur's (1978, 1986, 1992) narrative and metaphor theory and introduces aspects from the work of new materialist and post-humanist thinkers (e.g., Barad 2007; Coleman 2020; Stark 2016; Truman 2019) with a focus on the role of technology. In both the narrative and the new materialist/ post-humanist theoretical standpoints, agency is a critical notion. Based on these, the commentary draws from relevant concepts and metaphors in the keynote aiming to take further the narrative of agency in AI-enhanced learning.

According to Coleman (2020), agency becomes visible and understood through temporal, spatial, and material modalities. Modalities signal the ways agency is organized, distributed, and displayed. It is only natural then that in multimodal texts about AI-based environments (like the keynote and the interview are) the multiple modalities convey agency through a multiplicity of metaphors of learning.

The metaphor of agency, although not explicitly stated in the keynote and the interview, is an all-encompassing one. As multiple modalities converge in the audiovisual display, the scholarship, the background, and interests of the keynote and interview participants are revealed. The video of the keynote, for example, aims to communicate the speaker's message to researchers, scientists, teachers, and other audiences with whom an interest in the impact of AI on education in the future is shared. As Lester puts it in his keynote,

One is, ... I think [we will be] seeing fascinating developments in the upcoming five years or so in AI technologies to support education, which is really the focus of this talk. But it is also the case we are going to see some really interesting developments in 'AI education' per se, that is, AI as the subject matter for K-12 education.

The agency of the speaker in the study and research of AI, although outspoken in previous and later parts of the keynote text, here is resting between the lines. Nevertheless, it is underlying the considerations and imaginaries that the keynote expresses. The material dimension of AI will strongly impact the way education takes place in the actual, physical environment of the classroom and school in the future. The narrative of education will, therefore, change in the days to come with the use of AI. It is the directions of the change of narrative that the keynote aims to capture. Similarly, the interview questions target the visualizations of future changes and depart to bring into light their finer shades.

1.3 Modalities, Narrative, and Metaphors in AI for Learning Purposes

The role of the narrative, therefore, is double here. First, there is the narrative that the multimodal texts construct concerning the future of AI in schools. Second, there is the narrative-based approach that is integrated in applications of AI for pedagogical purposes. Indeed, as the work of the philosopher of language Paul Ricoeur has

shown, there are diverse forms and modes of narrative. According to Ricoeur (1986, 1992), despite the diversity, all narratives present universal elements. They perform a common function, namely, they mark, organize, and clarify temporal experience (Ricouer 1986).

The temporal experience that is organized and clarified through the conventions of the narrative does not concern the storytellers themselves whose agential knowledge, values, and practices are transferred through the storyline. It mainly concerns the lived experiences of the characters whose actions, events, and relations the stories are telling. The narrative of AI for learning purposes, therefore, emerges out of the agencies of its authors and tells the story of the agencies of its characters.

The plot of the narrative makes it possible to synthesize the experiences of the characters by organizing the story through, for example, expressions of time, descriptions of settings and backgrounds, and so on. In this way, through narrative plot, the meaning of persons, relations, and events that make up life affairs become visible. In this sense, the plot and the characters develop in a dialectical way. The development of the plot cannot happen without the actions, thoughts, decisions etc. of the characters. Neither can the characters grow outside the temporal and spatial configurations of the plot (Ricoeur 1986, 1992).

In this commentary, the plot of the narrative aims to make visible how students and teachers in K-12 education use AI for learning and what meanings emerge out of this use.

To make the multifunctional performance of the narrative possible, speakers and writers use metaphors. Metaphors can be of different types and so are metaphors of learning, multiple and shifting. How metaphors shift, for example, how novel or conventional they become, depends on the era and its socioeconomic and political developments.

As Lester explains,

There are many types, there are many *metaphors of learning*. I think it is fair to say that for the history of our field one of the most significant and powerful metaphors that we had since the beginning, since the 1970s, 50 years now, is tutorial dialogue. It is a very exciting area, it is an area that our group has worked in, and I know many people in conference, your labs are working on this too. I have seen your program, which looks fantastic. It is such a great metaphor. It is a really interesting development over, roughly the last 2,000 years that we have come to understand that human tutoring, where human tutoring engages with dialogue, in dialogue with the human student is incredibly effective.

It is arguably one of the most effective, if not the most effective approaches that we have. It is curious, we don't know exactly why this is. Right? It could be for the self-explanation effect. It could be because of very powerful learning mechanisms that are kind of released you might say when students engage in human dialogue with the tutor. There could be a very strong effect on components, for young learners. And likely it is a result of all of these and even more. This is one metaphor out of many, many possible metaphors and I would like to suggest one I think is particularly interesting and one we will be focusing on this morning's remarks which is known as narrative-centered learning.

Evidently, in the section above, Lester acknowledges the diversity of metaphors that relate to learning and the development of metaphorical language and its meanings as time progresses and technology advances. This reflects Ricoeur's (1986, 1978)

consideration that is built from the claim that metaphor should be grasped not as the substitution of one conventional name for a different one. When it comes to dialogue, tutoring and learning, we should go beyond the conventional meaning of the words by setting the ground for new imaginaries of tutorial dialogue. In this way, the metaphor of tutorial dialogue can evolve and transform into a metaphor of narrative-based learning. The question then arises: Is narrative-centered learning an innovation?

As Lester further elaborates,

Narrative-centred learning is in some ways not a new metaphor at all. The sort of recognition of the importance of story for human learning that sort of episodic memory that it triggers. The deep engagement that often contracept when students engage in it is a sort of hallmark of narrative-centered learning. But what I would like to suggest is that, in fact, because of the very recent developments in AI it is not going to be possible to really create an incredible powerful narrative-centered learning environment.

Narrative-centered learning links, therefore, with different theories of learning that have evolved in time and can have an impact on students' memory, engagement, and so on.

As Lester continues,

So really two parts of this discussion this morning. First is kind of looking at what you might call narrative-centered learning environments look like today. We look at one, and this is sort of an exemplar. It is kind of like a little case study. And what I would like for you to do when you look at this, think about how this kind of narrative-centered learning environment could in fact be kind of the laboratory for studying narrative-centered learning with 'AI full on', with fully supporting learning interactions.

It becomes evident, then, that while narrative-centered learning is not new (i.e., this is a conventional metaphor), the integration of AI system into the narrative approach can be innovative for learning and pedagogical purposes.

2 Crystal Island as a Metaphor for Learning with AI

To illustrate the innovative dimensions of narrative-centered learning with AI, Lester uses the example of Crystal Island in the following section:

Chrystal Island is a narrative-centered learning environment that has been under development by our group over many, many versions over many years. You can think of narrative-centered learning environments as a kind of an intelligent game-based learning environment. ... [T]here is a great attraction to having students to participate these story-centered activities that are fundamentally featuring problem solving in a way that fully integrates the story with the problem solving. And the students, ..., emerge themselves in these narratives. The narratives can be more or less powerful. They can be more or less well-designed; they can more or less effectively integrate pedagogical [purposes] into the learning experiences.

As it happens with well-organized narratives, there are specific elements that characterize well-designed interactive narratives for learning.

As Lester goes on to argue,

One is *believable characters*. So of course, enormous amount of work for many years and non-player characters (NPCs), and the words that are very expressive and captivating and then finally rich stories that unfold over time. So, these are core characteristics of narrative-centered learning environments which tend to have certain kinds of effects. One is that unlike many kinds of learning there is actually a very strong elicitation of learner affect in narrative-centered learning environments, and affect has a very strong impact on performance. It can be a positive impact. It can also be negative. Supporting effect is very important as we know in kind of more traditional tutoring, and it is really kind of core characteristic in many forms of learning that can contribute into effective learning. It is kind of particularly amplified in narrative-centered learning.

Indeed, in the Ricoeurian narrative theory, the character is not only an essential element. Most importantly, the character is in dialectical relation with the plot (Ricoeur 1992). This means that as the plot of the narrative evolves, the characters evolve as well. In addition, the events, actions, emotions, and relations that the characters are entangled with move the plot of the story forward. Therefore, beyond the expressiveness of words, the agency of the characters makes them rich and captivating, as stories unfold over time. The characters' agency is interconnected with whom those characters are. In the Crystal Island game-based learning environment, they represent different genders, racial and ethnic backgrounds. This attributes an innovative element to the game since the referential function of the Island narrative contributes to new imaginaries of AI-based design of games for learning purposes. This means that believable characters reshape the reality of games and display the world as multicultural and diverse.

And yet, Crystal Island represents only a small portion and, no matter how deeply we would wish for it, the world is not an island. How does then the Crystal Island metaphor speak to the rest of the world?

Under this lens, the interviewers ask,

Interviewers: So, . . . thinking then for the future, now [that] you have this knowledge from creating this wonderful environment... [B]ut... do you think that people in different countries could do something similar, based on what you have done during [these] fifteen years? Or should they do everything just from the beginning?

In response, Lester explains,

There are so many developments in the last, let's say, five, years or so, that I think is going to make it much, much, much easier to create these environments for everyone. One of the developments is that often at the sort of foundational infrastructure level, there are game technologies and there's—Finland of course is famous for this—such an enormous investment in the underlying technologies, for game engines that “for free” we researchers are able to leverage all of the 3D worlds, the characters, the game playing mechanics, all kinds of computational capabilities that these game engines offer and that's our starting point. Rather than starting from nothing, we can start from that, which is very helpful. Then there's a sort of collection of know-how or maybe best practices that have begun to evolve. So, we start seeing the literature, but we also start seeing in discussions and conferences. Shared interest makes it possible to not only do it kind of more efficiently, because of shared knowledge, but also more effectively. And the third and final thing I mention, which is in my own view the most exciting, is that over the next —let's say five years, seven years something in this time frame— we're going to be seeing the emergence of AI technologies

that underlie all of it. That will make it amazingly, if not easy, a lot easier to actually create these kinds of game-based learning environments. And that's the thing, we don't exactly know how that's going to happen, but it's very exciting.

This explanation signifies the need for a transdisciplinary approach to narrative-centered game design for learning. The consideration and integration of theories and practices from the literature is where perspectives from various scientific discourses, including computer science, human-computer interaction, and science education, intertwine. However, Lester's response in the section above brings forward mainly technological metaphors. These make visible the significance of the role of technology as game changer in the educational discourse of the future. Although the narrative-based game design should consider contextual, social, cultural, economic, historical, and other factors, the technology itself interacts with all of those. Technology, therefore, has an impact on the ways agency is organized, distributed, and displayed in space, time, and materiality.

In this sense, as many new materialist and post-humanist thinkers (e.g., Barad 2007; Truman 2019) would possibly agree, technology itself has agency. As Lester explains, different infrastructure will be needed to serve the needs of Finnish students if narrative game-based learning migrates to, for example, Finland. This would possibly include algorithmic configurations and design that consider the sociocultural dimensions of the learning context.

This speaks to the fact that techno-material (more-than-human or nonhuman) entities interact with the agency of humans. In this sense, techno-materialities bear their own agential qualities.

Moreover, this means that the integrated narrative has an impact on the ways the whole narrative plays and pushes the wider discourse of education and technology-enhanced learning forward.

The role of the characters also shifts, and new agents come into play in order to make possible the integration of Crystal Island in a context other than the one of its origins. For this kind of migration, a labor-intense process takes care the needs of the students on an individual basis and the new role of the drama manager is introduced. The drama manager plays a critical role here. As Lester goes on to explain,

So, in this approach we first create kind of a base line learning environment. It can be like Crystal Island. And then students one by one, typically in a laboratory setting in this approach, will interact with the game. So, they will solve a science mystery, they will talk to the characters, they will fill out diagnosis work sheets, if it's about sort of diagnostic task. So, sort of that kind of thing. But, unbeknownst to them, so they don't know this, but sitting often in another room is a kind of 'expert drama manager.' So, this is a person who is actually controlling when the character does this, or when a particular event in the world does that, so you can sort of imagine little switches been flipped so that the drama manager is actually the one creating a very personalized interactive narrative for the student. So, when you did that for many students, it's of course incredibly labor intense because you're doing it one by one and it's kind of interesting process.

3 Reversing the Double Narrative Process: The Agency of Students

In addition to adult human (e.g., data manager) and technological characters, the previous section introduces the agency of young students that comes into stage on an ongoing basis during the experimental phase of the game environment.

As the double narrative of actions, reactions, and interactions of technological and human entities unfolds, the agency of students as main characters becomes more visible in the feedback process of the experimentation.

As Lester goes on to argue,

... The long-term effects are having a strong potential for deeply motivating learning experiences and promoting learning characteristics for example like self-advocacy. These learning environments when they are done well have effective characters, and problem-solving guidance. Feedback is context-sensitive. Problems, which you can think of sort of narrative episodes, can be dynamically selected, and explanations can be tailored depending on the needs of students. So, this particular learning environment, Crystal Island, . . . , is one we have been working on for a very long time. And, in it the student plays a part of *protagonist* who actually goes to a remote island and finds out that members of their research team are falling ill.

As the integrated narrative process unfolds, the focus reverses into the wider context of the school, where students hold a protagonist (or main) role, as Lester argues. The agency of the students as main characters is manifested through opportunities to explore the environment as well as challenges underlying the learning situation. To deal with them, the students put their reasoning into action to come up with solutions. These actions match the needs and pedagogical objectives of science education. In the process, actions intertwine with the materiality of technology that itself acts to learn from the agency of the students in a situation that constitutes a differential diagnostic test, as Lester describes.

In the following section, Lester offers an account of the characteristics that make these environments attractive poles for thinking about how to integrate AI into learning.

One is that there is exploration of virtual environments. Two is that there are often very knowledge rich components in the environment. They can be sprinkled in to provide [. . .] resources for students and their problem solving. There can be arbitrarily a simple or complex kind of virtual equipment, in this case for *science education*. They can support very complex reasoning. In this case it is for differential diagnosis. There can be multiple subject matters integrated. In this case it is science and complex informational text comprehension. And then stealth assessment is [a] really important area. I think [what is] promising in this particular metaphor is being able to combine assessment into the narrative.

So really three kinds of promising ways of thinking about interactive narrative. One is that it is a laboratory of investigating learning, super important from a research perspective. Second, it is a great place to study AI learning analytics because of the enormous data that these things produce. Often on very granular levels. And finally, the one I am myself particularly excited about and I imagine you might be as well, which is that it is kind of a lab for investigating new and very, very promising AI learning technologies. I want just to quickly say that there are lots and lots of domains, and lots and lots of student populations and actually lots of settings too that narrative is kind potentially applicable too.

I just quickly mention passing, this is a narrative-centered learning environment for middle grade's computational thinking that we have been working on for many years.

Then, what you have is, you've got all this data from student problem-solving interactions, and it's all captured in the trace data. So, it's all in the way that the student moves around in the world and manipulates artifacts, interacts with characters, takes these little stealth-assessments and so forth, but you've also got the 'expert drama manager' as they're making the decisions about how the narrative should happen. And ... that's a supervised machine learning test.

This account speaks again to the need to pay close attention to the ways students and technology influence one another. In other words, how we relate with technology is a matter that matters, as it constitutes one ethical dimension of the role technology plays in AI-enhanced learning.

4 Agential Cuts in Narrative-Centered Learning Environments

As it was mentioned in previous sections of this chapter, the convergence of modalities in the audiovisual display allows the agency of the participants of the communicative event to emerge. Evidently, different forms and types of agency make an appearance here. As Stark (2016) argues, agency is, rather than constant, a fluid entity that intertwines and intra-acts in material objects and bodies through space and time. Under this lens, agential intra-action is seen as a dynamism of forces, rather than an inherent property of human beings (Barad 2007). It is this dynamism of forces that allows us to experience the world and, therefore, to relate with the world.

In a similar way, the perspectives that emerge though the convergence of modalities in this chapter are associated with a multiplicity of metaphors. These are metaphors of learning that, as Ricoeur has shown, make visible possibilities of reality that can orient agency and contribute to the effort to reshape reality. In this sense, the perspectives of the participants of the communicative events (e.g., keynote, interview, audiovisual, written text, etc.) actually constitute agential cuts.

Agential cuts are forces of bodies, objects etc. (Stark 2016) whose ongoing movement and intra-action transforms the way we understand the world. In AI-enhanced learning with narrative-based learning environments, Crystal Island is an example of agential cut that the agencies of both human (i.e., computer scientists, designers, researchers, teachers, other practitioners, students) and more-than-human (i.e., AI, digital technology, algorithms, etc.) entities both construct and transform.

Most certainly, there are issues for consideration here. It is debatable, for instance, whether the agency of technology is a valid notion. This discussion goes beyond the limits of this brief commentary. However, it might be worth mentioning here the example of the brittle fish (adapted from Barad 2007) as a response to the long-held belief that agency is associated with human consciousness only. The brittle star, a relative to the starfish, manages to develop a visual system to

avoid ocean predators without the aid of actual eyes and brain. Without the brain organ, it is hard to imagine that survival is possible. Despite this, the brittle star survives thanks to the spherical calcite crystals covering its limbs and central body, functioning as micro-lenses that collect and focus light directly onto its diffuse nervous system. In this way, even without a nervous system, the brittle star manages to escape its predators and survive. Agency, therefore, seems to not link with brain function and consciousness necessarily.

5 Summarizing Remarks

This brief commentary aims to bring together perspectives that arise from the double narrative of multimodal texts (keynote and interview) and the agential cuts that emerge from metaphors of AI-enhanced learning and the entanglement of experiences and actions of scholars, researchers, teachers, and students. In this way, it touches upon the ethical dimensions of technology in AI-enhanced learning.

The multiplicity of metaphors includes both older and newer, conventional and novel ones. The tutorial dialogue, a metaphor that Lester (2021) introduces early in his keynote, is not new. The tutorial dialogue is traced back in history with the Socratic dialogues being possibly the first notable example of teacher-student interaction, where Socrates teaches his students logic, reasoning, argumentation, and ethics. Later, the narrative-centered learning environment emerges through practice, in time. Even this is not a new metaphor at all, it acquires new meanings when associated with AI-enhanced learning metaphors.

Some metaphors seem to be in fluidity, as their meaning transforms in time. The discussion in this commentary shows that these are mainly metaphors associated with agential cuts, that is, dynamic forces of human and more-than-human entities that move, intra-act, and transform in time and space.

Another conventional metaphor is the student being a protagonist in school environments with student-centered orientations. However, how the narrative of student-centeredness becomes believable remains an issue when it comes to AI-based learning. The example of Crystal Island seems to offer possibilities for engaged learning with the spaces it opens for exploration, experimentation, and the new roles it generates in its experimental process. The role of drama manager, as Lester describes it, resembles that of the tutor. It could be the basis for an agential cut in the future.

Its current orientation, however, seems to be targeting the improvement of technology exclusively. In ancient Greece, “drama” is a type of narrative and, as such, refers to the actions, events, and relations of the characters that shape its plot. The noun “drama” is associated with the verb *δρω* (Greek for /dro/ meaning “act”). The drama manager should then take care of how students relate with the world rather than how they interact with technology only.

The visualization of the future, as eloquently expressed by Lester (2021), takes the metaphor of AI for learning forward with the multicultural, inclusive Crystal

Island. And yet, the technological metaphors are not enough. More thought and transdisciplinary discussions and collaborations are needed with scientists and pedagogues to articulate clearly how the agential roles of students and teachers are redefined in AI-enhanced learning.

As Lester rightly puts it, AI-enhanced learning will always be built on pedagogies of care and therefore the employment status of teachers is not threatened. Indeed, although the employment of workers has been the object of heated discussions since the 1950s brought up by the rapid advances of systems of automatization (Arendt 1998), the world will always need teachers who care.

In an era that is shaken by the COVID-19 pandemic and the larger questions concerning the sustainability of the planet, the role of teachers cannot be confined in the teaching of how technology functions. Teachers should be able to teach, among others, what environmental crisis means, what climate injustices are really about and who are most inflicted by them, what indigenous knowledges are and how they are downplayed. These could be part of science education curricula on the one hand. On the other hand, computer scientists need to think deeper how to integrate these realities into their algorithmic configurations. After all, science education does not happen in a vacuum.

And these can be some ways to move the narrative forward, having considered the crucial ethical questions that Lester poses at the very beginning of his talk, about «[W]hat happens when the AI, which this very powerful force, is kind of unleashed on the world».

Acknowledgments Many thanks to Professor James Lester, North Carolina State University, for the inspiring talk toward narrative-centered AI-enhanced learning environments. Also, for kindly reading and commenting on the text. Special thanks to Professor Hannele Niemi for the deep, thought-provoking discussions during the process of writing the commentary and beyond that.

References

- Arendt, H. (1998). *The Human Condition*. University of Chicago Press.
- Barad, K. (2007). *Meeting the universe halfway: Quantum physics and the entanglement of matter and meaning*. Durham, NC: Duke University Press.
- Coleman, F. (2020). Feminising politics: notes on material and temporal modal logics in action. *Matter: Journal of New Materialist Research*, 1: 1-22.
- Lester, J. (2021). AI and the Future of Education. AI in Learning – Shaping the Future!, University of Helsinki, Helsinki (virtual), 18-19 November 2021.
- Ricouer, P. (1986) [1991]. “De l’interprétation,” in *De Texte à l’action: Essais d’hermeneutique II*, Paris: Éditions du Seuil, 13–40. Translated as “On Interpretation,” in *From Text to Action: Essays in Hermeneutics II*, Evanston: Northwestern University Press, pp. 1–20.
- Ricoeur, P. (1978). *The Rule of Metaphor: Multi-disciplinary studies of the creation of meaning in language*. Transl. Robert Czerny with Kathleen McLaughlin and John Costello. London, UK: Routledge & Kegan Paul Ltd.
- Ricoeur, P. (1992). *Oneself as Another*, trans. Kathleen Blamey, Chicago: University of Chicago Press.

- Stark, W. (2016). Agential Cut. New Materialism - How matter comes to matter, Almanac. Available online: <https://newmaterialism.eu/almanac/a/agential-cut.html> Accessed 15 January 2022.
- Truman, S. E. (2019). Feminist New materialisms. In Paul Atkinson, Sara Delamont, M. A. Hardy, & Richard A. Williams. eds. *The SAGE Encyclopaedia of Research Methods*. London, UK: Sage.

Open Access This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.



Learning Career Knowledge: Can AI Simulation and Machine Learning Improve Career Plans and Educational Expectations?



I-Chien Chen, Lydia Bradford, and Barbara Schneider

Contents

1 Introduction	137
2 Making More Informed Career Choices: A Theoretical Framework	138
2.1 Why AI About Career Knowledge?	140
2.2 An Example of Gamified Career Knowledge: Init2Winit, an Overview	140
3 Opportunity for AI and Machine Learning (ML)	144
3.1 Machine Learning and Decision Trees	145
3.2 Empirical Example: Decision Trees Algorithm in Init2Winit	145
4 Result	147
4.1 Init2Winit Users' Profiles	147
4.2 Init2Winit Users' Classification for Multiple Goals	147
4.3 Alignment Knowledge of Decision Trees and Partition	148
4.4 Regression Decision Trees and Prediction of Educational Expectations	151
5 Strengths and Weakness of Current Design	153
6 Conclusion and Recommendation	153
Appendix A Record of e Daily Active Users	155
Appendix B Iterative Dichotomiser 3 (ID3) Algorithm	155
References	156

1 Introduction

Innovative automation advancements are profoundly affecting markets and societies in a rapidly changing information world (Arntz et al. 2016). Additionally, for young adults and those who have lost their jobs, the employment landscape is characterized by ambiguity and insecurity (Blustein et al. 2020a, b). Knowing the demands and requirements of specific jobs can be helpful for those seeking employment. How to align individual career goals and specific employment opportunities requires sophisticated information, guidance, and navigation (Kim et al. 2019; Nunley et al. 2016; Pinto and Ramalheira 2017). This process can become less complicated with

I.-C. Chen (✉) · L. Bradford · B. Schneider
Michigan State University, East Lansing, MI, USA
e-mail: ichiench@msu.edu; bradf134@msu.edu; bschneid@msu.edu

machine learning applications. Relying on several studies using game simulations and machine predictions to assist young adults in their career selections (Nie et al. 2020; Schumacher et al. 2010), this chapter explores the unique features of gamification in learning, machine learning, and artificial intelligence (AI) technology. The logic of gamification is described showing how these applications have been implemented to understand players' capacity, skills, and interests in selecting future occupations. This process includes machine learning decision tree algorithms that map out possible job selections, built upon players' career choices and opportunities given their background characteristics to increase the prediction precisions. Results from data insights can be implemented into a series of games to enhance users' knowledge of possible college and career choices. Finally, there are advantages of connecting mobile application, machine learning, and data insights used for predictions which extend user career knowledge, especially in domains where information is often ambiguous and inaccessible.

Unquestionably, young people today play games, often on their phones. Game technology has become a mainstay of entertainment, and it is a prime avenue for games that are challenging, fun, and transmit information at the same time. One area that has not yet been successfully designed and gamified is learning the link between education and career choices. This is a particularly important situation today, given that careers have expanded so rapidly, and vital information is not codified into an easily accessible place. Combining career and educational requirements—a combination where they can learn about emerging jobs, their corresponding educational requirements, and prospects for potential hiring, salaries, security, and advancement—is critical for adolescents' planning for the future.

Init2Winit was developed to fill this gap within smartphone technology. The gamified architecture follows the front-end, back-end system which introduces the participants into career knowledge and gameplay that is transferred to a confidential and de-identified individual database. The overall goal of Init2Winit is to help students learn more about the college-to-career process, which in turn will inspire students to improve their college applications and widen their college and STEM major choice options. The Init2Winit design combines a personalized exploration of career goals to assess individual-level alignment knowledge of the pathways from education to employment.

2 Making More Informed Career Choices: A Theoretical Framework

Recognizing the problem that limited information can create for informed college and career planning led to the creation of the theory of aligned ambitions (Schneider and Stevenson 1999). Alignment theory refers to a status of “aligned ambitions”

for young people who begin to develop an emerging understanding of the types of jobs they aspire to, how much education they need to attain these positions, and realistic projections on the annual salary. When young people are more aware of their abilities, strengths, and skills, they are more likely to develop a strategic plan that aligns education expectations and aspirations for their career goals.

Renbarger and Long (2019) find that a lack of access to information on financial aid and college programs has detrimental effects on college enrollment and completion. Cohodes and Goodman (2014) also find that students in disadvantaged schools have limited information on how to apply for college or meet important college-related deadlines. As a result, many students may not know how to make a smooth transition from education to employment nor how to navigate an educational system where choices have real consequences on postsecondary enrollment, degree completion, and employment (Castleman and Goodman 2018). Not having a realistic sense of aligned goals can keep students from being able to focus on the required courses, preparation, and skill development.

The consequence of misaligned knowledge has been shown to result in overestimating or underestimating requirements for college for a career pathway (Schmitt-Wilson and Faas 2016; Perry et al. 2016). Under-aligned high school students assume the pathway to specific jobs can be achieved without completing a postsecondary degree (Kim et al. 2019). The consequences of misaligned knowledge for low-income students can be costly, leading to financial debt or dropping out before obtaining a college degree (Morgan et al. 2013; Bettinger et al. 2012). A recent study has shown that nearly one-third of low-income students had under-aligned career expectations (Chen et al. 2020, 2021). Under-aligned students, while able to estimate a realistic salary range for a job, often were unaware of the educational requirements for a desired job. Students with misalignment knowledge in high school show significantly lower educational expectations, college preparation, and school GPA (Kena et al. 2016; Schneider 2009).

The prevalence of misalignment among low-income students also occurs for students outside the USA. PISA 2018 results show that one-third (30%) of young people from disadvantaged backgrounds are more likely to have misaligned career expectations than one-tenth of their advantaged peers across countries (Mann et al. 2020; Nedelkoska and Quintini 2018). The impact of misalignment has become a global issue due to uncertainty surrounding the job market and automation, and risks have risen in the digital era, particularly among lower-educated workers. Although past research has identified the gap between young people's desired jobs and employment realities, research is lacking on how these differences correspond to labor demands, college knowledge and eligibility, and an individual's needs on a case-by-case basis (Hoff et al. 2021; Schneider and Young 2019; Albion and Fogarty 2002). Career knowledge is critical to help guide an individual's efforts and decisions about college planning during high school.

2.1 Why AI About Career Knowledge?

The introduction of AI applications with megatrend data gathering and forecasting would benefit this decision-making process. Enrolling in a college or finding a job is not a simple cost/benefit question. As concerns rise over mismatched expectations, overqualified skills, or youth unemployment, making better decisions to optimize an individual's strengths, considerations, interests, and skillsets becomes imperative for young people. The decision-making process tends to rely on information and situational assessment to navigate a personalized college-to-work pathway, which is needed to warrant the success of college and work life (Reyna and Farley 2006; Clark et al. 2017; Bureau of Labor Statistics 2015).

2.2 An Example of Gamified Career Knowledge: Init2Winit, an Overview

Init2Winit integrates data-based analytics with occupational information algorithms that allow users to make choices with respect to their education planning and salary projection in visualizing themselves in a dream job. Init2Winit uses points as a feedback mechanism to encourage student participation and performance. Point feedback aims to motivate students to sustain their effort and continue their exploration across different jobs, even for those jobs or college majors that are beyond the students' current plans. To further motivate participation and build college knowledge, Init2Winit allows student performance to be translated into real-world rewards. For example, if a student remains a top five scorer for a week, he or she could earn a voucher for a college visit or an internship with a local company.

2.2.1 Game Design

The gamified architecture structure of the Init2Winit lays out front-end engagement features and a back-end database. The following is an example of the Init2Winit game, designed to motivate a personalized exploration of postsecondary planning and career goals.

2.2.2 Engagement and the Front-End Design

The front-end development focuses on those components of the game that the user sees and interacts with, such as the graphics, interactive user functions, and audio components. The importance of the Init2Winit user experience (UX) design is to keep students' attention on the college-to-career information that students may not know. The game mechanics are a set of rules that dictate the outcome

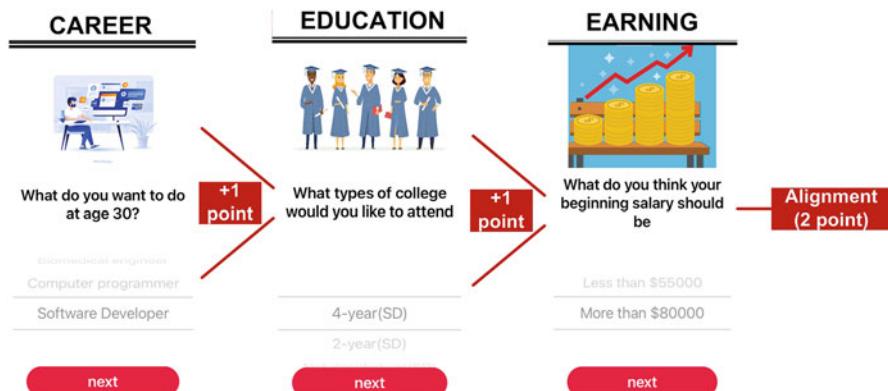


Fig. 1 The full-alignment scenario in the point reward process

of interactions within the system. The data collected are the users' responses to those mechanics. These coupled with an algorithm based on student responses was operated through an interactive interface – using points as real-time feedback on their level of alignment knowledge.

Alignment knowledge indicates that a student can visualize himself/ herself in a career pathway with aligned educational expectations and realistic salary projections. Figure 1 shows an example of how to earn full score points in one play. If a student chooses software developer as a career, he or she needs to know what the educational requirement is for this job and the yearly salary range. When the three informational pieces line up, the user earns the full-alignment score of 2 points. With this knowledge and preparation beforehand, the students are likely of knowing more about employment opportunities in the future.

A student with misaligned knowledge typically chooses either unaligned educational expectations or an unrealistic salary projection. These two types of misalignments cause different consequences to the student. Students with under-aligned knowledge are unaware of the requirements for a job or chose a lower yearly salary than reality. For example, Fig. 2 shows that a student who wants to be a “registered nurse,” selects a 4-year college degree, but incorrectly predicts earning less than a \$20K yearly salary, indicating a misunderstanding on the salary in the workforce for life science and health-related professionals.

Students with over-aligned knowledge expect to obtain more degrees than required or overestimates the potential annual salaries for their desired career choices. For example, Fig. 3 shows a student who wants to be a “police officer” chooses a 4-year university degree, and expects to earn more than \$100K. These choices indicate a misunderstanding of the required education or profession for being in the law enforcement institute (Schmitt-Wilson and Faas 2016). Students earned 0 points if their alignment between career and college planning and career and salary projections are over-aligned.

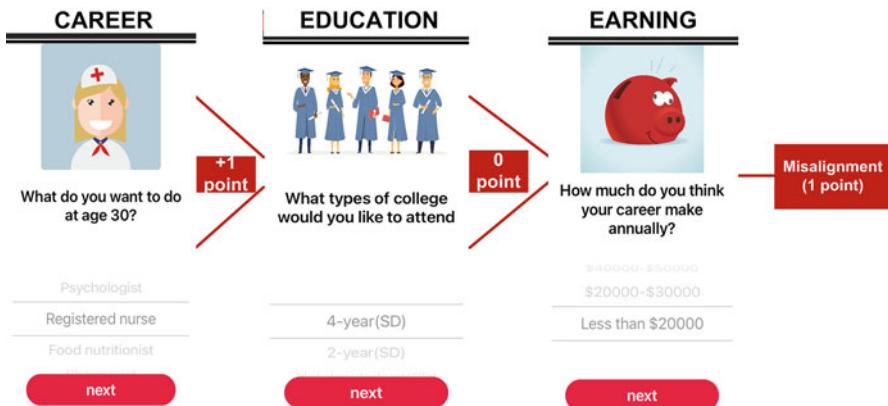


Fig. 2 The misalignment knowledge scenario in the point reward process

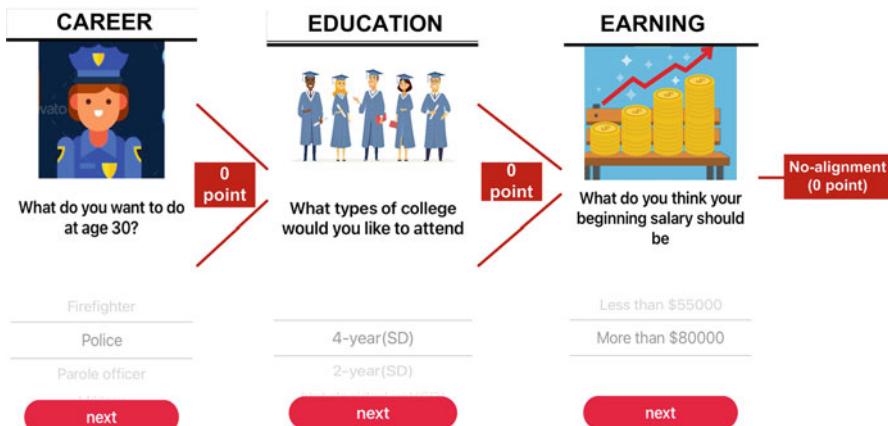


Fig. 3 No-alignment knowledge scenario in the point reward process

Computer-generated images (CGI) help to engage users during gameplay through augmented realities. Users can use forms, images, video, or visualized graphics to depict their stories, profiles, and imaginary selves. Every user can design his/her artwork to represent his/herself. All of this is under computer control and interactive with the servers (Fig. 4).

2.2.3 Design Component and Back-End System

The back-end development focuses on the “server side” of programming, where the connections between the server and the database are constructed. The Init2Winit

Fig. 4 Players' computer-generated characters during the gameplay

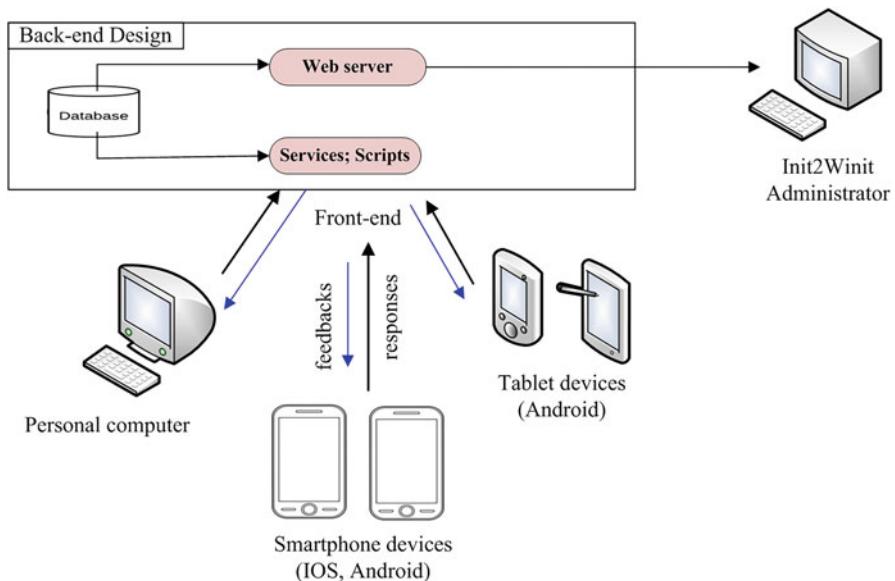
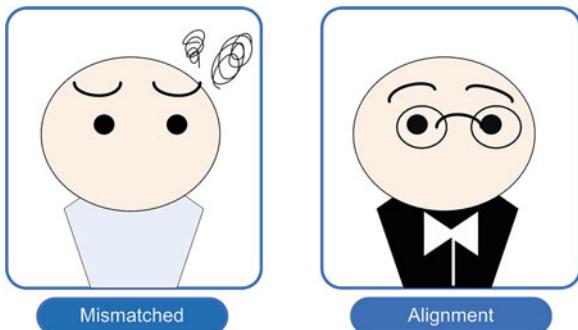


Fig. 5 Init2Winit system architecture

system architecture consists of the following components (Fig. 5): server-side computer system, web application, and mobile device users (including Android/ IOS application for smartphone and tablet). The operating system is a centralized data model which acts as a data hub that interacts with users and conducts data processing between the database and game mechanics as a set of rules and algorithms that guide the outcome of the user's interface interactions. The server-side computer system includes a relational database, user profile, web application, and services for communication with users or for retrieving users' previous records. Those four parts work together to allow for mega data storage and administration for both users and app administrators.

3 Opportunity for AI and Machine Learning (ML)

A broad definition of AI describes a computerized system which “... performs cognitive tasks, usually associated with human minds, particularly learning and problem-solving (Baker et al. 2019: p. 10).” AI and machine learning often refer to similar function as machine learning is a subset of AI, but they are not the same. Modern machine learning models have three types: (1) Supervised machine learning (ML) algorithms based upon existing labeled data or collected information to form a decision, recognizing a pattern, or predicting an outcome. For example, supervised ML can be used to predict dropping out from high school or a high rating score on a writing assignment. (2) Unsupervised classification and profiling are used to sort, identify, and filter unlabeled data based on structures, attributes, features, and densities of resolution. For example, unsupervised ML can be used for customer segmentation or to give recommendations on merchandise. (3) Semi-supervised ML classifies some of the unlabeled/ unidentified information along with labeled and categorized data. For example, semi-supervised ML can be used to classify and organize data, such as sorting writing assignments or job applications into a certain order.

In our case, Init2Winit app could design a function that can be easily integrated with artificial intelligence (AI) which has a broad multifaceted influence running from machine learning to data-based analytic algorithms. The algorithms can create a data feedback system and information loops that allow users to make choices and receive points for identifying correct answers, responses, and task values. The information that In2Winit feeds into the computational game program is based on several national databases. For example, students are asked to select an occupation to pursue, and then, the type of college and majors that they would have to attend to align with this goal in the “career tunnel.” The information on what types of degrees or certificates are needed for various occupations is derived from the Occupational Information Network (National Center for O*NET Department 2019), an occupational and STEM knowledge database that contains 974 occupation descriptions and a mix of required knowledge, education, skills, and abilities for each “person–occupation fit” choice.

The Init2Winit app with AI-enabled function could collect real-time information and misalignment patterns of students’ knowledge. This misinformation could incite a tool similar to an alarm system which alerts additional assistance and guidance by school counselors or the students’ own profiling. An AI-enabled function could also adopt adaptive job-specific or major-specific assessment by adjusting level of difficulty, number of questions, and crucial steps of reaching college-going eligibility and requirements. The Init2Winit app with AI features can also identify student usage behavior, knowledge profiles, and patterns, which can be used to train the machine to adjust the database, and further improve users’ personalized decision-making process (Sarker et al. 2019; Bashier et al. 2016).

3.1 Machine Learning and Decision Trees

The following explains how our small-scale pilot study on the Init2Winit prototype was used to understand students' college and career alignment. A small sample of 157, 10th to 12th graders volunteered to participate in the College Ambition Program (CAP). Two schools designed to assist upper secondary students find less costly, prestigious colleges that fit their academic and career interests. During the CAP program, the students completed a pre- and post-survey with valid app user records. Most users are 11th graders, minority, male with GPA ranging between 2.5 and 3.0, and have parents with less than a college education. The Daily Active Users (DAU) shows the frequency of records per user account of those who had at least one play of Init2Winit during 3 weeks of the prototype testing in 2019 (See Appendix A).

There are several algorithms that can be embedded in the operating system with regards to ML, such as linear regression, neural networks, logistic regression, random forest, decision trees, and support vector machines (SVMs). Decision trees are a type of supervised machine learning and can be divided into two major elements, decision nodes, and leaves. The leaves indicate the outcomes of a decision, and the nodes indicate a branch where the data is split. A simple example of a decision tree is to show how a tree grows in a binary regression. The decision nodes are a series of questions like "What major would you like to attend?", "What type of college would you like to attend?" "What do you think your beginning salary should be?". The leaves show the outcomes like "matched" or "mismatched." In the Init2Winit example, we can consider "matched" as a simple binary yes/no classification answer or a continuous classification answer that indicates the distance between desired goal and predictively matched goal.

3.2 Empirical Example: Decision Trees Algorithm in Init2Winit

Using our Init2Winit users as an example, Fig. 6 lays out the decision tree for predicting whether a student's career goal matched their college planning process given the information they obtained and whether their gameplay indicates a matched college degree and/or annual salary projection for the career they plan to pursue. The first decision test was based on the types of college students expect to attend. The sample included the 157 student users in the first job play as an example, 66% had matched college-going planning and 34% were mismatched. The second decision test identified accurate career knowledge of the annual salary in the targeted job. Here we tested the limited node (e.g., focusing on the nodes in the second decision test only) of the aligned college planners, 67% had a matched salary projection

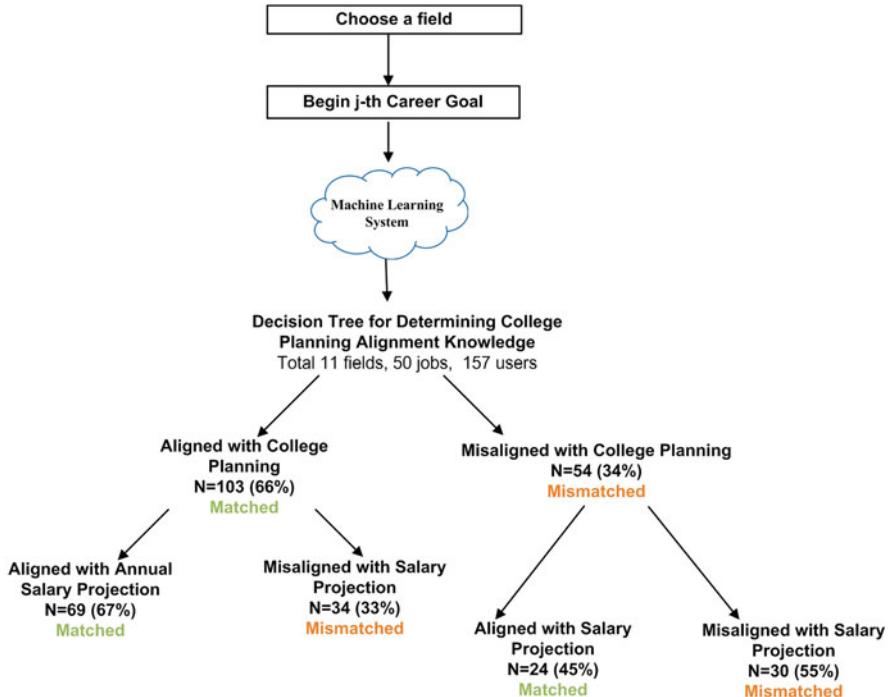


Fig. 6 Hypothetical decision trees using Init2Winit user data in the first job

and 33% were mismatched. On the contrary, when testing the limited node of the misaligned college planners, only 45% had a matched salary projection and 55% were mismatched. This result reflected the fact that users with misalignment knowledge in their college planning had a higher likelihood of having a wrong salary projection as well (55% versus 32%, $Z = 2.67$, $p = 0.0078$).

The decision tree method provides a predictive model in data exploration and training set for machine learning. Our goal is to create a system that models the value of target variables at the leaf of the tree based upon several input variables, including individual users' attributes, at the nodes of the tree. The decision trees in this study aim to identify the probability of a certain alignment results given a desired career choice. This method can also be used for classification and regression. There are several algorithms for decision trees, such as C4.5 (Quinlan 1993), CART (Breiman 2017), BehavDT (Sarker et al. 2019), and IntrudTree (Sarker et al. 2020a). In our example and in our prototype design, we use Iterative Dichotomiser 3 (ID3) algorithm and classification (James et al. 2013; details see Appendix B).

4 Result

4.1 *Init2Winit Users' Profiles*

Before we used the decision trees to predict users' attributes, we first explored user behavior to obtain prior known classified groups (Sarker 2019). We trained our ML model to be close to the reality of the users' behavior and their intention of exploring career-college planning pathways (Sarker et al. 2019, 2020b). To obtain some prior known classified group, we first looked at behavioral patterns of users' career goal-oriented responses in 3 weeks of playing. We restructured the activity record data into a user-specific data by generating indicators to represent the percent of play frequency in each career field (total 11 fields).

Our data shows that there are three patterns of behavioral career explorations. We named them as solo-goal explorers ($N = 67$), dual-goal explorers ($N = 46$), and multiple-goal explorers ($N = 44$). Solo-goal explorers only explored "one" career field and more than 80% of playing activities happened within one specific field. The top five career explorations for solo-goal users are 22% in Science and Technology careers, 20% in Health care careers, 20% in Business careers, 7% in Sport and Athletics, and another 7% in Media-related careers.

The dual-goal explorers choose only "two" career fields and nearly equal percentages of playing activities occurred between the two fields. For example, Kelly plays Init2Winit 12 times. Among those 12 times of plays, Kelly explores 50% (6 times) of career options in the Business field and another 50% (6 times) of career options in the Science and Technology. The top 3 of college planning and career exploration for dual-goal users are 8% in Business and Sport and Athletics careers, 8% in both Science and Technology and Transportation careers, and another 8% in both Science and Technology and Health care careers.

4.2 *Init2Winit Users' Classification for Multiple Goals*

The third pattern is the multiple-goal explorers, who explored "more than two" fields of career options. To allow multiple-goals users to explore nonexclusively career goals across 11 fields, we employ multi-label classification method to help classify their orientation in the training set of data. The multi-label classification can identify the association with several classes or labels, which could support mutually exclusive and nonexclusive classes or labels (Bashier et al. 2016; Hall et al. 2016).

Using 676 records in the data streams from 44 users, we built a classification model. After this multi-label classification, three classifications were identified and named: (1) Multiple field 1: Business, Media, and Healthcare ($n = 35$); (2) Multiple field 2: Education, Media, and Sports ($n = 3$); (3) Multiple field 3: Law, Healthcare, and Science Technology ($n = 6$) in Fig. 7. The models and model performance were examined for each classification.

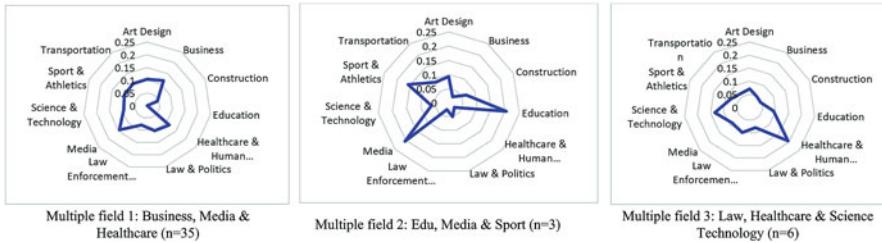


Fig. 7 User profiling results of the multi-label classification for multiple-goal explorers

Table 1 shows the descriptive statistics for the behavioral classifications. Parametric *t*-test and *z*-test are used to compare the means of two independent samples. In our case, we compare all subgroups with solo-goal explorers. Solo-goal explorers play the Init2Winit about 2 times with an average student GPA of 2.86. This group of explorers also has the highest percent of full alignment knowledge (56%) relative to other explorers (49% or 54%). Dual-goal explorers on average play the Init2Winit about 3 times with an average GPA of 2.81. As Table 1 shows, dual-goal and multiple-field explorers played Init2Winit more frequently than solo-goal users.

Multiple field 1 includes 35 users who mostly explored careers in the Business, Media, and Health fields, with the approximate proportion of playing in each field being 0.12, 0.12, and 0.14. This group of explorers also show interests in Art design, Law, and Sport and Athletics careers. Multiple field 2 includes only three users who mostly explored careers in Education, Media, and Sport and Athletics fields. The approximate proportion of playing in those field is 0.20, 0.20, and 0.16. On average, multiple field 1 explorers play Init2Winit 6 times and multiple field 2 explorers play Init2Winit 8 times. Multiple field 1 explorers play Init2Winit significantly more than sole-goal explorers. Multiple field 3 includes only 6 users who mostly explored careers in Law, Health care, and Science Technology. Students in this group of explorers have significantly higher GPA than solo-goal users ($M = 3.55$ versus $M = 2.86$, $p < 0.05$). Additionally, multiple field 3 explorers have relatively higher number of times played, percent of full alignment knowledge, and level of parents' education compared to other classifications.

4.3 Alignment Knowledge of Decision Trees and Partition

Before applying the tree-based prediction model, we explored the relationship between alignment knowledge and educational expectations after playing Init2Winit (using educational expectations in spring) by partitioning the three behavioral patterns and five career goal-oriented patterns. Due to the small sample size of multiple-field classification, we only report the partition results using the three behavioral patterns. In Fig. 8a, blue dots represent solo-goal explorers, pink dots represent dual-goal explorers, and green dots represent multiple-goal explorers.

Table 1 Descriptive statistics across user profiles in classification (five patterns)

	Solo-goal explorers (n = 67)	Dual-goal explorers (n = 46)	Multiple field: Business, Media and Healthcare explorers (n = 35)	Multiple field: Edu, Media and Sport explorers (n = 3)	Multiple field: Law, Healthcare and Science Technology explorers (n = 6)
	Mean/SD	Mean/SD	Mean/SD	Mean/SD	Mean/SD
Total N of play ^a	1.97 (1.37)	2.89*** (1.08)	6.69*** (4.28)	8.00*** (3.46)	69.67*** (46.39)
Student GPA	2.86 (0.79)	2.81 (0.81)	2.96 (0.73)	2.79 (0.86)	3.55* (0.32)
Parent education	2.18 (1.28)	2.17 (1.52)	2.26 (1.26)	2.00 (1.00)	2.50 (1.00)
Educational expectation in spring ^b	5.25 (1.32)	4.92 (1.57)	5.35 (1.29)	5.33 (0.57)	5.50 (1.00)
Female ^c	%	%	%	%	%
Female ^c	0.46	0.33	0.48	0.33	0.50
White	0.20	0.19	0.14	0.33	0.25
Black	0.07	0.10	0.07	0.33	0.25
Asian	0.22	0.36	0.31	0.00	0.25
Others	0.50	0.36	0.48	0.33	0.25*
Percent of no alignment	0.15	0.13	0.09	0.15	0.03
Percent of misalignment	0.29	0.38	0.37	0.29	0.24
Percent of full alignment	0.56	0.49	0.54	0.56	0.74
4-year college planning (yes/no)	0.65	0.60	0.66	0.67	0.50

Note: ^aFor continuous variables, two-tailed t-tests compares each classification groups with solo-goal explorers. ^{*p < 0.05 **p < 0.01 ***p < 0.001}

^bEducational expectations were measured by students' response to the question, "How far in school do you think you'll get?" in a survey administered in the spring semester of the 2018–2019 after playing Init2WinIt. The scale ranges from 1 (less than high school completion) to 7 (complete a Ph.D., M.D., law degree, or other high-level professional degree). A higher value indicated students' higher educational expectations

^cFor categorical variable, two proportion z-test compares each classification group with solo-goal explorers. ^{*p < 0.5}

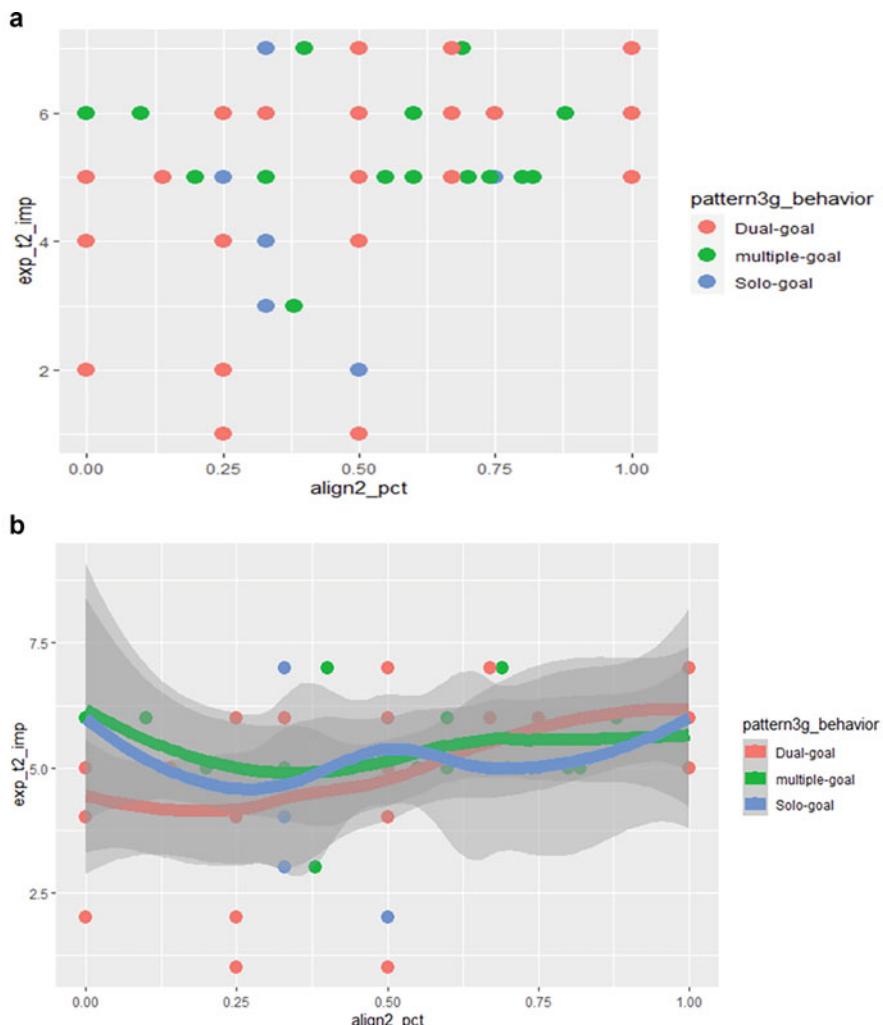


Fig. 8 (a) Partition results of three behavioral conditions: Percent of full alignment playing by educational expectations in spring. (b) Smoothing partition results of three behavioral conditions: Percent of full alignment playing by educational expectations in spring

The X-coordinate represents the percent of full alignment from the period of playing for 3 weeks, and the Y-coordinate represents users' level of educational expectations in spring. We assume *Init2Winit* users gain more alignment knowledge during the play, which in turn increases students' educational expectations in spring. We find that solo-goal explorers concentrate in the left middle of the partition space. The linear tendency is low and only happens in the middle level of alignment knowledge and expectations (expectation = 5, percent of alignment = 0.5). Most

multiple-goal explorers have relatively higher educational expectations in spring, and the linear tendency is moderate in the upper-right panel of the partition space (expectation > 5, percent of alignment > 0.5). Dual-goal explorers show more variation on the partition space, and the linear tendency is more robust and more responsive to the percent of full alignment knowledge in Fig. 8b. After viewing the partition plot above, we conclude that a regression decision tree is the more appropriate method to estimate our current sample.

4.4 Regression Decision Trees and Prediction of Educational Expectations

We then build a regression decision tree using four college-planning and salary-prediction questions in the first two gameplays to predict educational expectations in spring. The results of the regression decision tree have seven terminal nodes as shown in Fig. 9. Each node shows the predicted educational expectations of Init2Winit player in the growing trees and the number of observations from the training dataset located at that node in Table 2.

At the top of Fig. 9, the predicted educational expectations of the overall sample is 5.1. We have 92 users with completed alignment knowledge records in both the first and second careers. The first node asks whether the college planning matched with the first job goal is equal to 0. If no, then the users go down to the right node. The second node asks whether the college planning matched with the second job goal. If no, then the users go down to another right node. If the users have alignment knowledge on those two nodes, then the predicted educational expectations are 5.5 (ranged between a 4-year college degree and a master's degree). In this tree-based

Fig. 9 Decision tree of the alignment knowledge predicted educational expectation in spring

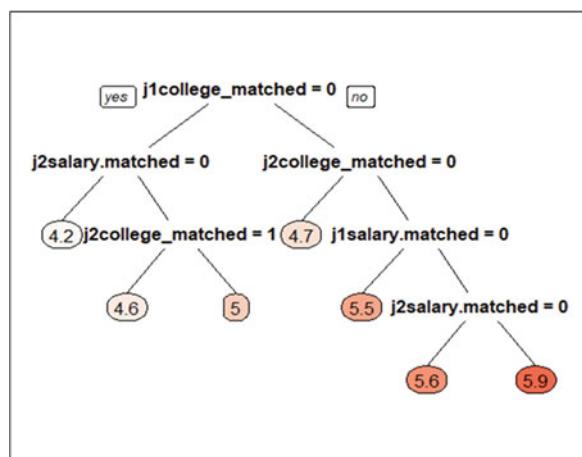


Table 2 Decision tree predicted rules, predicted expectations, and percent of sample

Percent of sample	Predicted educational expectation in spring	Prediction rule
14% (<i>n</i> = 13)	4.2	j1college_matched = 0 and j2salary.matched = 0
15% (<i>n</i> = 14)	4.6	j1college_matched = 0 and j2college_matched = 1 and j2salary.matched = 1
12% (<i>n</i> = 11)	4.7	j1college_matched = 1 and j2college_matched = 0
8% (<i>n</i> = 7)	5.0	j1college_matched = 0 and j2college_matched = 0 and j2salary.matched = 1
21% (<i>n</i> = 19)	5.5	j1college_matched = 1 and j2college_matched = 1 and j1salary.matched = 0
9% (<i>n</i> = 8)	5.6	j1college_matched = 1 and j2college_matched = 1 and j2salary.matched = 0 and j1salary.matched = 1
22% (<i>n</i> = 20)	5.9	j1college_matched = 1 and j2college_matched = 1 and j2salary.matched = 1 and j1salary.matched = 1

Note: Bold indicates an example we described in the main text

model, 19 users belong to this pathway. If the users did not have a matched college planning knowledge in the second job goal, the predicted educational expectations are 4.7 (ranged between some college and a 4-year college degree). We have 11 users who belong to this pathway. Our tree could grow and help us understand which primary alignment knowledge (college planning or salary prediction) impacts educational expectations prediction more.

To evaluate the prediction performance of the tree-based model, we split the current sample randomly by an 8:2 ratio into the training and testing sets. Then, we train our model on the training set and tested it. We used the averaged F1-score to measure the overall performance of the algorithm (Lipton et al. 2014). The F1 score is a weighted average of the precision rate for recall. The range of an F1 score is 0–1. Our current model has a F1 score of 0.72 using four college planning to salary prediction questions in the first two jobs of gameplay. We can increase the prediction performance to 0.85 by including more variables and questions, such as GPA, parent education, and students' characteristics. We report the simplest results in current study because the inclusion of more variables in the tree-based model also increases the number of missing cases (other decision trees results are available upon request).

5 Strengths and Weakness of Current Design

One of the strengths of the current design is the simplicity of the design and the effectiveness. The simplicity is the increase in students' alignment by playing the career exploration tunnel in the Init2Winit. The effectiveness is in predicting how student alignment knowledge corresponds to their educational expectations after game playing through the use of a decision tree. Using this prediction, the importance of increasing students' alignment knowledge and leading to increasing educational expectations after game playing becomes clear. Importantly, this prediction does not require a lot of users' background information or covariates but can still provide valuable data insights with a high prediction level. This feature is very useful with data where background data is not available or where there is over 10% of missing data.

Additionally, embedding the machine learning and decision tree algorithm in a mobile application is also quite useful with respect to users becoming more informed by the optimization students' college planning or forecasting the success rates for various career goals. Users' behavior patterns and goal-oriented explorations can also profile the individual's motivation and preparedness based upon a predetermined classification analysis. However, this design also leaves several open questions surrounding the factors which drive students' misalignment in their career/college knowledge, how to distinguish higher scorers between playing within the same career options versus playing across multiple career options, and the genuine learners of alignment.

The decision tree, as one of the simplest ML models, could incorporate several different functions to account for complex data structure and conditions, such as boosting when there is high variance in the outcomes. However, this method also has some limitations. First, decision trees are less efficient in estimation compared to other supervised ML methods, especially in big trees where increasing efficiency results in poor prediction accuracy (James et al. 2013). Second, large decision tree models cause high complexity in processing the data, increasing computation time, and difficulties in converging. More advanced methods, such as random forest, neural network, and support vector machines (SVMs), can be more computationally effective and handle nonlinear patterns and large samples (Puterman 2014). Third, the prediction of decision trees generally does not have comparable accuracy rate to other approaches, especially in a small sample (Wu et al. 2016).

6 Conclusion and Recommendation

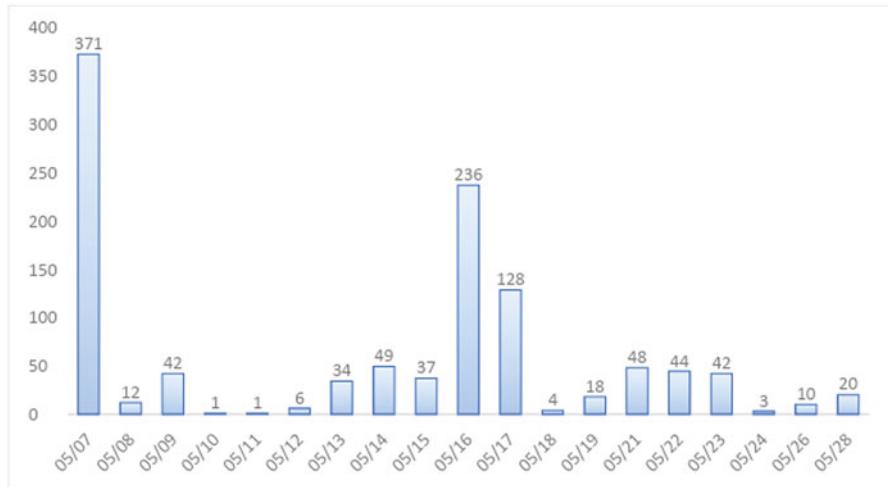
This study develops and tests the AI features of machine learning in Init2Winit, using the decision tree-based method, to identify users' usage behavior, goal-oriented patterns, and prediction of future educational expectations. Our results

show promise in terms of the prediction accuracy of educational expectations and users' behavioral classifications. Beyond this, machine learning could incorporate a game designed to measure students' strengths and weaknesses to give career recommendations and pathways. Init2winit can be an informational channel for low-income students who lack informal networks or whose parents have not earned college degrees. It also serves as a supplementary network supporting career/ college planning knowledge for students to make better education and employment decisions. This study is just one example of how AI and machine learning can help students explore careers and increase their educational aspirations and college-going choices. It shows how a mobile application can be built upon previous theory (alignment theory) to increase students' knowledge and educational expectations and to further flag students who may be mismatched, misaligned, or disoriented in their planning and decision-making for college and career choice.

The study has three primary goals, each of which informs the alignment theory of career-to-college explorations and applies efforts to strengthen the pipeline of STEM careers during high schools. First, we develop a mobile application Init2Winit to test theoretical assumptions about alignment knowledge. Second, we compare students' goal exploration behavior, orientation, and profile, which are important in shaping career choices and college decisions. Third, we provide data insights for school counselors, parents, and students to optimize their choices and college plans. Altogether, our study evaluates and recommends an outlook of Init2Winit in the coming decades.

We propose a few steps that should be considered to ensure that all students are served and provided with the information and social capital needed for college readiness and planning. The first suggestion is to consider the ways in which school counselors and homeroom teachers serve as role models and informational hubs in the lives of many students through the use of mobile technology and its applications. Teachers' participation can facilitate parents and students' knowledge, using machine learning to improve users' personalized decision-making (Thompson and Subich 2006). Another suggestion is to provide students with a real-time intervention and guidance even in resources-restricted schools. Educational technology can provide unlimited access to information and data feedback based on student usage behavior, goal-oriented profiles, and response patterns. Fundamentally, our goal is to use AI technology to formulate more realistic engaging tasks and scoring procedures that can provide improved college knowledge and career aspiration for students, their parents, and school professionals. The goal here is efficiency but not at the expense of students' interests or in trying to force career choice too early in a young person's life.

Appendix A Record of e Daily Active Users



Note: The numbers in each bar represent the total number of individual users per day.

Appendix B Iterative Dichotomiser 3 (ID3) Algorithm

The ID3 algorithm uses the most significant information gain after splitting the measure to partition the outcome and make each branch belong to the same classification. The criteria to separate the node is the Gini impurity and “entropy” for the information gain. Entropy measures the discriminatory power of an attribute in the classification task. It defines the amount of randomness in the attribution of classification or regression. Gini and Gini impurity are used to decide the best split. Gini ranges from 0–1. The higher the Gini coefficient, the more different instances within the node.

$$\text{Entropy} : H(S) = - \sum_{i=1}^n p(x_i) \log_2 p(x_i) \quad (1)$$

$$\text{Gini } (E) = 1 - \sum_{i=1}^c p_i^2 \quad (2)$$

Information gain defines as a set of S, which are effective changes in entropy after deciding on a particular attribute or goal. Information gain measures the relative changes in entropy conditional on the independent variables in the tree. A training set S could be a positive or a negative example. The indicates the probability of event x. Our goal is to use this method to train the machine to classify users' response patterns and provide predictive data insights for students and school counselors.

References

- Arntzi M, Gregory T, Zierahni U. (2016). The Risk of Automation for Jobs in OECD Countries: A Comparative Analysis. Paris: OECD.
- Albion MJ, Fogarty GJ. (2002). Factors Influencing Career Decision Making in Adolescents and Adults. *Journal of Career Assessment*, 10(1), 91–126. doi: <https://doi.org/10.1177/1069072702010001006>
- Baker, T., Smith, L., & Anissa, N. (2019). Educ-AI-tion rebooted? Exploring the future of artificial intelligence in schools and colleges. Retrieved from nesta.org.uk website: https://media.nesta.org.uk/documents/Future_of_AI_and_education_v5_WEB.pdf
- Bashier, E. B. M., Mohammed, M., & Khan, M. B. (2016). Machine learning: algorithms and applications. CRC Press.
- Bettinger, E. P., Long, B. T., Oreopoulos, P., & Sanbonmatsu, L. (2012). The role of application assistance and information in college decisions: Results from the H&R Block FAFSA experiment. *The Quarterly Journal of Economics*, 127(3), 1205–1242. doi: <https://doi.org/10.1093/qje/qjs017>
- Blustein, D. L., Duffy, R., Ferreira, J. A., Cohen-Scali, V., Cinamon, R. G., & Allan, B. A. (2020a). Unemployment in the time of COVID-19: A research agenda. *Journal of Vocational Behavior*, 119, 103436. doi: <https://doi.org/10.1016/j.jvb.2020.103436>
- Blustein, D. L., Erby, W., Meerkens, T., Soldz, I., & Ezema, G. N. (2020b). A Critical Exploration of Assumptions Underlying STEM Career Development. *Journal of Career Development*, 0894845320974449. doi: <https://doi.org/10.1177/0894845320974449>
- Breiman, L. (2017). Classification and regression trees: CRC Press.
- Bureau of Labor Statistics. Occupational employment projections 2014-2024. Monthly Labor Review. 2015. <https://www.bls.gov/opub/mlr/2015/article/occupational-employment-projections-to-2024.htm>
- Castleman, B., & Goodman, J. (2018). Intensive College Counseling and the Enrollment and Persistence of Low-Income Students. *Education Finance and Policy*, 13(1), 19–41. doi: https://doi.org/10.1162/edfp_a_00204
- Chen, I. C., Rocha-Beverly, C., & Schneider, B. (2020). Alignment of educational aspirations and career plans in high school with Mobile app technology. *Education and Information Technologies*, 1–19
- Chen, I. C., Rocha-Beverly, C., & Schneider, B. (2021) Learning by playing Init2Win: How alignment knowledge increases educational aspirations and college plans in high school, *Journal of Research on Technology in Education*, doi: <https://doi.org/10.1080/15391523.2021.1877225>
- Clark, B., Joubert, C., & Maurel, A. (2017). The career prospects of overeducated Americans. *IZA Journal of Labor Economics*, 6(1), 3. doi: <https://doi.org/10.1186/s40172-017-0053-4>
- Cohodes, S. R., & Goodman, J. S. (2014). Merit Aid, College Quality, and College Completion: Massachusetts' Adams Scholarship as an In-Kind Subsidy. *American Economic Journal: Applied Economics*, 6(4), 251–285. doi: <https://doi.org/10.1257/app.6.4.251>
- James, G M., Witten, D., Hastie, T., & Tibshirani, R. R. Tree-Based Methods. In: Casella G FS, Olkin I, ed. An introduction to statistical learning – with applications in R. New York: Springer; 2013:373–418.
- Hall, M. A., Witten, I. H., Frank, E., & Pal, C. (2016). Data Mining: Practical machine learning tools and techniques. San Francisco: Morgan Kaufmann.
- Hoff, K., Van Egdom, D., Napolitano, C., Hanna, A., & Rounds, J. (2021). Dream Jobs and Employment Realities: How Adolescents' Career Aspirations Compare to Labor Demands and Automation Risks. *Journal of Career Assessment*. doi: <https://doi.org/10.1177/10690727211026183>
- Kena G, Hussar W, McFarland J, Brey C, Musu-Gillette L. (2016). The condition of education 2016 (NCES 2016-144). Washington, DC: U.S. Department of Education, National Center for Education Statistics. 2016. Retrieved from <https://nces.ed.gov/pubs2016/2016144.pdf>

- Kim, S., Klager, C., & Schneider, B. (2019). The Effects of Alignment of Educational Expectations and Occupational Aspirations on Labor Market Outcomes: Evidence from NLSY79. *The Journal of Higher Education*, 90(6), 992–1015. doi: <https://doi.org/10.1080/00221546.2019.1615333>
- Lipton Z.C., Elkan C., Narayanawamy B. Optimal Thresholding of Classifiers to Maximize F1 Measure. In: Calders T., Esposito F., Hüllermeier E., Meo R. (eds) Machine Learning and Knowledge Discovery in Databases. 2014. ECML PKDD. Lecture Notes in Computer Science, vol 8725. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-662-44851-9_15
- Mann, A., Denis, V., Schleicher, A., Ekhtiari, H., Forsyth, T., Liu, E., & Chambers, N. (2020). Teenagers' career aspirations and the future of work: OECD.
- Morgan, S. L., Leenman, T. S., Todd, J. J., & Weeden, K. A. (2013). Occupational Plans, Beliefs about Educational Requirements, and Patterns of College Entry. *Sociology of Education*, 86(3), 197–217. doi: <https://doi.org/10.1177/0038040712456559>
- National Center for O*NET Development. (2019). Job zone. O*NET OnLine. Retrieved from <https://www.onetonline.org/find/zone>
- Nedelkoska, L., & Quintini, G. (2018). Automation, skills use and training OECD.
- Nie, M., Xiong, Z., Zhong, R., Deng, W., & Yang, G. (2020). Career Choice Prediction Based on Campus Big Data—Mining the Potential Behavior of College Students. *Applied Sciences*, 10(8). doi: <https://doi.org/10.3390/app10082841>
- Nunley, J. M., Pugh, A., Romero, N., & Seals, R. A., Jr. (2016). College major, internship experience, and employment opportunities: Estimates from a resume audit. *LABOUR ECONOMICS*, 38, 37–46. doi: <https://doi.org/10.1016/j.labeco.2015.11.002>
- Perry, B. L., Martinez, E., Morris, E., Link, T. C., & Leukefeld, C. (2016). Misalignment of Career and Educational Aspirations in Middle School: Differences across Race, Ethnicity, and Socioeconomic Status. *Social sciences* (Basel, Switzerland), 5(3). doi: <https://doi.org/10.3390/socsci5030035>
- Pinto, L. H., & Ramalheira, D. C. (2017). Perceived employability of business graduates: The effect of academic performance and extracurricular activities. *Journal of Vocational Behavior*, 99, 165–178. doi: <https://doi.org/10.1016/j.jvb.2017.01.005>
- Puterman, M.L. (2014). *Markov decision processes: discrete stochastic dynamic programming*. John Wiley & Sons.
- Quinlan, J.R. (1993). C4.5: programs for machine learning. *Mach Learn.*
- Reyna, V. F., & Farley, F. (2006). Risk and Rationality in Adolescent Decision Making. *Psychological Science in the Public Interest*, 7, 1–44.
- Renbarger, R., & Long, K. (2019). Interventions for Postsecondary Success for Low-Income and High-Potential Students: A Systematic Review. *JOURNAL OF ADVANCED ACADEMICS*, 30(2), 178–202. doi: <https://doi.org/10.1177/1932202X19828744>
- Sarker, I. H. (2019). A machine learning based robust prediction model for real-life mobile phone data. *Internet Of Things*, 5, 180–193. doi: <https://doi.org/10.1016/j.iot.2019.01.007>
- Sarker, I. H., Colman, A., Han, J., Khan, A. I., Abushark, Y. B., & Salah, K. (2020a). BehavDT: A Behavioral Decision Tree Learning to Build User-Centric Context-Aware Predictive Model. *Mobile Networks & Applications*, 25(3), 1151–1161. doi: <https://doi.org/10.1007/s11036-019-01443-z>
- Sarker, I. H., Alqahtani, H., Alsolami, F., Khan, A. I., Abushark, Y. B., & Siddiqui, M. K. (2020b). Context pre-modeling: an empirical analysis for classification based user-centric context-aware predictive modeling. *Journal Of Big Data*, 7(1). doi: <https://doi.org/10.1186/s40537-020-00328-3>
- Sarker, I. H., Kayes, A. S. M., & Watters, P. (2019). Effectiveness analysis of machine learning classification models for predicting personalized context-aware smartphone usage. *Journal Of Big Data*, 6(1). doi: <https://doi.org/10.1186/s40537-019-0219-y>
- Schumacher, P., Olinsky, A., Quinn, J., & Smith, R. (2010). A Comparison of Logistic Regression, Neural Networks, and Classification Trees Predicting Success of Actuarial Students. *Journal of Education for Business*, 85(5), 258–263

- Schmitt-Wilson, S., & Faas, C. (2016). Alignment of Educational and Occupational Expectations Influences on Young Adult Educational Attainment, Income, and Underemployment. *SOCIAL SCIENCE QUARTERLY*, 97(5), 1174–1188. doi: <https://doi.org/10.1111/squ.12244>
- Schneider, B., & Stevenson, D. (1999). *The Ambitious Generation: America's Teenagers, Motivated but Directionless*. New York: Yale University Press.
- Schneider, B. (2009). Challenges of transitioning into adulthood. In I. S. R. K. Silbereisen (Ed.), *Transitions from school to work: Globalization, individualization, and patterns of diversity* (pp. 265–290). Cambridge, England: Cambridge University Press.
- Schneider, B., & Young, L. (2019). Advancing workforce readiness among low-income and minority high school students. In F. L. Oswald, Behrend, T. S., & Foster, L. L. (Ed.), *Workforce readiness and the future of work*. New York: Taylor & Francis.
- Thompson, M. N., & Subich, L. M. (2006). The relation of social status to the career decision-making process. *Journal of Vocational Behavior*, 69(2), 289–301. doi: <https://doi.org/10.1016/j.jvb.2006.04.008>
- Wu, C.-C., Chen, Y.-L., Liu, Y.-H., & Yang, X.-Y. (2016). Decision tree induction with a constrained number of leaf nodes. *Applied Intelligence*, 45(3), 673–685. doi: <https://doi.org/10.1007/s10489-016-0785-z>

Open Access This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.



Learning Clinical Reasoning Through Gaming in Nursing Education: Future Scenarios of Game Metrics and Artificial Intelligence



Jaana-Maija Koivisto, Sara Havola, Henna Mäkinen, and Elina Haavisto

Contents

1 Introduction	159
2 AI in Healthcare Education	161
3 Exploiting AI Through Game Metrics	162
4 A Case Study of CR and the Use of Game Metrics in Nursing Simulation Games.....	164
5 Directions for Future Work	168
6 Conclusion	169
References	170

1 Introduction

The COVID-19 pandemic has challenged clinical practices, quality of care, and patient safety because of the uncertainties related to the virus itself and patients' clinical conditions, which deteriorate very suddenly. Working in such stressful and rapidly changing clinical situations challenges professionals' clinical reasoning (CR) (Audétat et al. 2020). CR is a complex cognitive process by which professionals use formal and informal thinking strategies to gather and analyze patient information, and evaluate the significance of this information and reflect on alternative actions. CR is one of the most essential competence areas in

J.-M. Koivisto (✉)

Tampere University, Tampere, Finland

Häme University of Applied Sciences, Hämeenlinna, Finland

e-mail: jaana-maija.koivisto@hamk.fi

S. Havola

Tampere University, Tampere, Finland

H. Mäkinen

Häme University of Applied Sciences, Hämeenlinna, Finland

E. Haavisto

Tampere University, Tampere, Finland

Tampere University Hospital, Tampere, Finland

clinical care (Hunter and Arthur 2016; European Parliament 2013). Good CR skills ensure patient safety (Mawhirter and Garofalo 2017), whereas incomplete CR skills are related to poor decision-making and even poor patient outcomes (Holder 2018; Simmons 2010). Clinical problems with COVID-19 patients have been ill-defined, and therefore decision-making may change from day to day and thus lead to errors in CR (Audéat et al. 2020). These unfortunate mistakes can be fatal for patients but also traumatic for healthcare professionals. The COVID-19 pandemic has highlighted the importance of professionals' CR skills and thus challenged organizations to consider methods for developing these crucial skills. Artificial intelligence (AI) is one solution for ensuring quality decision-making in challenging situations. Machine learning (ML), deep learning (DL), and natural language processing (NLP) methods can support healthcare professionals' clinical decisions. AI can also be used to support learning CR in healthcare professionals' education and training. Yet studies show that AI use in healthcare education is limited (Randhawa and Jackson 2020), whereas the use of technology utilizing immersive learning environments has increased. In medical education, AI has been used to an increasing extent recently (Sapci and Sapci 2020) but less so in nursing education (Randhawa and Jackson 2020).

CR skills play a major role in identifying and preventing the deterioration of patients' clinical conditions. The special focus of this chapter is nursing simulation games intended for that purpose. Although the use of AI is still limited in nursing education, there exists a positive attitude toward AI (Buchanan et al. 2021); thus, innovations in the field of AI are likely to be seen there in the near future. One potential area of application for AI is simulation games that automatically adjust to the player's abilities and needs. Game metrics could be used to develop adaptive features for educational games. The adaptivity of the content can be achieved by applying techniques from the field of AI, such as dynamic difficulty adoption (Streicher and Smeddinck 2016). Adaptivity refers to the ability of the system to identify the user's preferences or characteristics and customize the system accordingly by analyzing users' previous interactions with the system before making an automatic adjustment (Soflano et al. 2015).

The rapid development of technology has enabled the adoption of diverse types of simulation games in different areas of healthcare education, providing new ways to learn for various learners (McEnroe-Petitte and Farris 2020). These new approaches can offer opportunities for traditional and distance education in healthcare education. Simulation games promote motivation and improve problem-solving (Chang et al. 2020). For instance, simulation games have been used to prepare healthcare students for clinical practices or unexpected situations as well as to support maintaining skills (e.g., Besse et al. 2020; Breedt and Labuschagne 2019). Learning by playing simulation games is also fun and engaging. Engagement can be promoted by creating different and interesting scenarios (Ferguson et al. 2015). Previous studies have revealed that attitudes toward learning with simulation games are mainly positive (e.g., Foronda et al. 2020). CR skills are needed in clinical practice, and therefore, it is important to practice these crucial skills even before encountering patients in real life to avoid patient harm (Peddle et al. 2019).

The purpose of this chapter is to discuss the potential of exploiting AI through game metrics in nursing education for learning CR skills. The next section describes some examples of using simulation games in learning in healthcare education. Thereafter, the current state of using AI in healthcare education is discussed. The possibilities of leveraging game metrics in developing adaptive features for nursing simulation games are then examined. A case study of game metrics in nursing simulation games is presented, and finally, directions for further work are suggested.

2 AI in Healthcare Education

As immersive technologies develop, their use in healthcare education will increase significantly. The immersive technologies in the field of healthcare education include haptic device simulators, computer-based simulations, and head-mounted displays (HMDs), with haptic simulators being the most used and HMD devices the least used (Mäkinen et al. 2020). The use of completely immersive virtual reality (VR) simulations, which are used with HMDs and hand controls or haptics, is still quite rare (Fealy et al. 2019). In nursing education, computer-based simulations are used most often, and they are commonly used to develop clinical decision-making, situation awareness, stress management, and CR skills (Bracq et al. 2019a; Havola et al. 2020). Simulation games have also been used for evaluating nursing students' performance, for example, in resuscitation situations (Keys et al. 2021). Virtual reality simulations have been used to teach teamwork, communication and leadership skills (Bracq et al. 2019b; Kardong-Edgren et al. 2019, Pons Lelardeux et al. 2018), as well as clinical skills, such as urinary catheterization (Butt et al. 2018) and airway management (Botha et al. 2021).

Learners' experiences with using immersive technologies have been positive, and learners have perceived them to be useful in teaching and learning (Botha et al. 2021; Butt et al. 2018). Research has also shown that simulation games are effective learning methods (Chang et al. 2020, Koivisto et al. 2020, Keys et al. 2020, 2021). For instance, nursing students rated their CR skills better after playing a computer-based simulation game than before (Koivisto et al. 2020). Keys et al. (2020, 2021) found that students who played a virtual simulation game performed better in resuscitation situations than students who received traditional preparation. Similarly, in a study by Chang et al. (2020), students who played a simulation game indicated better learning performance, attitude, motivation, and critical thinking than students in the control group, who received only traditional instruction.

Although AI has a long history in healthcare and education, its application is quite limited in the education of healthcare professionals, especially in nursing education (Randhawa and Jackson 2020). In medical education, there have been some advancements in the use of AI. In their systematic review, Sapci and Sapci (2020) evaluated the current state of AI training and the use of AI tools to enhance the learning experience in both medicine and health informatics. AI use includes NLP application to medical education, ML algorithms used for evaluating technical

skills in VR simulators, AI analytics for personalizing the learning process, and AI algorithms for assessing surgical psychomotor skills. Shorey et al. (2019) used AI in nursing education by developing Virtual Patients (VP) with virtual counseling apps integrating AI for teaching communication skills. Google Cloud's Dialogflow NLP engine was used to train a voice chatbot that was visualized as a 3D avatar form using Unity 3D. In testing the application, technological limitations were encountered: the VPs were unable to adapt to the conversational context, the program did not recognize keywords to determine appropriate responses, not all computers or microphones were compatible with the app, and the program had difficulties recognizing some students' pronunciations or speech patterns, resulting in translation failures (Shorey et al. 2019). Such challenges may be overcome as technology advances.

Harmon et al. (2021) conducted a scoping review to explore the use of AI and VR in the context of clinical simulation for pain education in nursing. Only four studies utilizing AI within nursing pain education simulations were found, but the review did not report how AI was utilized in those articles. However, it was seen as playing an important role. A scoping review conducted by Buchanan et al. (2021) summarized the predicted influences of AI health technologies on nursing education. Most of the 27 articles reviewed were expository papers; only seven were empirical studies. The literature review indicated that predictive analytics, smart homes, virtual avatar apps such as chatbots, virtual or augmented reality devices, and robots were expected to have an influence in nursing education. In terms of simulation environments, humanoid robots and cyborgs were seen to complement existing high-fidelity simulators. VP gaming apps and virtual tutor chatbots were predicted to be useful for simulating clinical scenarios, and face tracker software using ML could be used to analyze students' emotions during simulation activities. ML could be used to enhance student engagement by analyzing student data and creating more personalized learning pathways. Furthermore, the use of AI health technologies, such as predictive analytics, could benefit nursing students' transition to clinical practice by improving their clinical judgment and CR skills (Buchanan et al. 2021). These prospects indicate that the use of AI in nursing education could have a positive impact on learning experiences, engagement, and learning outcomes.

3 Exploiting AI Through Game Metrics

First, this section introduces the concept of game metrics and their use in performance evaluation in education. Second, the section considers employing AI in game metrics by developing simulation games that adapt to the player's skill level. In previous studies, different game metrics, such as the number of played games, playing time and scores, have been of interest (Kiili et al. 2018; Hamdaoui et al. 2017; Drachen et al. 2013). Kiili et al. (2018) studied game metrics in assessing students at primary schools and their conceptual rational number knowledge skills. Game metrics consisted, for example, of overall game performance, effective playing time,

maximum level achieved, collected coins, estimation correctness, and the number of played games. In another study, total playtime per player, the number of quests or missions completed, location of the player at each time and interactions with other characters were investigated (Hamdaoui et al. 2017). Kim et al. (2020) investigated learners' behavior while using immersive virtual reality (IVR) applications in vocational education and training by analyzing the time spent, the number of objects placed, and the number of simulations run by the learners. They found that the quality of learning outcomes was positively correlated with the time spent and the number of objects placed in IVR, whereas a number of simulations were negatively correlated with learning outcomes. Soflano et al. (2015), on the other hand, found no correlation between completion time and learning effectiveness, but they found that adaptive game-based learning applications were better at allowing learners to complete the tasks faster than the nonadaptive game versions.

A closer look at different studies using game metrics shows that the definitions of terms differ. When considering the game metrics regarding time, for instance, Kiili et al. (2018) have used the term "effective playing time" to refer to "the summed-up time that a player took to complete all tasks." Hamdaoui et al. (2017), in turn, have used the term "total playing time," which is understood to mean "the sum of the duration of all played levels." They argue that when metrics regarding time have high value, they refer to players' deep immersion in the game. Since the definitions of game metrics differ, it is always necessary to determine the exact definitions of all game metrics in studies. Plass et al. (2013) highlighted that it is essential to know what data are being collected and to determine what is to be measured and why and how the variables are measured.

The use of AI techniques such as personalization and adaptivity in serious games enables meaningful learning experiences and can promote learning, motivation, and user acceptance by responding to the individual needs of the learner (Streicher and Smeddinck 2016). Game metrics could be used to develop adaptive features for nursing simulation games. Simulation games store a large amount of data about the students' game behaviors, including every action the player takes in gameplay, such as answering multiple-choice questions. The game system also stores how much time players spend interacting with different elements of the gaming environment, how many playthroughs they experience, and how many points they earn. Game analytics, learning analytics, and educational data mining enable monitoring interactions between the player and the gaming environment during gameplay and when analyzing usage data (Streicher and Smeddinck 2016). By calculating and analyzing performance according to specific game metrics, it is possible to demonstrate the player's learning, knowledge, and skills (Drachen et al. 2013; Plass et al. 2013). In other words, analyzing game metrics provides the opportunity to have specific data on how the player is engaged in the game (Drachen et al. 2013). Additionally, game metrics can be used to synthesize objective information about the progress of learners related to learning objectives. Game metrics are also essential when evaluating users' experiences (Hamdaoui et al. 2017). When using simulation games to learn CR skills, game metrics reveal how students interact with a VP. Furthermore, game metrics offer a new and objective

way of demonstrating and evaluating nursing students' CR skills (Drachen et al. 2013).

To guarantee efficient learning, simulation games should be able to adapt the gameplay and content of the game individually to all learners (Hamdaoui et al. 2017). An adaptive simulation game can react to learners' prior experiences by offering context-adaptive modifications (Streicher and Smeddinck 2016). One form of adaptivity is adapting the difficulty level of the learning content in simulation games to the current level of the learner based on predefined general parameters or according to a user model. By dynamically adjusting the difficulty level, learners' immersion and state of flow can be fostered. This, in turn, may promote learning outcomes. Adaptivity in learning games can also shorten the completion time of the game (Soflano et al. 2015).

In the initial phase of adaptation, simulation games must implement a performance evaluation to measure certain parameters of the player's performance. This is necessary because, when the player starts the game, the system does not yet have information about the player's skills (Streicher and Smeddinck 2016). Performance evaluation can be done by analyzing the game metrics stored in the game. Game metrics, as parameters, can be used for the classification of players' performances to determine the knowledge or skill levels of the users. Adjustments can be performed based on single or multiple parameters (e.g., game metrics). Difficulty adjustment based on performance may include decreasing the difficulty, not altering the difficulty, or increasing the difficulty (Streicher and Smeddinck 2016). In this case, students' performance in solving simulation game scenarios will respond to their own skill level, which increases motivation. This, in turn, may result in better learning outcomes.

Dynamic adaptive systems in simulation games benefit a heterogeneous group of learners with varying knowledge and skill levels, cultural backgrounds, and previous gaming experience. However, the use of adaptive features in simulation games for learning in a fully automated way in the field of nursing education is still limited, even though AI, including ML and data mining, creates opportunities for developing adaptive systems (Streicher and Smeddinck 2016).

4 A Case Study of CR and the Use of Game Metrics in Nursing Simulation Games

This section describes a case study conducted in Finland (Havola et al. 2021) that used game metrics to evaluate nursing students' scenario performance in simulation games. In this study, playing the simulation game was integrated into the students' studies as one method alongside other teaching methods. Game metrics included the number of playthroughs, the mean score, and the mean playing time.

The validated simulation game was previously developed in cooperation with researchers, nurse educators, nursing students, and game developers, and it has



Fig. 1 Screenshot of the simulation game

become an effective method for learning CR skills (Koivisto et al. 2020). In the game, players are engaged with different clinical situations, such as surgical and emergency settings. In each scenario, the common learning goal is to apply the “Airway, Breathing, Circulation, Disability, Exposure” (ABCDE) approach (Smith and Bowden 2017), which is a validated tool for identifying clinically at-risk patients. By using this approach in the game, a systematic way to assess a patient’s clinical condition can be practiced. A previous study has found that students feel that a simulation game allows for the internalization of different treatment protocols (Koivisto et al. 2017). Scenario-specific learning objectives included, for example, recognizing the symptoms of hypovolemia and knowing the right treatment methods for assessing the patient’s pain and implementing pain management.

The simulation game is a single-player game that can be played on a computer or with a VR headset (Fig. 1). The gaming environment is a 3D hospital environment, including a VP with specific animations indicating the clinical condition of the patient, such as difficulty breathing or chest pain. When gaming, the player takes on the role of a nurse. In every scenario, the player evaluates the patients’ clinical situation, collects and processes information, identifies problems, sets goals and acts in the right order based on the framework of the CR process (Levett-Jones et al. 2010) and ABCDE approach (Smith and Bowden 2017). More specifically, every action that the player wants to take is taken by choosing options from the multiple-choice menu. The nonlinear gameplay allows the player to take actions in patient care in the order determined by the players themselves corresponding to the real-life decision-making situation.

The difficulty level of the game is predetermined by the scenario creators. The level of difficulty is related to the challenge of the patient scenarios, which were

defined according to the students' study phase and learning objectives. The difficulty level of patient scenarios varied depending on the clinical situation of the patients (e.g., mild or severe shortness of breath), the various text-based and visual cues provided for players to identify patients' need for care, and the nursing intervention and treatment options available. The level of difficulty did not adapt to the skills of the users but remained the same throughout the playthrough. Furthermore, the difficulty level of the scenario did not change when a player played the same scenario repeatedly. In the game, the student received scores for performance so that each choice was scored: right actions earned points and wrong actions reduced points. Thus, the scores described the students' performance and competence in each scenario.

In the case study (Havola et al. 2021), the computer version of the simulation game, as well as the VR simulation with head-mounted display (HMD), was integrated into the studies of graduating nursing students in one university of applied sciences. The aim was to investigate the effect of simulation games on students' CR skills but also to increase understanding of the use of simulation games, and in particular the VR simulation, as an educational tool in modules. Altogether, 40 nursing students participated in the study. The computer version included nine clinical scenarios in surgical, internal medicine, emergency, and home healthcare settings. For example, in the postoperative observation scenario, the patient's surgical wound was bleeding, and the student needed to get the bleeding under control and prevent the patient from experiencing hypovolemia. The playing time was unlimited. The VR simulation included one scenario. In the scenario, the player had to assess a patient who was experiencing chest pain and administer the necessary treatment when the patient collapsed. At the end of the scenario, the player had to provide post-resuscitation care in the intensive care unit. In the VR simulation, students played the scenarios once with unlimited playing time.

First, graduating nursing students played the single-player simulation game independently using a computer at home. They had the opportunity to play as many times as they wanted. However, they were instructed to play every scenario at least once. The students got access to the simulation game from an electronic learning platform. Second, the students played the VR simulation. VR gaming sessions were conducted at the university of applied science in a game studio. When students arrived at the game studio, one researcher explained the use of a VR headset, and hand controllers were introduced. Students could practice how to navigate in the VR environment before an actual gaming session. One researcher helped students if they needed advice with game technology. Otherwise, help with the content of the scenario was not given by the researcher.

The data consisted of the game metrics stored in the simulation game (Table 1). The analyzed game metrics included the number of playthroughs, scores, and playing time. In every scenario, the maximum score was 100. The number of playthroughs was defined as the number of all playing sessions, whether the player played the scenario to the very end or not. The mean score referred to the mean score of all playthroughs by all players, whereas the mean playing time referred to the mean playing time of all playthroughs by all players (Havola et al. 2021).

Table 1 Playthroughs with the simulation games ($n = 36\text{--}40$ nursing students)

Variable	Label	Mean	SD	Minimum	Maximum
Playthroughs with the computer-based simulation game ($n = 494$)					
Mean score ^a	Score	67	8.7	43	82
Mean time (Min.) ^b	Time	4.2	1.0	2.4	7.1
Max score ^c	Score	100 ^f	1.4	91	100
Max time (Min.) ^d	Time	7.6	3.5	4.5	21
Average number of played scenarios ^e		13.7	6.6	4	29
Playthroughs with the VR simulation ($n = 40$)					
Mean score ^a	Score	95	9.9	66	100
Mean time (Min.) ^b	Time	16.0	4.2	8	30.5
Max score ^c	Score	95	9.9	66	100
Max time (Min.) ^d	Time	16.0	4.2	8	30.5
Average number of played scenarios ^e		1	0	1	1

^aMean score: The mean score of all playthroughs by all players

^bMean time: The mean time of all playthroughs by all players

^cMax score: The maximum score of all playthroughs by all players

^dMax time: The maximum time of all playthroughs by all players

^eNumber of played scenarios: Frequency of all playthroughs and all scenarios by all players

^fScore has round to two decimals

In addition, the students' demographics were collected using an electronic survey. Students also self-evaluated their CR skills in three phases using the Clinical Reasoning Skills scale (CRSSs) (Koivisto et al. 2020): before and after playing the computer version of the game and after playing the VR simulation.

In the study, 494 playthroughs were conducted by students with a computer, while there were 40 playthroughs with a VR simulation altogether (one per student). The main results demonstrated that students' CR skills were systematically improved after game playing. There was a systematic association between better mean scores and better CR skills in playing both with computers and with VR headsets. Students spent more time in the VR simulation than playing with the computer; the mean student playing time was over 4 min of computer play, with VR simulation play over 15 min. Interestingly, a better mean score was achieved by spending less time playing with the computer. When playing the VR simulation, in turn, a better mean score was achieved when playing longer. On average, the students' mean score was 67 out of 100 in the computer game, while the mean score was 95 when playing the VR simulation.

Taken together, some interesting findings were found in this case study. The notable finding was that students' CR skills improved after playing both games. A clear difference was found when considering the differences between the playing time with a computer and a VR simulation. It is essential to notice the possible effect of the researcher's presence in the VR sessions when considering the differences between gaming sessions with computers and VR simulations. Possibly, students may have felt some social pressure while gaming. However, it can be stated that

students were more immersed in playing the VR simulation than in playing with the computer (Hamdaoui et al. 2017).

When using both a computer simulation game and a VR simulation for learning CR, it is essential to examine the order in which the different versions should be used to achieve effective learning outcomes. For example, Kim et al. (2020) found that the effectiveness of immersive VR on learning outcomes was improved when it was carried out after the traditional method (paper-pencil). In this study, students achieved better scores by playing the VR simulation compared to the computer version. This could indicate that the students became familiar with the game's technology by playing first with the computer. Therefore, better scores may be achieved in the second playing session by playing the VR version, even though the content of the scenario was not the same.

5 Directions for Future Work

The purpose of the current chapter was to discuss the potential of exploiting AI through game metrics in nursing education for learning CR skills, since the use of AI is still limited in nursing education (Randhawa and Jackson 2020), even though immersive technologies provide promising opportunities. For good learning experiences and learning outcomes in simulation learning, the level of difficulty of the scenario must be proportional to the learner's competence to achieve optimal flow during the scenario (Csikszentmihalyi 2000), which in turn could promote intrinsic motivation and improve performance. In the best nursing simulation games, learners can achieve a flow state since, in the applications, the game elements and game mechanics familiar from entertainment games have been utilized (e.g., Koivisto et al. 2018). However, to maximize good learning experiences and effective learning outcomes in simulation games, they should provide more personalized content (Hamdaoui et al. 2017). One way to personalize simulation games could be to adapt them to the learner's level of skills, and dynamic difficulty adjustment techniques could be used for that purpose (Streicher and Smeddinck 2016).

Next, future work to utilize game metrics in developing simulation games that adapt to the player's skill level is discussed. The case study has provided preliminary information on how game metrics describe students' scenario performance in a simulation game (Havola et al. 2021). The future aim could be to create simulation games that are adaptive to the skill levels of the players in clinical patient scenarios. This could mean, for example, that the patient's clinical condition changes based on the student's competence level, so that the difficulty level of the scenario decreases, remains the same, or increases (Streicher and Smeddinck 2016).

The first step in developing simulation games into an adaptive system is to determine which aspects of the simulation game should be adaptive (Streicher and Smeddinck 2016). The difficulty adjustment of the patient scenarios based on performance could be selected as an adaptive element. Second, adjustable parameters should be defined, and when talking about game metrics, the parameters

could include scores, playing time, and playthrough quantity (Havola et al. 2021). These game metrics could be collected automatically in triggered positions or periodically with a time interval. The different difficulty levels could be determined using previous information about the relationship between playing time and the number of playthroughs with scores. The difficulty levels of the simulation game scenarios can be defined as easy, medium, or difficult. At the easy level, the time spent on playing is short, the number of playthroughs is low and the scores are low, while at the difficult level, a lot of time is spent on playing, the number of playthroughs is high, and the scores are high. To validate the different levels, they need to be tested on a large number of students and a large number of playthroughs.

Third, levels of automation in adaptability, such as adjustment automation, should be identified (Streicher and Smeddinck 2016). Adjustment automation can range from fully manual to fully automated. With a fully manual adjustment level, simulation games could be static games with predefined difficulty levels, as was the case in the case study presented (Havola et al. 2021). In this option, the students choose the level of difficulty themselves. When students play the game, the system collects information about the time spent on playing, the number of playthroughs, and scores, and when students start a new scenario, the game system recommends a level for the students based on their previous performance. However, learners still choose a predefined level.

In a manual adaptive level, the difficulty levels in simulation games could be determined in advance based on previous knowledge of the relationship between playing time and the number of playthroughs with scores. When students execute a scenario, the system automatically directs the players to a certain difficulty level based on their behavior in the game. A fully automatic adaptability level could be developed into simulation games when enough information has been obtained about the performance of a sufficient number of players in the game. When there are a lot of data, machine-learning techniques could be utilized to determine the difficulty levels automatically. In this case, the automation level is fully adaptive: the difficulty level of the game changes automatically during gameplay based on the players' behavior in the game, that is, the time spent playing, number of playthroughs, and scores. To achieve this kind of adaptability, which is based entirely on players' competence, more player data are needed to utilize, for example, ML methods. In addition, research is needed on how automatic difficulty adoption in simulation games affects the students' learning experiences as well as learning outcomes.

6 Conclusion

The COVID-19 pandemic has challenged the clinical reasoning of healthcare professionals in identifying and treating the various clinical symptoms caused by the virus. This global situation has highlighted the importance of CR skills for patient safety in a somewhat frightening way. As mentioned earlier, in clinical work, AI can be used to support decision-making. However, this chapter has concentrated on

the potential benefits of AI in healthcare education, especially the use of simulation games in learning CR skills in nursing education. The focus has been on adapting the difficulty level of simulation games based on the knowledge and skills of the learners and suggesting the use of game metrics for doing so. Game metrics have not yet been utilized very commonly in nursing simulation games, although research in other disciplines has shown that game metrics are suitable for demonstrating the achievement of learning outcomes. The empirical findings in the case study presented here create a new understanding of the possibility of game metrics to provide objective information on the CR skills of nursing students. To effectively achieve the learning outcomes for which the game has been developed, students must remain engaged in the game for a prolonged period. Dynamic adjustment of the difficulty level of the patient scenarios could keep students immersed and in a state of flow in clinical scenarios, which, in turn, could contribute to the achievement of learning outcomes, not frustration and boredom. Taking advantage of recent technological developments in AI, playing adaptive simulation games could enable nursing students to achieve even better CR skills for working life and for constantly challenging clinical situations. This ultimately benefits the patient.

References

- Audétat, M.C., Sader, J., Coen, M. (2020). Clinical reasoning and COVID 19 pandemic: current influencing factors Let us take a step back! *Intern Emergency Medicine*, 16(4), 1109–1111 (2020). <https://doi.org/10.1007/s11739-020-02516-8>.
- Besse, C., Vogelsang, L., Zdunich, J. (2020). The Simulation Games: Bringin' It to the Bedside. *Nurse Educator*, 45(6), 339–342 (2020). <https://doi.org/10.1097/NNE.0000000000000832>.
- Botha, B.S., de Wet, L., Botma, Y. (2021). Undergraduate Nursing Student Experiences in Using Immersive Virtual Reality to Manage a Patient with a Foreign Object in the Right Lung. *Clinical Simulation in Nursing*, 56, 76–83 (2021). <https://doi.org/10.1016/j.ecls.2020.10.008>.
- Bracq, M.S., Michinov, E., Jannin, P. (2019a). Virtual Reality Simulation in Nontechnical Skills Training for Healthcare Professionals. *Simulation in Healthcare: The Journal of the Society for Simulation in Healthcare*, 14(3), 188–194 (2019). <https://doi.org/10.1097/sih.0000000000000347>.
- Bracq, M.S., Michinov, E., Arnaldi, B., Caillaud, B., Gibaud, B., Gouranton, V., Jannin, P. (2019b). Learning Procedural Skills with a Virtual Reality Simulator: An Acceptability Study. *Nurse Education Today*, 79, 153–160 (2019). <https://doi.org/10.1016/j.nedt.2019.05.026>.
- Breedt, S., & Labuschagne, M.J. (2019). Preparation of nursing students for operating room exposure: A South African perspective. *African Journal of Health Professions Education*, 11(1), 22–26 (2019). <https://doi.org/10.7196/AJHPE.2019.v11i1.1072>.
- Buchanan, C., Howitt, M.L., Wilson, R., Booth, R.G., Risling, T., Bamford, M. (2021). Predicted Influences of Artificial Intelligence on Nursing Education: Scoping Review. *JMIR Nursing*, 4(1). <https://doi.org/10.2196/23933>.
- Butt, A.L., Kardong-Edgren, S., Ellertson, A. (2018). Using Game-Based Virtual Reality with Haptics for Skill Acquisition. *Clinical Simulation in Nursing*, 16, 25–32 (2018). <https://doi.org/10.1016/j.ecls.2017.09.010>.
- Chang, C-Y., Kao, C-H., Hwang, G-J., Lin, F-H. (2020). From experiencing to critical thinking: a contextual game-based learning approach to improving nursing students' performance in electrocardiogram training. *Educational Technology Research and Development*, 68, 1225–1245 (2020). <https://doi.org/10.1007/s11423-019-09723-x>.

- Csikszentmihalyi, M. (2000). *Beyond boredom and anxiety*. Jossey-Bass.
- Drachen, A., Seif El Nasr, M., Canossa, A. (2013). Game Analytics – The Basics. In Seif El-Nasr, M., Drachen, A., & Canossa, A. (Ed.), *Game Analytics. Maximizing the Value of Player Data* (pp. 13–40). New York: Springer-Verlag. https://doi.org/10.1007/978-1-4471-4769-5_2.
- Fealy, S., Jones, D., Hutton, A., Graham, K., McNeill, L., Sweet, L., Hazelton, M. (2019). The integration of immersive virtual reality in tertiary nursing and midwifery education: A scoping review. *Nurse Education Today*, 79, 14–19 (2019). <https://doi.org/10.1016/j.nedt.2019.05.002>.
- Ferguson, C., Davidson, P. M., Scott, P. J., Jackson, D., Hickman, L. D. (2015). Augmented reality, virtual reality and gaming: an integral part of nursing. *Contemporary Nurse*, 51(1), 1–4 (2015). <https://doi.org/10.1080/10376178.2015.1130360>.
- Foronda, C.L., Fernandez-Burgos, M., Nadeau, C., Kelley, C.N., Henry, M.N. (2020). Virtual Simulation in Nursing Education: A Systematic Review Spanning 1996 to 2018. *Simulation in Healthcare: The Journal of the Society for Simulation in Healthcare*, 15(1), 46–54 (2020). <https://doi.org/10.1097/sih.0000000000000411>.
- Hamdaoui, N., Khalidi Idrissi, M., Bennani, S. (2017). Adaptive Educational Games Using Game Metrics. Proceedings of the Third International Afro-European Conference for Industrial Advancement – AECIA 2016. https://doi.org/10.1007/978-3-319-60834-1_21.
- Harmon, J., Pitt, V., Summons, P., Inder, K.J. (2021). Use of artificial intelligence and virtual reality within clinical simulation for nursing pain education: A scoping review. *Nurse Education Today*. <https://doi.org/10.1016/j.nedt.2020.104700>.
- Havola, S., Haavisto, E., Mäkinen, H., Engblom, J., Koivisto, J-M. (2021). The effects of computer-based simulation game and virtual reality simulation in nursing students' self-evaluated clinical reasoning skills. *Computers, Informatics, Nursing*, 39(11), 725–735 (2021). <https://doi.org/10.1097/CIN.0000000000000748>.
- Havola, S., Koivisto, J-M., Mäkinen, H., Haavisto, E. (2020). Game elements and instruments for assessing nursing students' experiences in learning clinical reasoning by using simulation games: An integrative review. *Clinical Simulation in Nursing*, 46, 1–14 (2020). <https://doi.org/10.1016/j.ecns.2020.04.003>.
- Holder, A. (2018). Clinical Reasoning: A State of the Science Report. *International Journal of Nursing Education Scholarship*. <https://doi.org/10.1515/ijnes-2016-0024>.
- Hunter, S., & Arthur, C. (2016). Clinical reasoning of nursing students on clinical placement: Clinical educators' perception. *Nurse Education in Practice*, 18, 73–79 (2016). <https://doi.org/10.1016/j.nepr.2016.03.002>.
- European Parliament (2013). Directive 2012/55/EU of the European Parliament and the Council of 20 November 2013 amending Directive 2005/36/EC of the European Parliament and of the Council of 7 September 2005 on the recognition of professional qualifications. <https://eur-lex.europa.eu/legal-content/FI/LSU/?uri=celex:32005L0036>. Accessed 10 May 2021.
- Kardong-Edgren, S., Breitkreuz, K., Werb, M., Foreman, S., Ellertson, A. (2019). Evaluating the Usability of a Second-Generation Virtual Reality Game for Refreshing Sterile Urinary Catheterization Skills. *Nurse Educator*, 44(3), 137–141 (2019). <https://doi.org/10.1097/NNE.0000000000000570>.
- Keys, E., Luctkar-Flude, M., Tyerman, J., Sears, K., Woo, K. (2020). Developing a Virtual Simulation Game for Nursing Resuscitation Education. *Clinical Simulation in Nursing*, 39, 51–54 (2020). <https://doi.org/10.1016/j.ecns.2019.10.009>.
- Keys, E., Luctkar-Flude, M., Tyerman, J., Sears, K., Woo, K. (2021). The Integration of Virtual Simulation Gaming Into Undergraduate Nursing Resuscitation Education: A Pilot Randomised Controlled Trial. *Clinical Simulation in Nursing*, 54, 54–61 (2021). <https://doi.org/10.1016/j.ecns.2021.01.013>.
- Kiili, K., Moeller, K., Ninaus, M. (2018). Evaluating the effectiveness of a game-based rational number training - In-game metrics as learning indicators. *Computers & Education*, 120, 13–28 (2018). <https://doi.org/10.1016/j.comedu.2018.01.012>.
- Kim, K.G., Oertel, C., Dobrocki, M., Olsen, J.K., Coppi, A.E., Cattaneo, A., Dillenbourg, P. (2020). Using immersive virtual reality to support designing skills in vocational education. *Br. J. Educ. Technol.*, 51, 2199–2213 (2020). <https://doi.org/10.1111/bjet.13026>.

- Koivisto, J.-M., Haavisto, E., Niemi, H., Haho, P., Nylund, S., Multisilta, J. (2018). Design principles for simulation games for learning clinical reasoning: A design-based research approach. *Nurse Education Today*, 60, 114–120 (2018). <https://doi.org/10.1016/j.nedt.2017.10.002>.
- Koivisto, J.-M., Rosqvist, K., Buure, T., Engblom, J., Haavisto, E. (2020). The effectiveness of a simulation game on nursing students' self-evaluated clinical reasoning skills: A quasi-experimental study. *Hoitotiede*, 32 suppl, 36–45 (2020).
- Koivisto, J. M., Niemi, H., Multisilta, J., & Eriksson, E. (2017). Nursing students' experiential learning processes using an online 3D simulation game. *Education and Information Technologies*, 22, 383–398. <https://doi.org/10.1007/s10639-015-9453-x>.
- Levett-Jones, T., Hoffman, K., Dempsey, J., Jeong, S. Y.-S., Noble, D., Norton, C. A., Roche, J., Hickey, N. (2010). The “five rights” of clinical reasoning: An educational model to enhance nursing students’ ability to identify and manage clinically “at risk” patients. *Nurse Education Today*, 30(6), 515–520 (2010). <https://doi.org/10.1016/j.nedt.2009.10.020>.
- Mawhirter, D. A., & Garofalo, P. F. (2017). Expect the unexpected: Simulation games as a teaching strategy. *Clinical Simulation in Nursing*, 12(4), 132–136 (2017). <https://doi.org/10.1016/j.ecns.2015.12.009>.
- McEnroe-Petite, D., & Farris, C. (2020). Using Gaming as an Active Teaching Strategy in Nursing Education. *Teaching and Learning in Nursing*, 15(1), 61–65 (2020). <https://doi.org/10.1016/j.teln.2019.09.002>.
- Mäkinen, H., Haavisto, E., Havola, S., Koivisto, J.-M. (2020). User experiences of virtual reality technologies for healthcare in learning: an integrative review. *Behaviour & Information Technology*. <https://doi.org/10.1080/0144929X.2020.1788162>.
- Peddle, M., Mckenna, L., Bearman, M., Nestel, D. (2019). Development of non-technical skills through virtual patients for undergraduate nursing students: An exploratory study. *Nurse Education Today*, 73, 94–101 (2019). <https://doi.org/10.1016/j.nedt.2018.11.008>.
- Plass, J.L., Homer, B.D., Kinzer, C.K., Kyung Chang, Y., Frye, J., Kaczetow, W., Isbister, K., Perlin, K. (2013). Metrics in Simulations and Games for Learning. In Seif El-Nasr, M., Drachen, A., & Canossa, A. (Ed.), *Game Analytics. Maximizing the Value of Player Data* (pp. 697–729). New York: Springer-Verlag. https://doi.org/10.1007/978-1-4471-4769-5_31.
- Pons Lelardeux, C., Galaup, M., Panzoli, D., Lagarrigue, P., & Jessel, J.-P. (2018). A Method to Design a Multi-Player Educational Scenario to Make Interdisciplinary Teams Experiment Risk Management Situation in a Digital Collaborative Learning Game: A Case of Study in Healthcare. *International Journal of Engineering Pedagogy*, 8(2), 88–100 (2018). <https://doi.org/10.3991/ijep.v8i2.8140>.
- Randhawa G.K., & Jackson M. (2020). The role of artificial intelligence in learning and professional development for healthcare professionals. *Healthcare Manage Forum*, 33(1), 19–24 (2020). <https://doi.org/10.1177/0840470419869032>.
- Sapci, A.H., & Sapci, H.A. (2020). Artificial Intelligence Education and Tools for Medical and Health Informatics Students: Systematic Review. *JMIR Med Educ*. <https://doi.org/10.2196/19285>.
- Shorey, S., Ang, E., Yap, J., Ng, E.D., Lau, S.T., Chui, C.K., A. (2019). Virtual Counseling Application Using Artificial Intelligence for Communication Skills Training in Nursing Education: Development Study. *J Med Internet Res*. <https://doi.org/10.2196/14658>.
- Simmons, B. (2010) Clinical reasoning: concept analysis. *Journal of Advanced Nursing*, 66(5), 1151–1158 (2010). <https://doi.org/10.1111/j.1365-2648.2010.05262.x>.
- Smith, D., & Bowden, T. (2017). Using the ABCDE approach to assess the deteriorating patient. *Nursing Standard*, 32(14), 51–63 (2017). <https://doi.org/10.7748/ns.2017.e11030>.
- Soflano, M., Connolly, T.M., Hainey, T. (2015). An application of adaptive games-based learning based on learning style to teach SQL. *Computers & Education*, 86, 192–211 (2015). <https://doi.org/10.1016/j.compedu.2015.03.015>.
- Streicher, A., & Smeddinck, J.D. (2016). Personalized and Adaptive Serious Games. In Dörner R., Göbel S., Kickmeier-Rust M., Masuch M., Zweig K. (Ed.), *Entertainment Computing and Serious Games. Lecture Notes in Computer Science*, vol 9970. Springer, Cham. https://doi.org/10.1007/978-3-319-46152-6_14.

Open Access This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.



AI-Supported Simulation-Based Learning: Learners' Emotional Experiences and Self-Regulation in Challenging Situations



Heli Ruokamo, Marjaana Kangas, Hanna Vuojärvi, Liping Sun,
and Pekka Qvist

Contents

1 Introduction	175
2 Theoretical Framework	176
2.1 Self-Regulated Learning	176
2.2 Positive and Negative Emotions in Simulation-Based Learning	177
2.3 Simulation-Based Learning Situations	179
3 Research Questions	179
4 Method	180
4.1 Data Collection	180
4.2 Analysis	182
5 Results	183
5.1 Learners' Positive and Negative Emotional Experiences During Simulation-Based Learning	183
5.2 Self-Regulated Learning Operations in Challenging Situations	185
5.3 Toward Developing AI Tutors in Simulation-Based Learning	187
6 Conclusion	188
References	189

1 Introduction

Emotions—both positive and negative—play an important role in learning (McConnell and Eva 2012), and previous research has shown that using simulations can meaningfully enhance learning (Brewer 2011; Keskitalo et al. 2014; Konia and Yao 2013). Learners' emotional reactions to simulation-based learning have been shown to improve both learning and recall of experiences and information (DeMaria

H. Ruokamo (✉) · M. Kangas · H. Vuojärvi · L. Sun

University of Lapland, Rovaniemi, Finland

e-mail: heli.ruokamo@ulapland.fi; marjaana.kangas@ulapland.fi; hanna.vuojarvi@ulapland.fi;
Liping.Sun@ulapland.fi

P. Qvist

NAPCON, Neste Engineering Solutions, Turku, Finland

e-mail: pekka.qvist@neste.com

et al. 2010). In this study, a simulation is an imitation of reality as “a means to do something in the ‘as if’, to resemble ‘reality’, [and to] learn something without the risks or costs of doing it in reality” (Rall and Dieckmann 2005, p. 2). As such, simulation-based learning is regarded as an experiential and a fun and safe way to learn (Brewer 2011; Hope et al. 2011; Keskitalo and Ruokamo 2017; Konia and Yao 2013; Weller 2004).

In this multidisciplinary study, we will explore trainees’ emotional experiences and how they overcome stressful situations in a simulation-based learning environment (SBLE). The participants are operator trainees in oil production at Neste Engineering Solutions Ltd. The study integrates chemical engineering and educational sciences, and it concerns the learning of behavioral, emotional, motivational, and cognitive processes. In essence, we are interested in the key factors that either facilitate or inhibit the learning during the simulation.

2 Theoretical Framework

2.1 Self-Regulated Learning

Self-regulated learning (SRL) plays an important role in the learning process in helping learners to optimize their practice (Zimmerman 2006). The term *self-regulated learning* emphasizes learners’ responsibility and autonomy during their learning (Paris and Winograd 1998). According to Zimmerman (2000a), the term describes “self-generated thoughts, feelings, and actions that are planned and cyclically adapted to the attainment of personal goal” (p. 14). In the process of regulation, learners can plan, set goals, organize, self-monitor, and self-assess, which makes them self-aware and knowledgeable of the learning procedures. They employ effort and persistence rather than giving up when tasks are challenging. By taking strategic action, learners seek out appropriate and helpful advice, information, and strategies to support their learning, and they self-instruct and self-reinforce during performance enactments (Zimmerman 2000b; Perry and Rahim 2011; Pintrich 2003). The objects of the regulatory processes are the different behavioral, motivational, and emotional aspects of the learning process (Zimmerman 2006). In this study, we approach the topic of SRL from an emotional perspective and focus on emotional determinants through which the simulator trainees regulate their learning process.

Technological development, especially the adaptation of the intelligent tutoring system (ITS), can be a transformative factor for understanding learning patterns, and it can support SRL through discovering and responding to students’ emotional states during learning with AI systems (Channa et al. 2021; Kelly and Heffernan 2015). ITS provides a friendly platform to explore and encourage self-regulated behaviors, and it has an effect on students’ emotional states so as to facilitate reasoning deeply, such as critical thinking, problem-solving, and connecting previous knowledge with current problems (Channa et al. 2021; Kelly and Heffernan 2015; Sabourin et al.

2013). ITS, driven by AI technology, helps students perceive emotions as a way to encourage optimal learning, and it supports students to regulate their learning (Channa et al. 2021; Kelly and Heffernan 2015; Sabourin et al. 2013).

Previous research has identified the potential of AI tutors to facilitate students' learning progress and their skills mastery in ITS (Long and Aleven 2013; Koedinger and Aleven 2007). Unlike other computer-supported education systems, AI tutors can "respond dynamically to the individual learning needs of each student" (Johnson et al. 2009, p. 31). That is, an AI tutor can understand students' problems and assess their analyses; thus, they can structure a response immediately (Johnson et al. 2009; Lane et al. 2015; Koedinger and Aleven 2007). For example, an AI tutor can provide students with feedback and hints gradually based on specific analyses and difficulties in each student's response (Johnson et al. 2009; Lane et al. 2015). Johnson et al. (2009) indicate that an AI tutor acts as a human tutor. In this study, we focus on the situations when an AI tutor could promote simulator trainees' SRL.

2.2 *Positive and Negative Emotions in Simulation-Based Learning*

Emotions are always intertwined with learning (Engeström 1982; Immordino-Yang and Faeth 2010; Schutz and DeCuir 2002; Schutz et al. 2011), and they can strongly modulate learning outcomes and experiences (Tyng et al. 2017) and affect learners' motivation, their behavior in learning environments, and their recall ability (Damasio 2001; DeMaria et al. 2010; McConnell and Eva 2012; Schwabe and Wolf 2009; Trigwell 2012). Emotional experiences can have a crucial impact on other cognitive processes, such as attention, memory, reasoning, and problem-solving (Jung et al. 2014; Tyng et al. 2017; Um et al. 2012; Vuilleumier 2005). Understanding emotions and their relationship to learning may be key for the development of educational settings that are more conducive to the success of both learners and instructors (Trigwell 2012). Emotions—also referred to as moods, feelings, affects, or attitudes—are the affective contents, states, and lived experiences (McConnell and Eva 2012; Schnall 2011). They can both facilitate and hinder learning, and their effects on learning are mediated by several factors (Keskitalo and Ruokamo 2017; Vesisenaho et al. 2019).

Emotional experiences are situated—and socially and personally constructed—within sociohistorical contexts that emerge from conscious or unconscious appraisals of a particular event (Schutz et al. 2011); they are usually categorized as positive, negative (Fraser et al. 2012), or neutral (Nummenmaa et al. 2013). According to the literature, negative emotions hinder learning, while positive emotions facilitate learning. When feeling positive emotions, individuals are more likely to concentrate on the bigger picture, and when feeling negative emotions, they tend to focus on details (McConnell and Eva 2012). As McConnell and Eva

(2012, p. 1317; see also Fredrickson 2001) indicate, “Positive emotions encourage people to see the forest, whereas negative emotions lead them to focus on leaves.”

However, the relationship between emotions and learning is complex (Fraser et al. 2012; McConnell and Eva 2012; Peterson et al. 2015; Schutz et al. 2011). When learners perceive a learning situation as threatening or frightening, they may have a better memory of the emotional event because of their cognitive activity, but it may be more challenging for them to make broader connections and thus transfer the knowledge to other contexts (McConnell and Eva 2012).

According to many researchers, positive emotions were more likely to be as conducive to learning than negative emotions (Duffy et al. 2016; McConnell and Eva 2012; Postareff et al. 2017), and they were considered to “facilitate approach behavior” (Fredrickson 2001, p. 219). Learners who experienced positive emotions were found to be more likely to engage with their learning environment, and positive emotions were also connected with deep learning approaches (Trigwell 2012). They were found to increase cognitive flexibility and verbal fluency and facilitate decision-making and creative thinking. However, they could also reduce perseverance and exacerbate distractibility, while negative emotions tended to narrow thinking to a focus on details while facilitating more accurate decision-making (Dreisbach and Goschke 2004; Duffy et al. 2016; Fredrickson 2001; McConnell and Eva 2012; Staal 2004). Stress and anxiety both have negative connotations but may benefit learning in certain cases (DeMaria et al. 2010; Pekrun et al. 2006; Postareff et al. 2017). Overall, both positive and negative emotions can be harmful to learning when they focus the learner’s attention on something that is an irrelevant content. It also seems that both positive and negative emotions may benefit learning to some degree, but further research is needed to clarify this (Duffy et al. 2016; Keskitalo and Ruokamo 2021; Postareff et al. 2017).

Simulation-based learning is considered a fun, an experiential, and a safe way to learn (Brewer 2011; Hope et al. 2011; Konia and Yao 2013; Weller 2004). Research has shown that simulation-based learning is more than just fun (Rosen 2008); it is also an effective way to learn (Cook et al. 2011; McGaghie et al. 2010). Simulations can be more powerful experiences than traditional learning methods due to authentic connections to the emotions and the reflections that they stimulate, if these are debriefed (Silvennoinen et al. 2020). Essentially, simulation is an imitation of reality, and a simulation setting can be expected to arouse strong feelings and a motivation to learn (Dieckmann et al. 2007). In an SBLE, scenarios and materials are usually constructed to elicit particular emotions (DeMaria et al. 2010) because comparable real-life situations might be challenging and stressful or cause cognitive overload (Andreatta et al. 2010). Simulation-based learning is generally expected to provide learners with active and experiential learning opportunities to help them better integrate theory into practice (Cleave-Hogg and Morgan 2002; Gaba 2004; Keskitalo 2012; Keskitalo and Ruokamo 2016; Rall and Dieckmann 2005). However, simulation-based learning must be planned appropriately to be effective (Kneebone 2003; McGaghie et al. 2010), considering educational principles and human nature (Keskitalo 2015).



Fig. 1 Simulator training environment replicating the actual workstation of the operator

2.3 Simulation-Based Learning Situations

Simulation-based learning builds on learners' interaction with the facilitator, with other learners, with the simulator environment, and with and through other technical devices.

The trainees involved in the research experiment were participants in a basic training phase at Neste, and learning topics involved in operating a large-scale process industrial plant. These topics cover usage of different automation systems, basic controls, using automatic process controllers, and operating different process units. Additionally the trainees had been previously working as summer interns operating the real process plant, and during the simulation training sessions, they had to employ their accumulated knowledge in individual training scenarios. The operator training simulator (OTS) environment very closely replicates the actual workstation of the plant operator, allowing seamless transfer of knowledge from the simulator training to the day-to-day operations of the plant (Fig. 1).

3 Research Questions

On the basis of the theoretical framework and previous research, the research questions for this study are as follows:

1. What kinds of emotions do learners experience in simulation-based learning situations?

2. Through what kinds of SRL operations do learners aim to overcome challenging situations during simulation-based learning?
3. In what kinds of situations could an AI tutor be used to facilitate simulation-based learning?

4 Method

4.1 Data Collection

The data were collected in two phases. The first phase took place during a 1-week experiment conducted in August 2021. Four simulation-based learning sessions were organized in a simulation environment provided by Neste Engineering Solutions Ltd. in Finland. The four sessions were identical in terms of content and pedagogy. Each session was facilitated by two simulation instructors and lasted for 1 working day. The simulation environment was a classroom equipped with four workstations and an instructor observation room in the middle (see Fig. 2). In the workstations, learners used simulator software provided by NAPCON Neste. The simulator represents the operational software used in steering the chemical processes in the field. During the training, the simulator and the operations were first introduced to the trainees. Next, the trainees operated the system (i.e., the simulator) independently and learned how to operate in typical error situations that may occur in the processes of the chemical industry. In these challenging situations, instructors provided help when needed.

NAPCON Simulator classroom experiment setup Aug. 16-19, 2021

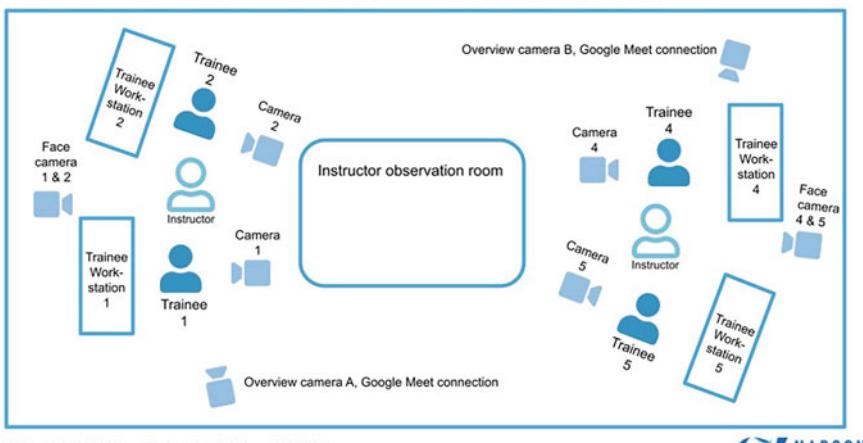


Fig. 2 Data collection setup in the simulation environment

The participants of this study ($N = 12$; nine males and three females) were summer employees at Neste Engineering Solutions Ltd. at the time of data collection. Each of them participated in one of the four training sessions. Participants were asked to provide informed consent to take part in the study. The data were collected in two phases. The first data collection phase was carried out through online observations and video recordings in the simulation environment (Fig. 2).

The setup included two overview cameras (A and B) with a Google Meet connection. These overview cameras were used by authors 1, 2, and 3 of this article to collect online observation data. On-site observations were not allowed, due to COVID-19 pandemic restrictions. Researchers observed each of the training sessions from beginning to end and wrote field notes during observations either by hand or using a word processor. Additionally, there was one over-the-shoulder camera recording the activities on each workstation (cameras 1–4) and two face cameras that recorded two workstations each. The first data collection phase yielded 161 h and 42 min of video data and 77 pages of observation notes written either by hand or a word processor.

The data collected in the first phase were used in preparing and conducting the second phase of data collection (i.e., the dSTR interviews). During the 2–3 weeks following the simulation training sessions, the researchers viewed the videos and read their field notes to identify situations that seemed to be challenging to participants. *Challenging learning situations* here mainly refer to cognitive learning challenges (Zimmerman 2011) that involve difficulties in understanding the concepts and solving the problems at hand. Motivational challenges (Zimmerman 2011) were not seen as relevant to the study, because all the participants practiced in the simulation environment to better succeed in their future work, so it could be assumed that they were motivated to enhance their knowledge and skills. Focusing on the challenging situations was considered important in determining how students aimed to overcome challenging situations and in determining in what kinds of situations an AI tutor could be used to facilitate learning. After getting familiarized with the video data and field notes, dSTR interviews were organized. Of the 12 participants in the first phase of data collection, 6 volunteered to take part in the second phase.

The basic idea of the STR interviews is that learners can relive the original situation with vividness and accuracy when presented with several cues or stimuli that occurred in the original situation (Bloom 1953). STR is an advanced interview method (Alexandersson 1994) that can be approached from different methodological perspectives and can produce an interpretation of the situation as the learners themselves conceive and understand it (Calderhead 1981). STR may also be elicited introspectively, with learners observing their internal processes in the same way they observe external real-world situations (Gass and Mackey 2000). STR involves the verbal reporting of learners' thinking processes in decision-making and problem-solving situations, and it is related to a variety of process tracing methods, including *think aloud* methods, and retrospective interviews (Shavelson and Stern 1981; Shavelson et al. 1986; Vesterinen et al. 2010).

In the dSTR interviews, the participants were first asked about their learning aims and general experiences in the simulation training. Next, the interviewees watched video clips from the situations identified as challenging. The researchers then asked questions to elicit participants' thoughts on those situations, as well as their actions and emotions when experiencing them (Keskitalo and Ruokamo 2017). At the end of the interviews, participants were presented with a short online questionnaire that included a list of 36 emotions, and they were asked to estimate how strongly they felt them during the simulation training on a scale from 1 (*not at all*) to 5 (*very strongly*). The questionnaire was designed using the Webropol online survey tool. The dSTR interviews lasted 22–45 min each. The interviews were recorded, and the data were transcribed verbatim, yielding 16,973 words of interview data.

4.2 Analysis

The researchers involved in data collection were also responsible for data analysis. The interview data was analyzed through a deductive thematic analysis process (Terry et al. 2017). The first step of the analysis was creating an analysis framework. The framework included three categories according to the research questions (Maguire and Delahunt 2017): emotions, operations, and experienced challenges. Second, each of the three researchers read through their interview data to get an overall picture of learners' experiences and to become familiarized with the data. The third phase of the analysis consisted of coding the data and marking everything related to the analysis framework. This included learners' expressions of thoughts and emotions during simulation-based learning, their descriptions of the operations through which they aimed to overcome challenging situations, and descriptions of situations experienced as challenging. Any expressions of experienced deficiencies in their own skills or the simulator software were also coded.

The fourth phase of the analysis began by combining the coded data extracts from the three researchers. All data extracts with the same code were aggregated, and the codes were collated into potential themes. Next, the collated data were reread, some of the coded data extracts were reorganized, and potential differences in interpretations were negotiated within the team. After that, sub-themes were created on the basis of the coding. The final step of the analysis included combining the sub-themes into primary themes and ensuring that each theme was justified and addressed to the research questions. Despite the linear presentation here, the analysis process involved moving back and forth between steps, which is common in qualitative research (Maguire and Delahunt 2017).

5 Results

5.1 Learners' Positive and Negative Emotional Experiences During Simulation-Based Learning

Research question 1 is "What kinds of emotions do learners experience in simulation-based learning?" Results from the emotions survey that participants completed during dSTR interviews show that positive emotions seem to be emphasized in learners' experiences. The five most reported and the five least reported emotions are presented in Fig. 3.

It seems that the simulation-based training was generally a positive experience for the learners. All five most reported emotions presented in Fig. 3 can be interpreted as positive, and the five least reported can be interpreted as negative. To get a deeper understanding of participants' experiences, their expressions regarding their emotions were coded from the data. Tables 1 and 2 below present examples of these codings and the emotions interpreted from them.

Learners experiencing positive emotions are more likely to engage with their simulation-based learning environment (SBLE) (Trigwell 2012). Positive emotions may increase learners' cognitive flexibility and verbal fluency and may facilitate decision-making and creative thinking.

Both positive and negative emotions can facilitate and hinder learning (Keskitalo and Ruokamo 2017; Tyng et al. 2017). The difference in the effects of positive and negative emotions is dependent on the learner's state of mind (McConnell and Eva 2012).

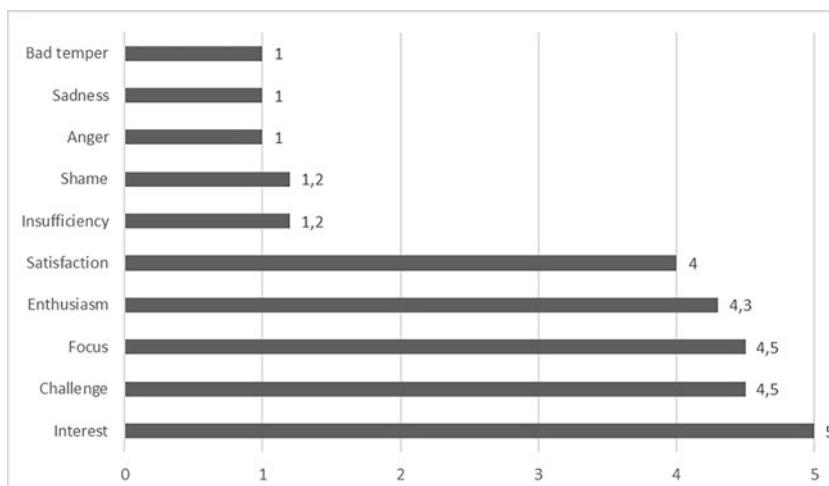


Fig. 3 Five most reported and five least reported emotions

Table 1 Data examples of positive emotional experiences during simulation-based learning

Data example	Positive emotion
That is a good place to practice then, and I was quite enthusiastic ... [Trainee 6]	Enthusiasm
But if there comes only one, so it is a bit different, you can take it more relaxed, and you can act as though. [Trainee 6]	Relaxed
That for sure in that situation was relieved per se, that you were quite in the same situation with others considering your own way of thinking. [Trainee 5]	Relieved
I was feeling super good because I didn't care that I made a mistake and because I was making at first an even bigger mistake ... so I was not so disturbed ... because I was really happy that I did the right thing. [Trainee 2]	Happy
Yeah, I have maybe never thought about how I feel when I am at work. It was not at all stressful ... [Trainee 2]	Non-stressful
I did not recognize that I would have been especially excited, neither before nor in that situation. [Trainee 2]	Non-excited
I'm not panicking because I'm used to it, that it sounds an alarm ... so those alarms in that kind of situation when we are in the simulator, those for sure have not frightened me. [Trainee 2]	Confidence
Kind of start to focus, and for sure a bit excited about what is going to show up. [Trainee 6]	Focused
In other words, the stress level decreased. I was really able to concentrate on what I was doing, and I was able to get a lot out of that. ... That was the first time in that situation I was able to apply what I had learned already. [Trainee 3]	Concentrated

Table 2 Data examples of negative emotional experiences during simulation-based learning

Data example	Negative emotion
I remember my heartbeat was very high, and I was very excited because I didn't know at all how I would react to that ... thought that was only a simulation ... I haven't had so much experience with it, so it was very exciting. [Trainee 3]	Excitement
So my ideas were escaping, and I was in that phase still a bit excited, so kind of impatient ... [Trainee 3]	Impatience
Here comes just the uncertainty. I realized that I was not exactly sure how I should act. I had done too little with those pumps. [Trainee 3]	Uncertain
anxiety that, but ... [Trainee 6]	Anxiety
I might have been a bit frustrated when I wasn't able to put the pump on ... [Trainee 2]	Frustration
that it is always very bad that you have to be able to concentrate ... [Trainee 6]	Challenged
It took quite a long time to understand what you should have done when you were only reading those guidelines ... [Trainee 1]	Confusion
So it was quite an empty feeling when you couldn't or felt you couldn't do anything, yet you should then do something anyway. [Trainee 4]	Emptiness

5.2 *Self-Regulated Learning Operations in Challenging Situations*

In this section we answer research question 2., "Through what kinds of SRL operations do learners aim to overcome challenging situations during simulation-based learning?" During simulation-based learning, the trainees met several challenging situations related to chemical engineering and process operating. These tasks were often experienced as stressful, and emotional regulation was needed to cope with the situation. The findings show that to overcome challenging situations, the trainees resorted to the following SRL operations: (1) metacognitive monitoring, (2) social scaffolding, (3) cognitive operations, and (4) emotional regulation.

First, *metacognitive monitoring* (Zimmerman 2008) occurred in the situations when trainees did not know what to do or expect. During the simulation, unexpected situations were faced, and the trainees needed to solve emergency problems using their own screens. The metacognitive monitoring strategies they used included intensively studying at the charts on the screen, going through working phases in their mind, prioritizing tasks, and predicting and envisaging forthcoming problems and challenges.

I prepared and anticipated which screens [of eight screens] those [changes in chemical processes] would come. As you can notice [from the video], I moved the small screen boxes here to make room for . . . well, there, I assumed the alarm would come; I made room for the screen so that I could see what would be happening there, because I had earlier been in an operating room, so I roughly knew or guessed and presumed what the instructor was aiming to do, and I prepared for that so I would be instantly there when something would happen. [Trainee 3]

Well, I looked at the chart that was there. Then I tried to go through those operational phases . . . you know, in my head—where to start, and what to do first. [Trainee 4]

Signs of metacognitive monitoring in the trainees' responses were verbs such as *predicting, assuming, knowing, guessing, figuring out, and thinking about*. Metacognitive monitoring enables learners to plan and monitor their own knowledge and skill levels, thus helping them to proceed in the task (Tzohar-Rosen and Kramarski 2014; Zimmerman 2008). Metacognitive monitoring can be seen as a systematic form of self-observation in an endeavor to understand the problem, devise a plan to proceed, implement a strategy, and check the accuracy of one's own thinking (Tzohar-Rosen and Kramarski 2014).

Second, although the trainees had to take active charge of their learning, the instructors provided them with help if needed. In addition, other trainees provided help in challenging situations. These strategies are called *social scaffolding* (Naukkarinen and Sainio 2018; Pea 2004) and include social support received to overcome the situation. The trainees asked questions to the instructors, or the instructors provided them with help and feedback if they noticed the trainees were stuck. The following excerpts illustrate the learning situations where social scaffolding was received.

Yes, it was [the trainer's help]; it was really good. Without it I wouldn't have noticed that point there. [Trainee 4]

I remember that the instructor came and said straightforwardly that I should do this, and then I moved forward from there. I was a bit in trouble there. The instructor said that I was on the right track, that I just needed to finish what I was doing. I had made a mistake, and he told me to fix that . . . so the instructor gave me the final solution. . . . As you can see [from the video] I have quit touching my hair and mask. [Trainee 3]

The last excerpt shows that the trainee noticed that her nonverbal communication no longer appeared restless after receiving social scaffolding and getting back on track. In the SBLE, the trainees felt it important to receive social support and feedback, even though they also had some ideas for developing SBLEs digitally so that scaffolding could be provided by an AI tutor. Wood et al. (1976) coined the term *scaffolding* for the first time and stated that scaffolding enables a novice to solve a problem and achieve a goal that would otherwise be beyond their unaided ability.

Third, to overcome stressful situations, the trainees also leaned on *cognitive operations*. Here, *cognitive operations* refer to cognitive processes and operating actions in the SBLE. The trainees reflected that by focusing and concentrating on those operations, they could go on and overcome difficult situations. Those activities included both mind-on activities, such as reasoning and problem-solving, and screen-on activities, such as reading through the alarm list, looking through the regulators, and checking the status of the regulators. The following excerpts illustrate the trainees' experiences:

At first, I was really confused trying to figure out what would be the first task. Then I realized I had to increase the gas intake to fuel the fire and increase the air level simultaneously to maintain balance. I got the hang of it there; honestly, I was quite confused. Of course, I checked the alarm from the list to find out which regulator the alarm was about and then I checked the regulator, what's the situation there. If the alarm is red or blinking the situation is quite bad, and one should really react and figure out what to do with it. [Trainee 6]

Well, I can remember I couldn't get the point directly, when there were many notifications at the same time. And they [the instructors] did not say exactly what the problem was, so I needed to sort out a bit before you realized that, okay, the incinerator is out of gas. . . . It took a while to understand that this is . . . this is the matter. [Trainee 5]

After those alarms I saw what started to happen, and it took a couple of minutes to figure out what I can and cannot do. Then the tension stopped and I was able to use my brain normally and think normally. [Trainee 3]

Fourth, to overcome stressful situations, emotional regulation was needed. This manifested as accepting a possible failure, understanding realities, or taking a time-out.

At that point I had a blackout. I knew in principle what to do but wasn't sure at all if I was on the right track. You know what's right and what's left but suddenly get all mixed up and can't show where right is. That's why I was quiet for a while. I gathered my thoughts and waited . . . counted how to justify myself that my decision was right. That is why I'm quiet here for quite a long time as I was calculating that, yes, this is what I have to do, and I have to close those vents. I don't remember what I said to the radio earlier, but I guess I asked to close that vent or something. [Trainee 3]

This excerpt shows that the trainee was aware of her stress reactions and that she needed to gather her thoughts to calm down and think clearly. This example shows that, in this case, negative feelings and feelings of stress hindered the learning process. As earlier research demonstrated, when feeling positive emotions, individuals are more likely to be cognitively flexible, open to information, and able to concentrate on the bigger picture; when feeling negative emotions, they tend to focus on details associated with a learning scenario, which may be beneficial in tasks that require a strong attention to detail (McConnell and Eva 2012).

5.3 Toward Developing AI Tutors in Simulation-Based Learning

Next, we will answer the research question 3.: “In what kinds of situations could an AI tutor be used to facilitate simulation-based learning?” The findings of this study reveal that AI could support the learning and operating processes in the following ways: (1) by providing decision-making aid, (2) by visualizing critical spots in the system, and (3) by asking questions to help check the system and make decisions.

First, it was evident that an AI tutor could provide support for making decisions (i.e., it could act as a decision-making aid for the learner). One option would be to provide a list of possibilities concerning how to continue when a difficult situation is faced. The trainees considered it important, however, that they could make the final decision by themselves, based on clues provided by the system.

At that point I faced another problem: how to open that vent, as it was automatically closed. I had to do something before I could open it, but I didn't know what that something would be. So there was a bit of a blackout. [Trainee 4]

what to do. Could there be for example a list of choices or just everything you need to . . . yes, there could be a list of all the possible choices, and then you could figure out what to do and in what order. Then you would know all the things you should do but would need to figure out the order by yourself. [Trainee 2]

The second way to facilitate the learner's process through an AI tutor would be to provide visual clues of the critical spots in the system. This would help the learners to focus their attention on the relevant things in the situation.

I couldn't check the route on the computer; that all would be green, and the pump could be started. That's why I couldn't make the final decision. [Trainee 4]

That [leaking pump] should have been shut down, but there was some obstacle for that, and I just couldn't see what it was. [Trainee 4]

The third possible way to use an AI tutor in the process would be through presenting the learner with questions during the process. Through well-formulated questions, the learner could check the system and make decisions.

Yeah, well, he didn't exactly say I should do this or that, but he just asked those right questions, and I started to think that of course that would be it. [Trainee 5]

Previous research shows that the dynamic features of an AI tutor can provide many benefits for students to regulate their own learning behaviors and emotions (Koedinger and Aleven 2007; Long and Aleven 2013). The instruction and feedback provided by the AI tutor are immediate and designed to further the process and outcomes of problem-solving simultaneously; they are thus adapted to individual students' needs (Johnson et al. 2009; Koedinger and Aleven 2007; Lane et al. 2015). These interventions can also teach learners to assess their learning performance and to select appropriate strategies in response to those assessments (Long and Aleven 2013). Zheng et al. (2021) state that learners have different emotions when experiencing these interventions, which thus play a part in self-regulating their learning.

6 Conclusion

The results of this study support the earlier findings of McConnell and Eva (2012): emotions are deeply connected with how learners use available information and with how they act on that information in learning and practice scenarios. During simulation-based learning, learners experience various positive and negative emotions that can both enhance and hinder learning. Further research is needed to describe these connections in more detail.

The ability to use metacognitive monitoring strategies (Zimmerman 2008) is evident from the progress made in simulation-based learning, and when receiving social support from others (i.e., social scaffolding; Naukkarinen and Sainio 2018; Pea 2004), these strategies enable learners to overcome challenging situations. Cognitive operations and emotional regulation are also important in all simulation-based learning to enable learners to proceed. The results of this study suggest three ways to involve an AI tutor in the simulation-based learning process. An AI tutor can provide help for decision-making, visualize critical points in the system, and ask questions that help the learner to check vital points in the system.

This study has some limitations. First, the number of participants is rather small. However, the group was self-selective, as the participants were summer employees at Neste at the time of data collection, and additional participants were not available. The data were gathered by three researchers, and they all analyzed their own interview data, which may have caused variation in the interpretation. This variation effect was minimized through negotiations and discussions during the analysis process. Collecting data through online observation and analyzing participants through videos may have caused misinterpretations, but watching the video clips together with the interviewees helped us to clarify those interpretations. Having video cameras on-site may have caused disturbances during observation, but having researchers present may have had the same effect.

References

- Alexandersson, M. (1994). Metod och medvetande [Method and consciousness] [Unpublished doctoral dissertation]. University of Gothenburg.
- Andreatta, P. B., Hillard, M., & Krain, L. P. (2010). The impact of stress factors in simulation-based laparoscopic training. *Surgery*, 147(5), 631–639.
- Bloom, B. S. (1953). Thought-processes in lectures and discussions. *Journal of General Education*, 7, 160–169.
- Brewer, E. P. (2011). Successful techniques for using human patient simulation in nursing education. *Journal of Nursing Scholarship*, 43(3), 311–317.
- Calderhead, J. (1981). Stimulated recall: A method for research on teaching. *British Journal of Educational Psychology*, 51, 211–217.
- Channa, F. R., Sarhandi, P. S. A., Bugti, F., & Brohi, I. A. (2021). Supporting self-regulated learning by affect detection and responding in AI-driven learning systems. *Ilkogretim Online*, 20(5), 3205–3211.
- Cleave-Hogg, D., & Morgan, P. J. (2002). Experiential learning in an anaesthesia simulation centre: Analysis of students' comments. *Medical Teacher*, 24, 23–26.
- Cook, D. A., Hatala, R., Brydges, R., Zendejas, B., Szostek, J. H., Wang, A. T., et al. (2011). Technology-enhanced simulation for health professions education: A systematic review and meta-analysis. *Journal of the American Medical Association*, 306(9), 979–988.
- Damasio, A. (2001). *Descartesin virhe. Emootio, järki ja ihmisen aivot* [Descarte's mistake. Emotions, intelligence and human brains]. Terra Cognita.
- DeMaria, S., Bryson, E. O., Mooney, T. J., Silverstein, J. H., Reich, D. L., Bodian, C., et al. (2010). Adding emotional stressors to training in simulated cardiopulmonary arrest enhances participant performance. *Medical Education*, 44(10), 1006–1015.
- Dieckmann, P., Gaba, D., & Rall, M. (2007). Deepening the theoretical foundations of patient simulation as social practice. *Simulation in Healthcare*, 2, 183–193.
- Dreisbach, G., & Goschke, T. (2004). How positive affect modulates cognitive control: Reduced perseveration at the cost of increased distractibility. *Journal of Experimental Psychology: Learning, Memory, Cognition*, 30(2), 343–353 (2004). <https://doi.org/10.1037/0278-7393.30.2.343>
- Duffy, M. C., Lajoie, S., & Lachapelle, K. (2016). Measuring emotions in medical education: Methodological and technological advances within authentic medical learning environments. In S. Bridges, L. K. Chan, & C. E. Hmelo-Silver (Eds.), *Educational technologies in medical and health sciences: Advances in medical education* (pp. 181–213). Springer.
- Engeström, Y. (1982). *Mielekäs oppiminen ja opetus* [Meaningful learning and teaching]. Valtion koulutuskeskus.
- Fraser, K., Irene, M., Teteris, E., Baxter, H., Wright, B., & McLaughlin, K. (2012). Emotion, cognitive load and learning outcomes during simulation training. *Medical Education*, 46(11), 1055–1062.
- Fredrickson, B. (2001). The role of positive emotions in positive psychology. *American Psychologist*, 56(3), 218–226.
- Gaba, D. M. (2004). The future vision of simulation in health care. *Quality and Safety in Health Care*, 13, 2–10.
- Gass, S. M., & Mackey, A. (2000). *Stimulated recall methodology in second language research*. Lawrence Erlbaum.
- Hope, A., Garside, J., & Prescott, S. (2011). Rethinking theory and practice: Pre-registration student nurses experiences of simulation teaching and learning in the acquisition of clinical skills in preparation for practice. *Nurse Education Today*, 31, 711–715.
- Immordino-Yang, M. H., & Faeth, M. (2010). The role of emotion and skilled intuition in learning. In D. Sousa (Ed.), *Mind, brain and education: Neuroscience implications for the classroom* (pp. 69–83). Solution Tree Press.

- Johnson, B. G., Phillips, F., & Chase, L. G. (2009). An intelligent tutoring system for the accounting cycle: Enhancing textbook homework with artificial intelligence. *Journal of Accounting Education*, 27, 30–39 (2009). <https://doi.org/10.1016/j.jaccedu.2009.05.001>
- Jung, N., Wranke, H. K., & Knauff, M. (2014). How emotions affect logical reasoning: Evidence from experiments with mood-manipulated participants, spider phobics, and people with exam anxiety. *Frontiers in Psychology*, 5, 570.
- Kelly K., & Heffernan N. (2015). Developing self-regulated learners through an intelligent tutoring system. In C. Conati, N. Heffernan, A. Mitrovic, & M. Verdejo (Eds.), *International Conference on Artificial Intelligence in Education. Lecture Notes in Computer Science*: Vol. 9112. Springer, Cham. https://doi.org/10.1007/978-3-319-19773-9_128
- Keskitalo, T. (2012). Students' expectations of the learning process in virtual reality and simulation-based learning environments. *Australasian Journal of Educational Technology*, 28, 841–856.
- Keskitalo, T. (2015). *Designing a pedagogical model simulation-based healthcare education* (Publication No. 299) [Doctoral Dissertation, University of Lapland]. <https://lauda.ulapland.fi/handle/10024/61885>
- Keskitalo, T., & Ruokamo, H. (2016). Students' expectations and experiences of meaningful simulation-based medical education. *Seminar.net*, 12(2). <https://journals.oslomet.no/index.php/seminar/article/view/2331>
- Keskitalo, T., & Ruokamo, H. (2017). Students' emotions in simulation-based medical education. *Journal of Interactive Learning Research*, 28(2), 149–159.
- Keskitalo, T., & Ruokamo, H. (2021). Exploring learners' emotions and emotional profiles in simulation-based medical education. *Australasian Journal of Educational Technology*, 37(1), 15–26 (2021). <https://doi.org/10.14742/ajet.5761>
- Keskitalo, T., Ruokamo, H., & Gaba, D. (2014). Towards meaningful simulation-based learning with medical students and junior physicians. *Medical Teacher*, 36(3), 230–239 (2014). <https://doi.org/10.3109/0142159X.2013.853116>
- Kneebone, R. (2003). Simulation in surgical training: Educational issues and practical applications. *Medical Education*, 37, 267–277.
- Koedinger, K. R., & Aleven, V. (2007). Exploring the assistance dilemma in experiments with cognitive tutors. *Educational Psychology Review*, 19, 239–264.
- Konia, M., & Yao, A. (2013). Simulation a new educational paradigm? *Journal of Biomedical Research*, 27(2), 75–80.
- Lane, H. C., Core, M. G., Hays, M. J., Auerbach, D., & Rosenberg, M. (2015). Situated pedagogical authoring: Authoring intelligent tutors from a student's perspective. In C. Conati, N. Heffernan, A. Mitrovic, & M. F. P. Verdejo (Eds.), *Proceedings of the 17th International Conference on Artificial Intelligence in Education* (pp. 195–204). Springer. <http://www.springer.com/series/1244>
- Long, Y., & Aleven, V. (2013). Supporting students' self-regulated learning with an open learner model in a linear equation tutor. In H. C. Lane, K. Yacef, J. Mostow, & P. Pavlik (Eds.), *Proceedings of the 16th International Conference on Artificial Intelligence in Education* (pp. 219–228). Springer. https://doi.org/10.1007/978-3-642-39112-5_23
- Maguire, M., & Delahunt, B. (2017). Doing a thematic analysis: A practical, step-by step guide for learning and teaching scholars. *All Ireland Journal of Higher Education*, 9(3), 3351–3364.
- McConnell, M. M., & Eva, K. W. (2012). The role of emotion in the learning and transfer of clinical skills and knowledge. *Academic Medicine*, 87(10), 1316–1322.
- McGaghie, C., Issenberg, S. B., Petrusa, E. R., & Scalese, R. J. (2010). A critical review of simulation-based medical education research: 2003–2009. *Medical Education*, 44, 50–63.
- Naukkarinen, J., & Sainio, T. (2018). Supporting student learning of chemical reaction engineering using a socially scaffolded virtual laboratory concept. *Education for Chemical Engineers*, 22, 61–68.
- Nummenmaa, L., Gleean, E., Hari, R., & Hietanen, J. K. (2013). Bodily maps of emotions. *Proceedings of the National Academy of Sciences*, 111(2), 646–651.

- Paris, S. G., & Winograd, P. (1998). The Role of Self-Regulated Learning in Contextual Teaching: Principles and Practices for Teacher Preparation. In *Contextual teaching and learning: Preparing teachers to enhance student success in the workplace and beyond* (Information Series No. 376) (pp. 219–252). ERIC Clearinghouse on Adult, Career, and Vocational Education, Columbus, OH.
- Pea, R. (2004). The social and technological dimensions of scaffolding and related Theoretical concepts for learning, education, and human activity. *The Journal of the Learning Sciences*, 13(3), 423–451.
- Pekrun, R., Elliot, A. J., & Maier, M. A. (2006). Achievement goals and discrete achievement emotions: A theoretical model and prospective test. *Journal of Educational Psychology*, 98(3), 583–597 (2006). <https://doi.org/10.1037/0022-0663.98.3.583>
- Perry, N. E., & Rahim, A. (2011). Studying self-regulated learning in classroom. In B. J. Zimmerman, & D. H. Schunk (Eds.), *Handbook of self-regulation of learning and performance* (pp. 122–136). Routledge.
- Peterson, E. R., Brown, G. T. L., & Jun, M. C. (2015). Achievement emotions in higher education: A diary study exploring emotions across an assessment event. *Contemporary Educational Psychology*, 42, 82–96 (2015). <https://doi.org/10.1016/j.cedpsych.2015.05.002>
- Pintrich, P. R. (2003). A motivational science perspective on the role of student motivation in learning and teaching contexts. *Journal of Educational Psychology*, 95(4), 667–686.
- Postareff, L., Mattsson, M., Lindblom-Ylänne, S., & Hailikari, T. (2017). The complex relationship between emotions, approaches to learning, study success and study progress during the transition to university. *Higher Education*, 73(3), 441–457 (2017). <https://doi.org/10.1007/s10734-016-0096-7>
- Rall, M., & Dieckmann, P. (2005). Simulation and patient safety: The use of simulation to enhance patient safety on a systems level. *Current Anaesthesia & Critical Care*, 16, 273–281.
- Rosen, K. R. (2008). The history of medical simulation. *Journal of Critical Care*, 23, 157–166.
- Sabourin, J. L., Shores, L. R., & Mott, B. W. (2013). Understanding and predicting student self-regulated learning strategies in game-based learning environments. *International Journal of Artificial Intelligence in Education*, 23, 94–114 (2013). <https://doi.org/10.1007/s40593-013-0004-6>
- Schnall, S. (2011). Affect, mood and emotions. In S. Järvelä (Ed.), *Social and emotional aspects of learning* (pp. 59–64). Oxford.
- Schutz, P. A., & DeCuir, J. T. (2002). Inquiry on emotions in education. *Educational Psychologist*, 37(2), 125–134.
- Schutz, P. A., Quijada, P. D., de Vries, S., & Lynde, M. (2011). Emotion in educational contexts. In S. Järvelä (Ed.), *Social and emotional aspects of learning* (pp. 64–69). Oxford.
- Schwabe, L., & Wolf, O. T. (2009). Stress prompts habit behavior in humans. *Journal of Neuroscience*, 29(22), 7191–7198.
- Shavelson, R. J., & Stern, P. (1981). Research on teachers' pedagogical thoughts, judgments, decisions, and behaviour. *Review of Educational Research*, 51, 455–498.
- Shavelson, R. J., Webb, N. M., & Burstein, L. (1986). Measurement of teaching. In M. Wittrock (Ed.), *Handbook of research on teaching* (pp. 50–91). Macmillan.
- Silvennoinen, M., Vesisenaho, M., Manu, M., Kullberg, T., Malinen, A., & Parviainen, T. (2020). Methodology development in adult learning research: Combining physiological reactions and learning experiences in simulation-based learning environments. In L. Gomez Chova, A. Lopez Martinez, & I. Candel Torres (Eds.), *EDULEARN20 Proceedings. 12th International Conference on Education and New Learning Technologies* (pp. 5037–5046). IATED. <https://doi.org/10.21125/edulearn.2020.1316>
- Staal, M. A. (2004). *Stress, cognition, and human performance: A literature review and conceptual framework*. NASA Ames Research Center.
- Terry, G., Hayfield, N., Clarke, V., & Braun, V. (2017). Thematic analysis. In C. Willig, & W. Stainton Rogers (Eds.), *The SAGE handbook of qualitative research in psychology* (pp. 17–37). SAGE.

- Trigwell, K. (2012). Relations between teachers' emotions in teaching and their approaches to teaching in higher education. *Instructional Science*, 40(3), 607–621.
- Tyng, C. M., Amin, H. U., Saad, M. N. M., & Malik, A. S. (2017). The influences of emotion on learning and memory. *Frontiers in Psychology*, 8, 1454.
- Tzohar-Rosen, M., & Kramarski, B. (2014). Metacognition, motivation and emotions: Contribution of self-regulated learning to solving mathematical problems. *Global Education Review*, 1(4), 76–95.
- Um, E., Plass, J. L., Hayward, E. O., & Homer, B. D. (2012). Emotional design in multimedia learning. *Journal of Educational Psychology*, 104, 485–498.
- Vesinenaho, M., Juntunen, M., Häkkinen, P., Pöysä-Tarhonen, J., Fagerlund, J., Miakush, I., et al. (2019). Virtual reality in education: Focus on the role of emotions and physiological reactivity. *Journal of Virtual Worlds Research*, 12(1). <https://doi.org/10.4101/jvwr.v12i1.7329>
- Vesterinen, O., Toom, A., & Patrikainen, S. (2010). The stimulated recall method and ICTs in research on the reasoning of teachers. *International Journal of Research & Method in Education*, 33(2), 183–197.
- Vuilleumier, P. (2005). How brains beware: Neural mechanisms of emotional attention. *Trends in Cognitive Sciences*, 9(12), 585–594.
- Weller, J. M. (2004). Simulation in undergraduate medical education: Bridging the gap between theory and practice. *Medical Education*, 38(1), 32–38.
- Wood, D., Bruner, J., & Ross, G. (1976). The role of tutoring in problem solving. *Journal of Child Psychology and Psychiatry and Allied Disciplines*, 17, 89–100.
- Zheng, J., Huang, L. Y., Li, S., Lajoie, S. P., Chen, Y. X., & Hmelo-Silver, C. E. (2021). Self-regulation and emotion matter: A case study of instructor interactions with a learning analytics dashboard. *Computer & Education*, 161.
- Zimmerman, B. J. (2000a). Attaining self-regulation. A social cognitive perspective. In M. Boekaerts, P. R. Pintrich, & M. Zeidner (Eds.), *Handbook of Self-Regulation* (pp. 13–39). Academic Press.
- Zimmerman, B. J. (2000b). Becoming a self-regulated learner: An overview. *Theory into Practice*, 41(2), 64–70.
- Zimmerman, B. J. (2006). Development and adaptation of expertise: The role of self-regulatory processes and beliefs. In K. A. Ericsson, N. Charness, P. J. Feltovich, & R. R. Hoffman (Eds.), *The Cambridge handbook of expertise and expert performance* (pp. 705–722). Cambridge University Press.
- Zimmerman, B. J. (2008). Goal setting: A key proactive source of academic self-regulation. In D. H. Schunk & B. J. Zimmerman (Eds.), *Motivation and self-regulated learning. Theory, research and applications* (pp. 267–295). Routledge.
- Zimmerman, B. (2011). Motivational sources and outcomes of self-regulated learning and performance. In B. Zimmerman & D. Schunk (Eds.), *Handbook of self-regulation of learning and performance* (pp. 49–64). Routledge.

Open Access This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.



Part III

AI Technologies for Education

and Intelligent Tutoring Systems

Training Hard Skills in Virtual Reality: Developing a Theoretical Framework for AI-Based Immersive Learning



Tiina Korhonen, Timo Lindqvist, Joakim Laine, and Kai Hakkarainen

Contents

1 Introduction	195
2 From ITS to AI Tutor	197
2.1 Intelligent Tutoring System (ITS)	197
2.2 Observability	198
2.3 Modifiability	198
2.4 AI Tutor	199
3 Grounded Cognition	199
4 VR-Native AI Tutor Framework	203
4.1 Situated Conceptualizations	203
4.2 Physical Environment Simulation	205
4.3 Specifiers	205
4.4 Learner and Tutor Simulations	206
4.5 Implementing the Framework	206
5 Toward VR-Native Pedagogy	207
5.1 Simulation Environment	207
5.2 Task Sequencing	208
5.3 Scaffolding	208
6 Discussion	210
References	211

1 Introduction

This chapter explores the pedagogical setting of hard skills training that takes place in immersive virtual reality (VR), guided by artificial intelligence (AI) tutoring software. Since the commercial introduction of sophisticated but affordable

T. Korhonen (✉) · J. Laine · K. Hakkarainen
University of Helsinki, Helsinki, Finland

e-mail: tiina.korhonen@helsinki.fi; joakim.laine@helsinki.fi; kai.hakkarainen@helsinki.fi

T. Lindqvist
Upknowledge, Helsinki, Finland
e-mail: timo.lindqvist@upknowledge.com

immersive virtual reality hardware around 2016, immersive VR technology has generated widespread interest for training practical skills. This could be due to the technology's profound educational affordances: (1) providing a strong sense of presence and (2) affording agentic embodiment of operational activity (Johnson-Glenberg 2018). As such, VR is considered to have promise across many educational domains. In the same time frame, a family of machine learning methods based on deep hierarchies of neural network layers, known as "deep learning," has made major advances in enabling practical AI systems that act on classifications of large amounts of observed data. The key aspect of deep learning is that the constituent features making up different classes are not engineered by humans but learned from training data (see e.g., Goodfellow et al. 2015).

Our research aims to examine how the inherent features of VR, such as *modifiability* and *observability*, could benefit AI-based tutoring software. A software program performing real-time inference on models observed from a learner's behavior in VR – what we call an *AI tutor* – could observe more patterns in learner activity than what is within the capabilities of a human trainer. Based on these observations, the tutor could modify the VR environment dynamically to support learning. Also, unlike most human trainers, an AI tutor could maintain its attention on the learner constantly.

The focus of our current work is training hard skills in industrial settings. In these settings, immersive virtual training environments (IVRTEs) are used to simulate real-life operational environments where learners can practice the use of equipment or the mechanisms of machinery and perform safety and work procedures. This domain offers an interesting area for research as: (1) knowledge to be learned is mostly procedural, allowing experimental setups that better isolate phenomena attributable to VR technology and (2) in this domain VR training is increasingly seen to address current hard skills training challenges of timeliness, cost, authenticity, accuracy, and scalability, and thus many industrial organizations are already implementing training using VR technology.

Realizing AI-based tutoring software that can produce richer and more consequential learning in an IVRTE requires both extensive development and experimental studies. In this chapter, we elaborate on a theoretical framework that could inform such work. We will first explore the application of intelligent tutoring systems (ITS) to immersive learning, then review applicable learning theory and conceptualize it within a proposed AI tutor framework, and finally suggest reasonable VR-native pedagogical approaches that could inform empirical research.

2 From ITS to AI Tutor

2.1 Intelligent Tutoring System (ITS)

In the industrial training domain, hard skills learning takes normally place under a human trainer's supervision and control. The need for better scalability calls for computer systems that would allow learners to learn or assess their knowledge and skills by themselves, without human trainer guidance. However, without personalized pedagogical guidance on both selecting the next learning task and completing the task, the learner may not achieve the performance and conceptual learning goals, or fail to do so within a target time, negatively impacting the very scalability. A computer system providing such pedagogical guidance is known as an intelligent tutoring system (ITS).

A large body of research exists since the late 1960s to inform the construction of an ITS (Alkhatlan and Kalita 2018). The canonical structure of an ITS divides its functions between four interconnected modules (Wenger 1987), see Fig. 1. *The expert knowledge module* (or domain model) serves as a repository of expert knowledge about the task being tutored. In the procedural training context, this knowledge, captured from subject matter experts, defines the steps of the procedure to the learned. *The student model module* (or learner model) enables personalized learning by capturing the system's current understanding of the learner's mastery of

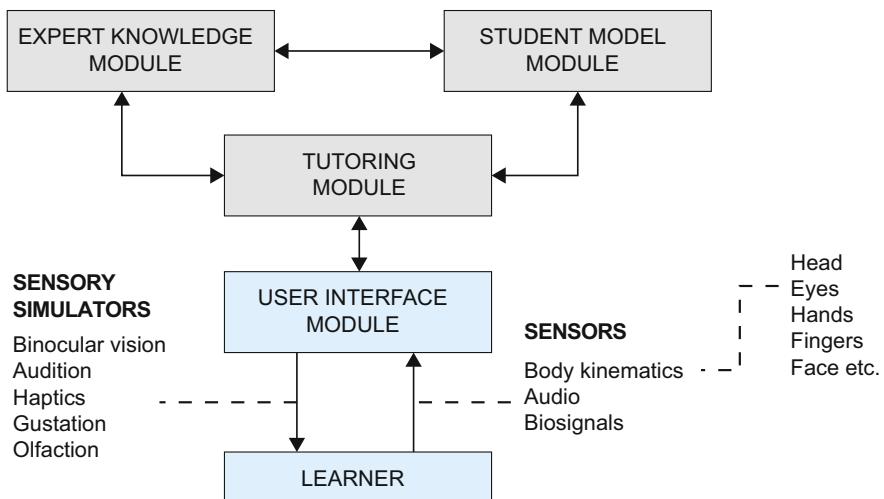


Fig. 1 Traditional ITS architecture. (Adapted from Wenger 1987) augmented to show the interaction modalities with the learner when the user interface module is provided by an IVRTE. Sensors such as head, eye, and face trackers provide the computer information about the learner. Sensory simulators, including head-mounted binocular displays, headphones, and haptic vibrators, simulate sensory experiences for the learner

the domain model tasks and the student's cognitive state. The ITS takes decisions in *the tutoring module*, which, following the tutoring strategies known to the ITS, executes two decision loops: (1) *outer loop*, selecting a task that would best help the learner learn and (2) *inner loop*, guiding the learner by instruction through the right steps constituting a task (VanLehn 2006). Each of these modules has spawned its own rich research topic and literature.

2.2 Observability

To adapt an ITS assuming the canonical structure to an IVRTE, one needs to replace the fourth module, the *user interface module* with the IVRTE user interface (see Fig.1). Recent ITS research has been naturally directed toward the user interface with widest availability, a web browser or mobile app. As such, the input from the learner consists of typed keyboard input, pointing using a mouse and selections through mouse clicks/taps. In addition, directional input through device acceleration sensors has been utilized. While some systems allow user audio input, the predominant method of conversational input is typing. Additional sensors, such as eye tracking or heart rate monitors, have been used in experiments that aim to enrich the student model with information on learner affect, with the aim of implementing the principles of *affective computing* (Picard 1997).

In contrast, a standard IVRTE user interface in 2021 consists of sensors that provide kinematic tracking of the user's head position and rotation, as well as the position and rotation of controllers the user is attached to or holds in each hand. A standard VR headset also includes headphones and a noise-cancelling microphone for audio input. Eye and face tracking as well as heart rate tracking are readily available as commercial options. Tracking of finger joints is available in some hand controllers as well as a camera-based option if controllers are not used. As such, the input from an IVRTE provides much more data than what is utilized by a traditional ITS that tracks learner interactions with a graphical user interface. As the learner's representation in the virtual-physical space is mediated through sensor hardware, the VR environment uniquely affords extensive *observability* of the learner's location, posture, and interaction.

2.3 Modifiability

While the output of a traditional ITS user interface is a two-dimensional page or screen, the output of an IVRTE is generated by devices that simulate the learner's sensory experience. The main modality is vision through a head-mounted binocular display, supported by spatially simulated audio sources and haptic stimulators in the hand controllers. With sufficient presence (Slater and Wilbur 1997), the world sensed by the learner – an imagined sociotechnical space – becomes fundamentally

different compared to real-life experience. As this space is generated by a computer program, it exhibits inherent *modifiability*.

By modifying the learner's simulated experience, they can potentially be assisted in reaching their dynamic zones of proximal development (Vygotsky 1978). Toward that end, tasks and scenarios can be presented with variation, refining their features until sufficient skills are demonstrated. These manifestations of modifiability explain VR's popularity for traditional simulator training targeting special learner groups, such as pilots, astronauts, soldiers, and athletes. A less obvious manifestation of modifiability is the capability to modify the experience in subtle ways to support or scaffold, the learner's cognitive processes during learning tasks.

2.4 AI Tutor

Various IVRTEs have been implemented in the industrial training context, but very few identify using an ITS (Laine et al. 2022). Typical solutions that assume a self-study setting (e.g., Hirt et al. 2019) guide the learner using authored hard-coded logic or branched programming (Pavlik et al. 2013), with no learner-specific adaptation. Some systems repurpose an ITS originally designed for traditional user interfaces (e.g., Ashenafi et al. 2020), limiting its pedagogical capabilities.

Examples of ITSs specifically designed for controlling procedural training in immersive VR do exist, for example, STEVE (Rickel and Johnson 1998) and PEGASE (Buche et al. 2010). However, while these systems achieve impressive functionality in adapting to learner actions, they do that by producing actions based on rules that trigger on changes in world simulation state.

With our notion of an AI tutor, we aim for more meaningful learning than what is possible with such triggers. We look for a framework that would assume a model of learner cognition based on emerging theories of grounded cognition. In such a framework, tutoring logic could modify the learner's experience on a fine-grained level based on its observations of the learner's cognitive state.

3 Grounded Cognition

Learning that takes place in virtual reality is immersive in nature (Dede 2009); this means learning through diving into a simulated environment that provides a strong sense of presence together with affordances of acting and functioning in the artificial environment. The “imagined” property of VR allows us to simulate any immersive physical experience. Such immersion appears to also require expanded ways of understanding the cognitive processes involved in learning. Toward that end, the 4E approach to cognition appears to provide important resources (Newen et al. 2018). This framework assumes that cognition does not only take place in the human

head but that it is distributed (Clark 2003; Pea 1993), i.e., embodied, embedded, enacted, or extended across external tools, processes, structures, and environments. The term 4E cognition, attributed to Mark Rowlands (2010), stands for “embodied, embedded, enacted, and extended (4e) cognition.” The 4E approach on cognition involves a collection of interrelated but also conflicting viewpoints, which highlight the materially and socially distributed aspect of cognition (Pea 1993).

Embodied cognition Investigations of hard skill training highlight the importance of embodied cognition. Skills and their training are inherently dependent on the human body and the tools manipulated, and learned skills become “carved” or “sculpted” into the body (e.g., bicycling and boxing). Embodied cognition succeeded the computational theory of mind (Fodor 1981) that replaced behaviorism in the 1950s. Embodied cognition is anti-dualistic in nature; it claims that psychological processes (“software”) cannot be investigated without the “hardware” that the human body provides. Varela et al.’ (1991, revised 2016) book is commonly seen as a starter for the “embodied cognition movement.” Pioneering research of Pea (1993) and Hutchins (1995) established the distributed cognition approach, which has long roots in sociocultural psychology (Rogoff and Lave 1984; Vygotsky 1978) and philosophy (Clark 2003; Clark and Chalmers 1998). The embodied approach builds on phenomenological tradition of philosophy, such as Merleau-Ponty (1945), according to which cognition is grounded in “lived experiences.” Moreover, many cognitive scientists have rejected the computational theory of cognition according to which human mind processes abstract (“amodal”) symbols independent from the modalities of perception, action, and self-reflection. Knowledge is grounded in sensorimotor routines and experiences (Barsalou 1999, 2008, 2020; Lakoff and Johnson 1999) that forms the basis for language and “wording.” Accumulating behavioral and neural evidence across research on perception, memory, knowledge, language, thought, social cognition, and human development supports this view. Lakoff stated, in his foreword to Bergen (2012), “the ball game is over; the mind is embodied.”

Embodied learning The role of active bodily engagement has been highlighted in learning (Stoltz 2015; Shapiro and Stoltz 2019). It is argued that the practice of teaching [declarative] knowledge first before it can be applied (formalisms first) is rooted in the dualistic view of knowledge; in this view intellectual work is associated with the “mind” and practical work with the “body.” Separating knowledge from activity and application leads easily to inert knowledge that cannot be applied in context. Shapiro and Stoltz (2019, p. 27) anchor embodied learning on an assumption summarized from the Maturana and Varela (1998) account on embodied cognition: “learning is contingent upon the cognitive activity that is triggered by the environment and is determined by the dynamic nature of living beings engaged in the self-organizing activities by which they sustain themselves.” Learning conceptual knowledge should be integrated with firsthand (direct experience) and secondhand (description of experience) experiences and with both physical and imagined manipulation. Anchoring learning on physical

manipulation is critical because it assists experiential grounding of abstract symbols that are used to build embodied mental models (Glenberg 2008). The other three Es (embedded, enacted, extended) are more or less “breaking out” some of the aspects of the original “embodied” thinking into separate areas (see e.g., Newen et al. 2018).

Embedded cognition Embedded cognition may be seen as the aspect of embodied learning that describes how the environment is partially involved in cognitive processing. For instance, when an outfielder in baseball catches a fly ball, it may appear that they are dependent on sophisticated cognitive operations, when in fact they are exploiting features of the environment in a way that reduces cognitive load (Shapiro and Stoltz 2019). Human activity in general takes place in deliberately designed and built cultural environments (e.g., schools, learning labs) fostering learning and development. Embedded cognition can be harnessed by creating artificial worlds open to exploration, designing complex open-ended challenges and tasks that can be worked with virtual physical and semiotic tools, and by manipulating the environment so that desired aspects become opaque or transparent, depending on the purpose. Through deliberate and iterative design efforts, it is possible to create structures, functions, and processes that support training activity, adapt to learners’ developing competences, and foster building and stretching the skill being developed.

Enacted cognition This perspective emphasizes real-time dynamic interaction between a human and the environment as a crucial aspect of cognition. The world is experienced through exploratory sensorimotor interaction with the environment. Learning is not a property of mind or located at a person but enacted through dynamic interaction between learners and environments. Enaction refers to a dynamic process in which a learner adaptively couples their actions to the requirements of unfolding situations. One aspect of enaction is gesturing. Gestures used in conversations (even in telephone calls) may, for instance, be considered as a form of communication (Shapiro and Stoltz 2019). Also, certain gestures may signify the readiness to learn (Shapiro and Stoltz 2019, p. 28). “A living organism enacts the world it lives in; its effective, embodied action in the world actually constitutes its perception and thereby grounds its cognition (Stewart et al. 2010).” From the enactive perspective, learning is not the passive reception of information but involves active and deliberate exploration of the environment, entailing motivation and planning activity and observing and transforming the environment as emphasized by Bruner (1966). Interacting with one’s cultural environment structures experiences according to patterns of sociocultural practices (see e.g., Nasir et al. 2020).

Extended cognition The extended mind thesis assumes that rather than being encapsulated within the brain or the body, cognitive processes extend into the physical world (Clark 2003; Clark and Chalmers 1998). Learners can off-load their cognitive work to the environment (Donald 1991; Wilson 2002), for example, use a paper and pencil as external memory field to support calculation. The human and the

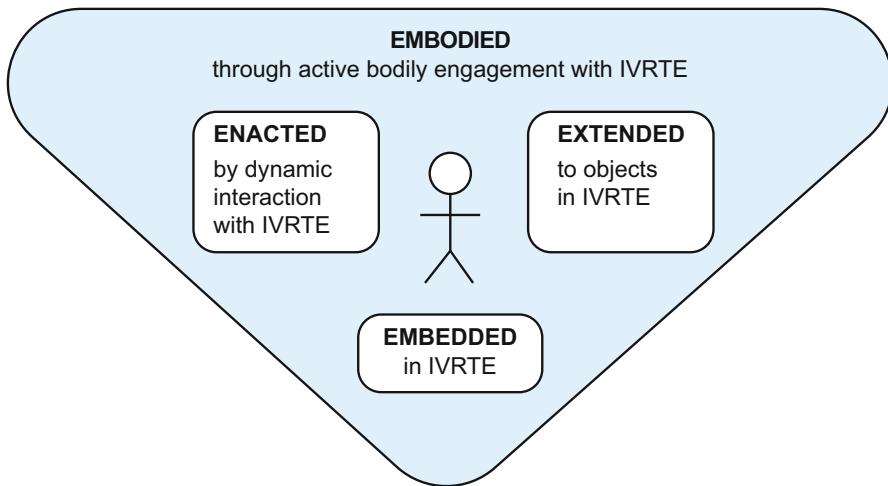


Fig. 2 Summary of 4E cognition for a learner immersed in a task in an IVRTE. The learner's cognition is embodied through their active bodily engagement with the IVRTE. Breaking out aspects of the original embodied thinking, the learner's cognition is embedded in the virtual world generated by the IVRTE, enacted by their dynamic interaction with the virtual world, and extended to objects in the virtual world

environment of their activity develop gradually to support one another and constitute a coupled cognitive system. As far as the IVRTE structures support their activity, such as reminding about the purpose of the tasks, they do not have to invest so much effort in the cognitive task of remembering. The environment can also represent the tools and objects needed for subsequent tasks, as in allowing the learner to pick up the parts and tools they intend to use next. Here the learner is engaged in a developmental process of appropriating and internalizing tools used in the activity to the extent that the tools become a part of their minds (Galperin 1992) and invisible in their hands (they are aware of the object of activity rather than tool that is seamlessly integrated with their activity).

The above examination, summarized in Fig. 2, indicates that learning in general and hard skills learning in particular is an embodied, embedded, enactive, and extended process. While embedded in an IVRTE, the learner does not employ an isolated set of processes. Instead, cognition emerges from interactions of processes in the domains of the modalities, the body, the physical environment, and the social environment with processes traditionally associated with solo cognition, such as knowledge, attention, memory, thought, and language (Barsalou 2020). Barsalou (2020, p. 2) summarizes this interplay of processes as *grounded cognition*:

From the 4E perspective, cognition, affect, and behavior emerge from the body being embedded in environments that extend cognition, as agents enact situated action reflecting their current cognitive and affective states.

It follows that the research and development of digital tools and environments does not represent the creation of neutral and external instruments but may instead radically remediate a learner's cognitive processes; the same concerns also apply to the creation of IVRTEs that reshape embodied, embedded, enactive, and extended processes and provide resources for training. Integration of external tools with the human activity is, however, a developmental process of its own, called instrumental genesis (Rabardel and Bourmaud 2003; Ritella and Hakkarainen 2012). Only after the tools have been seamlessly merged and fused with the human activity system are they likely to enhance various aspects of 4E cognition. Organizational researchers use the concept of sociomateriality (Orlikowski and Scott 2008) to examine how epistemic, social, and material processes of using technologies are intertwined. Such entanglement of technology and human activity also concerns immersive virtual technology.

4 VR-Native AI Tutor Framework

4.1 *Situated Conceptualizations*

To develop a conceptual framework for an AI tutor that could natively utilize observability and modifiability in an IRVTE, we assume the 4E cognition perspective that the learner's cognitive state emerges from interactions between cognitive activity domains in terms of grounded cognition (Barsalou 2020), see Fig. 3. In this perspective, the *physical and social environment* domains of the learner's cognition form conceptualizations of the virtual world they are embedded in. Whether these are represented as amodal symbols or through some other knowledge representation is an open area of research. However, considerable evidence shows that sensory-motor modalities become active as people process conceptual and semantic information, a phenomenon known as multimodal simulation.

Barsalou's (2020) examination of the accumulation of memories that underlie skill acquisition inspires the following example of how a learner in an IVRTE could form a multimodal simulator for the concept of electric screwdriver. When a learner encounters a task requiring a tool, their cognitive processes in different modalities that would normally process the tool's features become active. These can include how the tool looks (vision) and what it feels like (tactile). Importantly, these activations are not only limited to static ontological representations of the tool concept but span multiple domains of cognition that participate in the cognitive processing while a person is working with the tool. Barsalou offers the "situated action cycle" as one account of the involvement of different cognitive domains in the sequence of processing phases from observing the environment to taking action and ultimately reaching an outcome (reward, punishment, prediction error). According to this account, *situated conceptualizations* are formed in memory during the processing cycle, recallable when the cycle runs again in similar manner (Barsalou 2020).

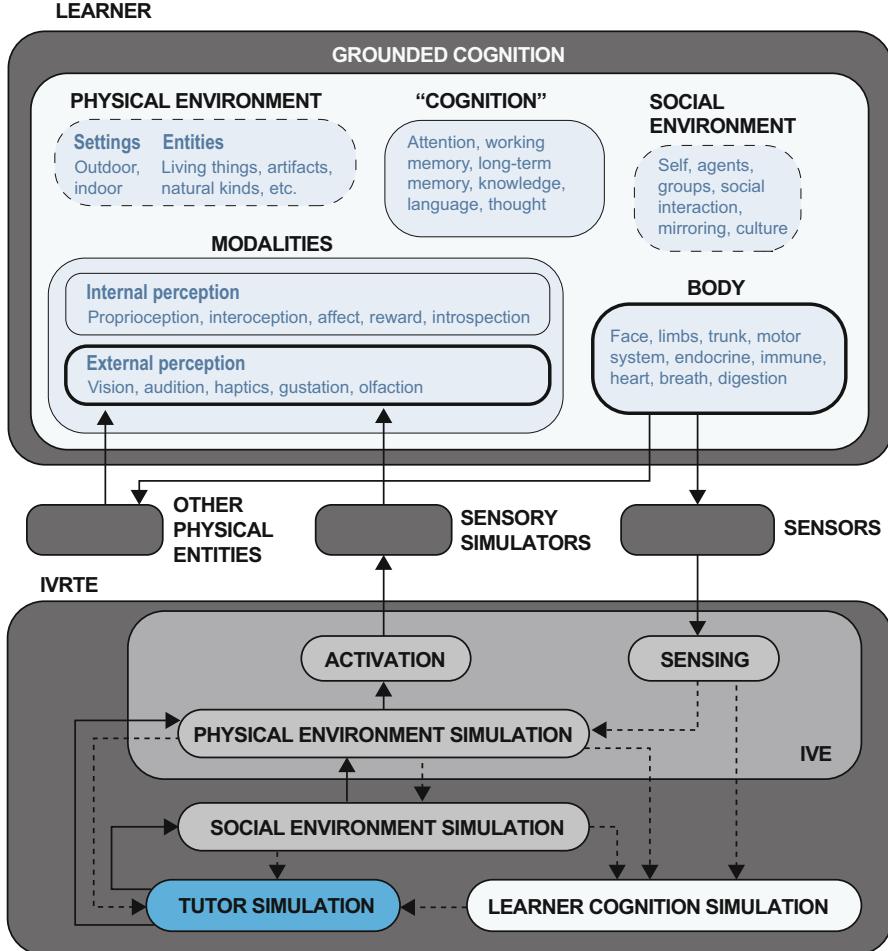


Fig. 3 Domains of grounded cognition. (Adapted from Barsalou (2020) for a learner mapped to the IVRTE functions that attempt to simulate and sense them. The conceptualizations in the physical and social environment domains arise through grounded simulators (Barsalou 1999). The learner's external perception is partially replaced by simulated sensory perceptions generated by a physical environment simulation generated through sensory simulator devices, with input from sensors that quantify the learner's body kinematics. This part of the IVRTE constitutes a minimal IVE. The social environment experienced by the learner is formed through physical environment percepts generated by a social environment simulation. A simulated tutor adapts the physical and social environments for the learner, based on a simulation of the learner's cognition informed by sensing of the physical and social environments and additional sensors. Dashed arrows indicate inputs)

4.2 Physical Environment Simulation

To outline a systemic view of the interaction between the learner and the proposed components of an IVRTE featuring an AI tutor (see Fig. 3), we first recognize that the learner's cognition must necessarily interface with the external world through the learner's body, which provides for action and mediates the external modalities. The body interacts physically with *sensory simulators* provided by the IVRTE, primarily the binocular vision simulators (displays), that activate external perception. The sensory simulation is generated in software controlled by *sensors* that sense the learner's body kinematics, creating an illusion of the virtual-physical space. This part of the IVRTE, providing a *physical environment simulation*, essentially describes any VR-based immersive virtual environment (IVE).

4.3 Specifiers

Physical environment simulation elicits external perception activations that, through interacting cognitive processes, form the learner's perceived environment. However, the same IVRTE simulation may not result in the same concept in the learner unless it also incorporates features that sufficiently activate all cognitive domains that contribute properties of the concept. To invoke or form a situated conceptualization, the physical environment simulator in the IVRTE should thus be instructed to add physical phenomena with features that would be expected to activate the multimodal sensory experiences.

For the social environment, such additions are provided as part of the *social environment simulation*. To elicit recall of a social situation, it may not suffice to show the appropriate visual representations we normally associate with the situation (such as avatars for the participants). In addition, the social simulation may need to instruct the simulation of physical representations such as additional objects, sounds, or interactions that for an outside observer would seem to be extraneous but which, when perceived by the learner, would be essential for invoking the correct situated conceptualization.

We call these extra pieces of simulation added for the purpose of forming the desired cognitive state "specifiers," as without their presence the situated conceptualizations formed in the learner in response to the physical simulations may remain unspecific, differing considerably from what was intended for the purpose of supporting learning. Specifiers need not be purely visual; for example, additions to the simulation that elicit gesturing action may offer a way to guide the learner's cognition toward the intended situated conceptualization (Goldin-Meadow 2011).

4.4 Learner and Tutor Simulations

The responsibility to add the correct specifiers should lie with a function that models the learner's grounded cognition state. The *learner cognition simulator* provides this model, utilizing cues from the current physical and social environment simulation states as well as from non-kinetic sensing inputs.

The remaining function, which we call a *tutor simulation*, is analogous to the tutoring module in a traditional ITS. Based on the current state of the simulations, the tutor simulation instructs the creation of appropriate specifiers needed to invoke the correct conceptualizations. Extending the terminology of traditional ITSs, we denote the tutor simulation as operating in the *innermost loop*, compared to the *inner* (guiding through task steps) and *outer* (selecting tasks) loops of the traditional ITS. The target of this additional loop is to select from a repertoire of specifiers the ones that are most likely to elicit the intended situated conceptualizations in the current learner, allowing the tutor simulation to build or modify situated conceptualizations that may be necessary and/or sufficient for skill acquisition.

ITs providing a conversational interface that mimics the conversation between a learner and a human tutor have achieved significant improvements in learning effectiveness. The most well-known of such efforts is AutoTutor (Graesser et al. 1999). In an IVRTE, such an interface could be implemented as part of the social environment simulation. Learner utterances recognized by the simulation could be used as inputs for the learner cognitive model. Correspondingly, when selecting specifiers that would elicit the desired situated conceptualization, the tutor could instruct the social environment simulation to produce the appropriate conversational utterances. Here it should be noted that while typing is impractical with current VR technology, we may be able to infer "mute" learner such as hedges, pauses, and disfluencies, which allow the tutor to infer more information about learner cognition (Pon-Barry et al. 2004). This approach could work especially in our domain (industrial setting) where learners may not be comfortable with having a conversation with a computer. Any conversational approach should consider the cultural traditions of the learning domain (compare Pea 2004).

4.5 Implementing the Framework

The functional arrangement described above could form the basis for the implementation of VR-native pedagogical agent software or AI tutor. Deep learning-based AI methods promise powerful ways to implement key parts of the framework. The extensive sensor data already used to inform the physical environment simulation can be utilized to train machine learning models that may be able to recognize specific learner cognitive states. Additional sensors such as eye and face trackers as well as bio-signal sensors may improve the models.

Experimental results suggesting the feasibility of making inferences from learner cognitive state using sensor data are already available. User body tracking data has

been used to identify individual users (Miller et al. 2020, Moore et al. 2021). Pfeuffer et al. (2019) identified characteristic behavior for users in VR from monitoring their head, hand, and eye motion data. Holzwarth et al. (2021) correlated head yaw in VR with user's affective state. Won et al. (2014) were able to automatically distinguish between low and high success learning interactions by monitoring body movement. Marín-Morales et al. (2018) used electroencephalography (EEG) and electrocardiography (ECG) sensors to distinguish between emotional states of users embedded in a virtual environment. Hussain et al. (2011) used machine learning methods to detect learners' affective states from multichannel physiological data, including heart rate, respiration, facial muscle activity, and skin. In social psychology, VR-based behavioral tracing has been operationalized for quantifying social approach and avoidance, evaluation of a social other, and engagement and attention (Yaremych and Persky 2019).

5 Toward VR-Native Pedagogy

In this section we provide a preliminary outline of principles based on the proposed framework that can guide the pedagogical design of an IVRTE and its AI tutor functionality.

5.1 *Simulation Environment*

The physical environment simulation in an IVRTE for procedural learning is built to simulate the mechanisms and causal relationships involved in the procedure. The extent to which a simulator attempts to imitate the real world is determined by task analysis. Time and cost concerns often necessitate the prioritization of simulating the parts of the environment the learner is most likely to interact with. However, the learner should be able to freely manipulate the simulation toward the desired end state of the procedure, possibly taking pathways that prevent further progress or cause known problems.

The highest achievable fidelity (both in terms of visual and task fidelity) may not always be desired. While a higher-fidelity simulation adds to learner presence (Dalgarno and Lee 2010), it may impact learning negatively from the grounded cognition point of view as the learned situated conceptualizations may not transfer to semantically similar but different situations exhibiting altered details. When designing specifiers that can be added to the physical world simulation, one consideration is the learner's emotional and aesthetic engagement with the world (Stolz 2015). The VR environment should simulate professionally adequate ways of working with tools. An object becomes instrument (and, therefore, “invisible” tool in hand) only through learning and internalizing the IVRTE system (instrumental

genesis); when disturbances or breakdowns occur, the instrument, again, becomes an object of deliberate inspection (Engeström 1987).

If a learner can achieve the desired performance just by interacting with the IVRTE simulation and the simulation has been implemented to account for the failure modes identified during task analysis, the learner has effectively demonstrated their possession of the targeted knowledge and skills. Such a simulation with no tutoring actions can still make use of the inherent observability of the VR environment by producing a detailed analysis of the learner performance, as well as suggestions for improvement where the learner has exhibited weaker results.

5.2 Task Sequencing

Grounded cognition principles can be already considered when the tutor is selecting the next tasks for the learner from the available tasks created by the instructional designer (ITS outer loop). Existing instructional design guidelines prescribe a theory-first approach (Fowler 2015). However, this approach may not allow the learner to benefit from the IVRTEs' ability to ground the theoretical concepts as part of the learner's situated conceptualizations. We may be able to get better results by transforming theory topics into experiences where the learner engages in goal-directed but open-ended operational procedures anchored on their cognitive domains. Grounded cognition emphasizes the importance of affording the learners the opportunities to be active in a congruent way, i.e., allow and encourage movements and interactions that resemble the actual operational procedures and mechanisms (Johnson-Glenberg 2018). Whenever the learner thinks about something (tries to build a mental model or solve a problem), their cognitive process is impacted by the virtual environment they are located in and the affordances they interact with (Newen et al. 2018).

5.3 Scaffolding

In normal circumstances a learner is not expected to succeed in the IVRTE simulation without external help. Thus, the key function for the inner loop becomes the selection of appropriate scaffolding actions for the learner. The scaffolding provides structures and guides the learners' activity without necessarily prescribing only one or a few "correct" lines of activity. Accordingly, there are likely to be several pathways to the desired learning outcome (reaching the currently targeted step completion). It is also critical to engage the learners themselves in agentic efforts of analyzing situations, selecting promising lines of activity, and assessing their advancements of their efforts. The effectiveness of any proposed scaffolding actions needs to be assessed for their impact on learner performance by design-based or experimental research.

As pointed out by Pea (2004), scaffolding is a complex theoretical concept related to relations between people, tools, and environments (Engeström 1987) rather than anchored on analyses of disconnected cognitive tasks. If we subscribe to the grounded cognition perspective and model the learner through a simulation of the learner cognitive state from observational data (learner model), the scaffolding activation function within the tutor simulation should map the cognitive state and the desired domain model state to scaffolding actions. To implement such a function, it becomes necessary to express the domain model using concepts that are compatible with the learner model. Thus, the inner loop could consider what the learner has already experienced and what should they experience next – following, for instance, the theory of comprehensive learning by Jarvis (2012). Notice that this does not contain an assumption of one normatively corrected performance because there can be multiple pathways to the targeted objective.

In an industrial setting, work instructions and other task and performance support provide distributed cognitive resources (Pea 1993); such resources include manuals, labels, checklists, and affordances of tools that prevent them from being used in the wrong ways. A key function of such resources is quality control, but, simultaneously, they may also be used to scaffold the learning. Sometimes scaffolds are a part of the procedural instruction, but professionals tend to adapt and “devise their own aides” by arranging their tools, materials, and workspaces.

As the learner demonstrates through the learner model that a specific scaffold is no longer needed, the scaffold is faded, and the learner is expected to continue achieving the same performance without the scaffold. As follows, we list possible VR scaffolds that could be tested:

- Abstracting the world (making physics, mechanisms, structures, affordances simpler)
- Automating things (mechanisms, other person’s actions, processes etc.)
- Providing the learner new abilities (X-ray vision)
- Augmenting the world with text and graphics overlaid on objects
- Extending the learner’s body (remote manipulation of objects)
- Allowing the learner to move large distances effortlessly (teleport)
- Allowing the learner to manipulate large or heavy objects
- Selectively silencing or emphasizing sounds
- Emphasizing or hiding objects and affordances

Pea (2004) asks a good question – if a scaffold improves learning, why should it become faded? Why cannot scaffolds just become an aspect of accepted performance support and a part of the distributed system of intelligence? The inherent modifiability of VR provides a straightforward answer to this question, and a key principle for designing scaffolds for an IVRTE; to improve learning beyond what can be done by non-fading performance support, an IVRTE should aim to primarily provide “impossible” scaffolds, actions, and events that could not be implemented in the real world. These scaffolds must fade as they cannot be realized in the real world to continue providing the relevant performance support. What are such impossible scaffolds? The exciting opportunity of VR technology is that within the bounds of achievable presence, anything can be implemented and tested.

6 Discussion

In an article titled “Where’s the pedagogy?,” Fowler (2015) calls for working out missing pedagogical principles in VR-based self-study training solutions. The suggested solution is to add pedagogy through a step-by-step design process. Similarly, although recognizing the unique modifiability afforded by VR, Johnson-Glenberg (2018) focuses on giving guidelines on how to design better VR learning experiences. In general, there may exist a tendency to address VR technology as another medium to apply the “pedagogically well-designed interaction” tradition from web-based self-study and, going further back, from the proper organization of textbooks. However, this approach may fail to produce the learning results expected from the increasingly complex simulations of real-world tasks and associated hard skills training afforded by IVRTEs. In these contexts, we should also ask “where is the teacher?” and focus more on automatic systems that can support the learner in a personalized manner through real-time scaffolding decisions.

Determining the learning benefits of an VR-native ITS that utilizes the observability and modifiability of the VR environment requires design-based and experimental work on the ability to automatically infer learner cognitive state and the situational scaffolding needs from real-time sensor data. Also, further research and development work is needed to assess the learning benefits of any proposed automatic scaffolding interventions based on the general principles presented. Our work resides at an intersection between IT, psychology, and learning science. Each field is approaching experiments on VR technology from its tradition, which complicates the interpretation and application of existing experimental results.

The conceptual work is not without challenges either. Despite large evidence of the existence of grounded principles of human cognition, an understanding of actual working of the cognitive principles, for example, how knowledge is represented under these premises, remains as elusive as ever. On the other hand, a full account of the mechanisms underlying grounded cognition may not be necessary for practical applications of the concept, as demonstrated by the largely unobservable inner workings of highly useful deep neural network architectures.

To the extent that the presented AI tutor framework proves implementable and its theoretical underpinnings have merit, one must also raise the question of ethical use of such technology. Should a VR-native tutor implementation prove to be capable of modifying situated conceptualizations for skill acquisition, it may be able to modify such conceptualizations for any other purpose. Those purposes may be highly beneficial (modifying adverse habitual learning) but also questionable (making learning overly dependent on tutoring software).

Further, we should examine how IVRTE mediated hard skills learning complements conventional training with human educators. A critical concern is to what extent VR training transfers to working with conventional tools and instruments and how VR and regular training support one another. We expect VR training to assist learners in developing the orienting basis (Galperin 1992) for training and further refining their vocational skills. Our work is focused on a specific domain

(procedural training in industrial settings), which may not always present many of the typical challenges faced in other educational settings (e.g., social interaction, developmental psychology, abstract concept formation). However, the core insights of the work may be applicable to other educational settings.

Acknowledgements Our work is supported by the AI Learn project coordinated by the University of Helsinki and funded by the Finnish innovation funding agency Business Finland.

References

- Alkhatlan, A., & Kalita, J.K. (2018). Intelligent tutoring systems: A comprehensive historical survey with recent developments. ArXiv, abs/1812.09628.
- Ashenafi, M.M., Wolf, S.D., Barbosa, M., Jones, N., & McLaren, B. (2020). Work-in-progress: A generalizable virtual reality training and intelligent tutor for additive manufacturing. In *2020 6th International Conference of the Immersive Learning Research Network (iLRN)*, 355–358.
- Barsalou, L.W. (1999). Perceptual symbol systems. *Behavioral and Brain Sciences*, 22, 577–609.
- Barsalou, L.W. (2008). Grounded cognition. *Annual Review of Psychology*, 59, 617–645.
- Barsalou, L.W. (2020). Challenges and opportunities for grounding cognition. *Journal of Cognition*, 3, 1–24.
- Bergen, B. K. (2012). *Louder than words: The new science of how the mind makes meaning*. Basic Books (AZ).
- Bruner, J. S. (1966). *Toward a theory of instruction* (Vol. 59). Harvard University Press.
- Buche, C., Bossard, C., Querrec, R., & Chevallier, P. (2010). PEGASE: A generic and adaptable intelligent system for virtual reality learning environments. *Int. J. Virtual Real.*, 9, 73–85.
- Clark, A. (2003). *Natural-born cyborgs: Minds, technologies intelligence*. Oxford University Press.
- Clark, A. & Chalmers, D. (1998). The extended mind. *Analysis*, 58(1), 7–19.
- Dalgarno, B. & Lee, M. (2010). What are the learning affordances of 3-D virtual environments? *British Journal of Educational Technology*, 41, 10–32.
- Dede, C. (2009). Immersive interfaces for engagement and learning. *Science*, 323, 66–68. <https://doi.org/10.1126/science.1167311>
- Donald, M. (1991). *Origins of the modern mind*. Harvard University Press.
- Engeström, Y. (1987). *Learning by expanding an activity – theoretic approach to developmental research*. Helsinki: Orienta-Konsultit Oy.
- Fodor, J. (1981). Methodological solipsism considered as a research strategy in cognitive psychology. In J. Fodor (Ed.), *Representations: Philosophical essays on the foundations of cognitive science* (pp. 225–253). MIT Press.
- Fowler, C. (2015). Virtual reality and learning: Where is the pedagogy? *British Journal of Educational Technology*, 46(2), 412–422.
- Galperin, P. I. (1992). Stage-by-stage formation as a method of psychological investigation. *Journal of Russian and East European Psychology*, 30(4), 60–80.
- Glenberg, A. (2008). Embodiment for education. In: Calvo, P. and Gomila, A. (Eds.), *Handbook of cognitive science: An embodied approach*, pp. 355–371. London: Elsevier Science.
- Goldin-Meadow, S. (2011). Learning through gesture. *Wiley interdisciplinary reviews. Cognitive science*, 2(6), 595–607.
- Goodfellow, I.J., Bengio, Y., & Courville, A.C. (2015). Deep Learning. *Nature*, 521, 436–444.
- Graesser, A.C., Wiemer, K., Wiemer-Hastings, P.M., & Kreuz, R.J. (1999). AutoTutor: A simulation of a human tutor. *Cognitive Systems Research*, 1, 35–51.
- Hirt, C., Holzwarth, V., Gisler, J., Schneider, J., & Kunz, A.M. (2019). Virtual learning environment for an industrial assembly task. In *2019 IEEE 9th International Conference on Consumer Electronics (ICCE-Berlin)*, 337–342.

- Holzwarth, V., Schneider, J., Handali, J.P., Gisler, J., Hirt, C., Kunz, A.M., & vom Brocke, J. (2021). Towards estimating affective states in Virtual Reality based on behavioral data. *Virtual Reality*, 1–14.
- Hussain, M. S., AlZoubi, O., Calvo, R. A., & D'Mello, S. K. (2011, June). Affect detection from multichannel physiology during learning sessions with AutoTutor. In *International Conference on Artificial Intelligence in Education* (pp. 131–138). Springer, Berlin, Heidelberg.
- Hutchins, E. (1995). *Cognition in the wild*. MIT Press.
- Jarvis, P. (2012). *Towards a comprehensive theory of human learning*. Routledge.
- Johnson-Glenberg, M. (2018). Immersive VR and education: Embodied design principles that include gesture and hand controls. *Frontiers in Robotics and AI*, 5, 81.
- Laine, J., Lindqvist, T., Korhonen, T., & Hakkarainen, K. (2022). Systematic Review of Intelligent Tutoring Systems for Hard Skills Training in Virtual Reality Environments. *International Journal of Technology in Education and Science*.
- Lakoff G. and Johnson M. (1999) *Philosophy in the flesh: The embodied mind and its challenge to Western thought*. New York: Basic Books.
- Marín-Morales, J., Higuera-Trujillo, J.L., Greco, A., Guiñez, J., Llinares, C., Scilingo, E.P., Alcañiz, M., & Valenza, G. (2018). Affective computing in virtual reality: Emotion recognition from brain and heartbeat dynamics using wearable sensors. *Scientific Reports*, 8(1), 1–15.
- Maturana H. and Varela F. (1998). The tree of knowledge: *The biological roots of human understanding*, revised edition. Boston, MA: Shambhala.
- Merleau-Ponty, M. (2013, originally published 1945). *Phenomenology of perception*. Routledge.
- Miller, M.R., Herrera, F., Jun, H., Landay, J.A., & Bailenson, J.N. (2020). Personal identifiability of user tracking data during observation of 360-degree VR video. *Scientific Reports*, 10(1), 1–10.
- Moore, A.G., McMahan, R.P., Dong, H., & Ruozzi, N. (2021). Personal identifiability of user tracking data during VR training. In *2021 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW)*, 556–557.
- Nasir, N.S., Lee, C.D., Pea, R.D., & De Royston, M.M. (Eds.). (2020). *Handbook of the Cultural Foundations of Learning*. New York: Routledge.
- Newen, A., De Bruin, L. & Gallagher, S. (2018). 4E cognition: Historical roots, key concepts, and central issues. In A. Newen, L. De Bruin., & S. Gallagher (Eds.), *The Oxford handbook of 4E cognition* (pp. 1–16). Oxford University Press.
- Orlikowski, W.J., & Scott, S.V. (2008). 10 Sociomateriality: Challenging the Separation of Technology, Work and Organization. *The Academy of Management Annals*, 2, 433–474.
- Pavlik, P. I., Brawner K., Olney, A., & Mitrovic, A. (2013). A review of student models used in intelligent tutoring systems. In R. Sottilare, A. Graesser, X. Hu, & H. Holden (Eds.), *Design recommendations for intelligent tutoring systems: Volume 1-Learner Modeling* (pp. 135–142). US Army Research Laboratory.
- Pfeuffer, K., Geiger, M.J., Prange, S., Mecke, L., Buschek, D., & Alt, F. (2019). Behavioural biometrics in VR: Identifying people from body motion and relations in virtual reality. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (pp. 1–12).
- Pea, R. D. (1993). Practices of distributed intelligence and designs for education. In G. Salomon (Ed.), *Distributed cognitions: Psychological and educational considerations* (pp. 47–87). Cambridge University Press.
- Pea, R.D. (2004). The social and technological dimensions of scaffolding and related theoretical concepts for learning, education, and human activity. *Journal of the Learning Sciences*, 13, 423–451.
- Picard, R.W. (1997). *Affective Computing*. MIT Press.
- Pon-Barry, H., Clark, B., Schultz, K., Bratt, E.O., & Peters, S. (2004). Advantages of Spoken Language Interaction in Dialogue-Based Intelligent Tutoring Systems. In *International Conference on Intelligent Tutoring Systems* (pp. 390–400). Springer, Berlin, Heidelberg.
- Rabardel, P., & Bourmaud, G. (2003). From computer to instrument system: A developmental perspective. *Interacting with Computers*, 15, 665–691.

- Rickel, J. & Johnson, W. (1998). STEVE: A Pedagogical Agent for Virtual Reality. In *proceedings of the Second International Conference on Autonomous Agents*, ACM Press, Minneapolis/St. Paul.
- Ritella, G. & Hakkarainen, K (2012). Instrumental genesis in technology mediated learning: From double stimulation to expansive knowledge practices. *International Journal of Computer-Supported Collaborative Learning*, 7, 239–258 (<https://doi.org/10.1007/s11412-012-9144-1>).
- Rogoff, B. E., & Lave, J. E. (1984). *Everyday cognition: Its development in social context*. Harvard University Press.
- Rowlands, M. (2010). *The new science of the mind: From extended minds to embodied phenomenology*. MIT Press.
- Shapiro, L., & Stoltz, S. A. (2019). Embodied cognition and its significance for education. *Theory and Research in Education*, 17(1), 19–39.
- Slater, M., & Wilbur, S. (1997). A Framework for Immersive Virtual Environments (FIVE): Speculations on the Role of Presence in Virtual Environments. *Presence: Teleoperators & Virtual Environments*, 6, 603–616.
- Stewart, J., Gapenne, O., & Di Paolo, E. (2010). Introduction. In J. Stewart, O. Gapenne & E. Di Paolo (Eds.). *Enaction: The new paradigm for cognitive science*. Cambridge, MA: The MIT Press, 7–17.
- Stoltz, S. A. (2015). Embodied learning. *Educational Philosophy and Theory*, 47(5), 474–487.
- VanLehn, K. (2006). The behavior of tutoring systems. *Int. J. Artif. Intell. Educ.*, 16, 227–265.
- Varela, F., Thompson, E., & Rosch, E. (1991). *The embodied mind: Cognitive science and human experience*, Revised edition. The MIT Press.
- Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes*. Harvard University Press.
- Wenger, E. C. (1987). *Artificial intelligence and tutoring systems: Computational and cognitive approaches to the communication of knowledge*.
- Wilson, M. (2002). Six views of embodied cognition. *Psychonomic Bulletin and Review*, 9: 625–636.
- Won, A.S., Bailenson, J.N., & Janssen, J. H. (2014). Automatic detection of nonverbal behavior predicts learning in dyadic interactions. *IEEE Transactions on Affective Computing*, 5, 112–125.
- Yaremych, H.E., & Persky, S. (2019). Tracing physical behavior in virtual reality: A narrative review of applications to social psychology. *Journal of Experimental Social Psychology*, 85.

Open Access This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.



Multiple Users' Experiences of an AI-Aided Educational Platform for Teaching and Learning



Shuanghong Jenny Niu, Xiaoqing Li, and Jiutong Luo

Contents

1 Introduction	215
2 Study Background and Research Questions	216
3 Description of the AI-Aided SLP Educational Platform	217
4 Methodology	221
5 Findings	222
5.1 Students' Self-Reported Experiences	222
5.2 Teachers' and Principal's Self-Reported Experiences	224
6 Discussion and Learning from This Case Study	226
6.1 Major Functions Favored by Students and Teachers	226
6.2 New Ways and Possibilities in Learning and Teaching	228
6.3 Positive Experiences and Changes	228
6.4 The Importance of Learning Theories Applied in AIED Applications	228
6.5 Continuous Improvements and the Social Nature of Learning	229
7 Conclusion and Recommendation	229
References	230

1 Introduction

Currently artificial intelligence (AI) has attracted enormous attention in the media and in public discussion. AI has had a huge impact on societies, organizations, work, and education. Applying AI in learning and education has a long history, going back to at least the 1960s (Minsky and Papert 1968). Driven by the fast advancement of AI technologies, many new ways and possibilities were found to apply AI in education and in supporting students' learning. How to use AI technologies to better support teaching and learning has become one of the main developments in the educational field.

S. J. Niu
University of Helsinki, Helsinki, Finland
e-mail: Jenny.niu@helsinki.fi

X. Li · J. Luo (✉)
Beijing Normal University, Beijing, China

There have been several common AI-associated themes widely used in education, such as robot teachers, intelligent tutoring systems (ITS), massive online learning courses (MOOCs), etc. (Stone et al. 2016). These applications have been widely used in education throughout the world. A typical scenario of these applications is a student working with a digital device to solve or learn domain-level knowledge (e.g., VanLehn 2006). However, this kind of use case does not sufficiently reflect the recent development in practices and theories of education, such as the learning of skills and competencies, students' motivation and agency, the importance of social interaction, and the active role of learners. Additionally, compared to unified "one-size-fits-all" courses, there is an urgent requirement for individualized and/or various ways of teaching and learning based on students' needs and strengths. Therefore, both students and teachers are in need of better personalized support and social interactive learning environments in AI-aided platforms in learning and teaching.

Furthermore, there is also a major concern how to utilize the available educational resources to benefit more schools, especially schools in less advanced areas. The fast development in information communication and AI technology creates possibilities to provide high-quality educational resources to a large number of schools. In this way more students and schools can have a chance to access high-quality educational resources even in less advanced or less developed areas. The current educational platform should be created by using AI technology to meet these needs.

The purpose of this study is to investigate the experiences of students, teachers, and principals in using an AI-aided educational platform and their suggestions for future platform development. This chapter consists of sections of the background, methodology, findings, discussions, learning, and recommendations.

2 Study Background and Research Questions

AI technology has been widely used in many fields, as well as in learning and education. Lorenz and Saslow (2019) refer to AI as "the scientific pursuit of teaching machines to think like humans, or more simply, the automation of cognitive processes." Lorenz and Saslow (2019) consider machine learning (ML) to be a subdiscipline of AI. Renz and Hilbig (2020) state that ML consists of "data and learning algorithms that are fed into a software program able to create patterns, summaries, or conclusions about certain phenomena." Renz and Hilbig (2020) believe that "ML is only possible if big datasets are available." Gartner (2012) defines big data as "high-volume, high-velocity, and high-variety information assets that demand cost-effective, innovative forms of information processing for enhanced insight and decision making." ML and big data are the basic conditions for AI-supported applications or platforms. In the last few decades, the use of AI, especially ML and big data with educational methods, has grown rapidly in AI tutoring system (ITS). This enables ITS to provide customized tutoring functions based on learners' needs (Keleş et al. 2009).

Many studies (e.g., Baker and Inventado 2014; Fischer et al. 2020) show that ML, LA, big data, and educational data mining (EDM) have been important tools for personalized learning and assessment tools in the current use of AI in education (AIED). Several researchers (e.g., Labarthe et al. 2018; Renz et al. 2020) point out that AIED, LA, and EDM are the essential concepts of technology-enhanced learning by using available digital data and the results of analyzing the data to provide more options and improve the quality of education. The main applications of AIED are to provide intelligent agents and tutors services through AI-supported platforms (Alexander et al. 2019; Labarthe et al. 2018; Renz et al. 2020).

There are numerous studies on the system design and functions of AI-supported ITS. In a systematic overview of 57 papers related to ITS (Mousavinasab et al. 2021), researchers found that the major factors examined in those papers were applied AI techniques, the purpose of AI techniques, learners' characteristics, educational fields, evaluation, and user interface of ITS. However, there is fewer research investigating the multiple users' experiences. In this study, we will introduce the AI-aided Smart Learning Partner (SLP), which is designed as an ITS with AI technology to support teaching and learning at schools. The SLP educational platform adopts a number of AI technologies, and its design uses a number of pedagogical and learning theories. The aim of this study is to investigate multiple users' experiences of using SLP which support teaching and learning at schools. We can learn from these cases, and the learning can be used for future development. In this case study, we focus on the students, teachers, and the school principal's self-reported experiences using the AI-aided SLP. The research questions are the following:

1. In what ways does the AI-aided SLP platform assist students in their learning? What are students' self-reported experiences of using this platform?
2. In what ways does the AI-aided SLP platform assist the teachers and the principal in their work at school? What are the teachers' and the principal's self-reported experiences of using this platform?
3. What can we learn from the students', teachers', and the school principal's self-reported experiences for further development of AI-aided educational platforms?

In the next section, this AI-aided SLP educational platform will be presented. The methodology will also be described, followed by the main findings from this case study. Finally, the conclusions and recommendations are given.

3 Description of the AI-Aided SLP Educational Platform

In this section, we will explore the purpose, design structure, functions, and the current uses of the SLP platform. The case description is based on the materials, documents, and articles as well as interviews from the platform designing and developing team.

This AI-aided SLP educational platform has been developed by the Advanced Innovation Center for Future Education at Beijing Normal University. It can be easily accessed by students and teachers using any smart device, such as computers, iPads, and mobile phones. According to the data from the platform retrieved on 1st of June 2021, there were over 200 schools that were using SLP in five different provinces in China. Over 20,000 teachers and more than 250,000 students have used the platform. To have a better understanding of the SLP platform, we conducted in-depth interviews with four SLP platform designers, developers, and researchers. Additionally, we also investigated 35 documents (platform descriptions, PowerPoint presentation slides, journal articles, user experience reports, etc.) which gave detailed descriptions of the platform. We sought to identify the main purposes, major functions, and ways to support the students' learning in the SLP platform from the developers' perspectives.

Based on the interviews with the designers and developers of SLP, we identified two main purposes for creating the AI-aided SLP educational platform. One purpose is to expand possible ways of teaching and learning, especially providing additional resources for students' self-study, and for individualized teaching and learning. Another purpose is to provide more educational resources to schools, especially to bring high-quality educational resources to schools located in less advanced locations. This includes exurban or rural areas that have fewer teachers and lack high-quality educational resources. As one of the SLP platform designers stated in an interview:

We intend to use AI and ICT technology to provide more possibilities to students and teachers. On the one hand, we strive to build a database with high quality educational assessment tools and resources created by the best teachers and educators. These high-quality educational materials can be utilized by any Chinese schools regardless of their locations. On the other hand, students' real inputs are collected and analyzed to construct individual students' learning reports that include several dimensions, such as knowledge and competencies, strengths, and weaknesses, learning paths and learning progress . . . these kinds of learning reports can be used either by teachers or students for the students' further development.

Technically, this SLP platform adopts machine learning techniques to build the student model, especially the knowledge-tracing model for estimating the individual students' knowledge proficiency at the concept level (Chen et al. 2018). Furthermore, the specifically designed algorithms have been deployed to recommend the multimodal learning resources. Graph convolutional network models have been designed to grade both text-answer math questions and formula-answer questions (Tan et al. 2020). In addition, a cognitive graph is used to support the learner's self-awareness and reflective thinking, which consists of a proper form of knowledge representation and the individual learner's cognitive status (Pian et al. 2019). Recently, the SLP research team has attempted to adopt explainable AI techniques to better support and interpret different decisions made by the platform (Lu et al. 2020). Besides the desktop and mobile version, the SLP educational platform also provides the robot version. Lu et al. (2018) state that the robot version "provides the personalized learner-robot interaction services by leveraging on the

latest techniques, typically including the conversational agent, question-answering system and emotion recognition." (pp. 447).

This SLP platform is intended as a learning assistant at school (Lu et al. 2018). The platform provides different levels of resources which satisfy different learners' needs and competency levels. It also periodically gives positive feedback when learners make progress in their learning topics or tasks. The platform enhances learners' relatedness to the platform through a conversational agent which can chat with the learners. All the assessment tools are built on Bloom's learning pyramid at various levels (Bloom 1956). The learning reports show the students' learning capabilities in remembering, understanding, applying, analyzing, evaluating, and creating. An adaptive learning cognitive map model (Wan and Yu 2020) is also applied in this platform. The platform continuously adjusts appropriate learning resources and recommendations with learning contents, learning activities, learning paths, and learning partners to the learners based on the learners' knowledge structure and cognitive state. Therefore, this platform has used several learning theories to increase learners' motivation, active role and agency, progressive learning, and competencies when using the platform.

This platform in its block diagram has two modules (see Fig. 1). One is the *data aggregation module* which refers to how the data are collected and managed in the platform. It can construct a personalized knowledge graph according to the students' personal assessment results and the interaction data. Another is the *human-machine (learner-machine) interaction module*. It is mainly in charge of how the human interacts with the platform. These two modules establish the block diagram of this SLP platform.

The data aggregation part in this SLP platform continuously collects educational data and resources, including the data on students' learning. The continuously evolving educational data are based on existing and new educational data and resources, the continuous data collection from students, and continuous inputs from educational experts and resources. Also worth mentioning is the fact that the students' data is not limited to the knowledge-level learning and assessment information in different subjects; it also includes students' core literacy related to these subjects, such as their math literacy and reading literacy. All the students' data can be utilized to better serve the students' learning and development. This provides the foundations of *big data* for the AI-aided SLP platform. The platform incorporates uses of *learning analytics*, *machine learning*, and *educational data mining*.

The human-machine interaction part in this SLP platform continuously interacts with the users. The platform provides various *assessment tools* which can be used by students and teachers. Based on the assessment results, data are analyzed and *learning reports* are provided to the users. The platform then sends *resource recommendations* to the users based on the users' learning reports. The platform uses AI technology to build visualizations of the students' learning progress diagram and the students' learning competencies level module, as well as students' strengths and weaknesses. Based on these data, the platform provides information and suggestions to students for *learning enhancement*.

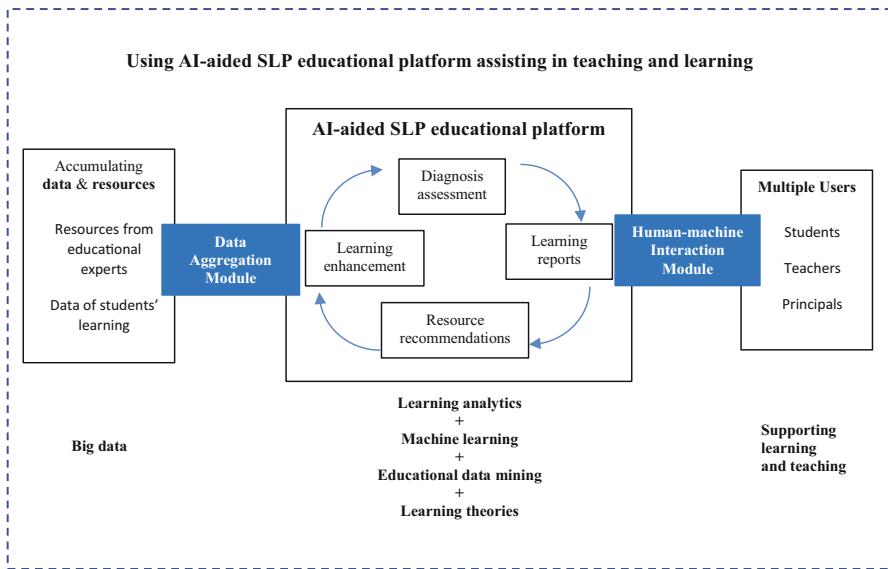


Fig. 1 Design structure of the AI-aided SLP platform

We identified four major functions of the SLP platform. The first function is to provide various assessment tools and tests for teaching and learning purposes. Students can access the tools in the platform to carry out self-diagnosis assessment whenever they want, while teachers can use the tools to do diagnosis assessment to assess the students' learning level or learning outcomes (Chen et al. 2018). Teachers gain a good overview of the learning situation of all students as well as individual students to provide appropriate teaching and individualized teaching for the students. The second function is to produce various learning analytical reports with instant feedback as well as learning progress over time. Students can get reports on their learning situation as well as their learning progress over a long period. The report can also show the students what they are good at and what they need to work on to improve themselves. Teachers can get reports on individual students' learning as well as the whole class's learning situations. In this way teachers can better plan their teaching and courses to suit their individual students' needs as well as the whole class situation. Principals have access to the overview report of the whole school teaching and learning situation, so that they can provide better support and resources for teachers and students. The third function of the platform is to provide recommendations and suggestions to the students and teachers. Students receive recommendations from the platform to improve their learning. Teachers also receive suggestions from the platform for their teaching to better support their students' needs. The fourth function of the platform is to provide a resource pool with various micro lectures. The teachers can use these micro lectures as part of their teaching. Students can watch these micro lectures according to their interests or based on the recommendations from the teachers or from the platform. All these four functions

from this AI-aided SLP platform provide many options and possibilities for teaching and learning at schools to better support the students' learning.

4 Methodology

The school in this case study is in an exurban area near Beijing, China. The school has been using the SLP platform since 2017. The main users were the principal as well as the students and teachers who were from grade 7 to grade 9 with students from 13 to 15 years. In-depth interviews were conducted with two students, two teachers, and one principal. A questionnaire with background information and six open-ended questions were collected on paper from seventh to ninth grade classes. The students could either voluntarily reply to the questionnaire or choose not to. Fifteen fully supplied responses were received. The participants in this study is shown in Table 1.

The background information included gender, grade or teaching position, as well as how many years they have used the platform. We made comprehensive interviews with students, teachers, and the principal at schools to investigate their experience of using the AI-aided SLP platform. We strove to understand and obtain their best experiences, challenges, and suggestions for this AI-aide SLP platform. We also sought to identify the major functions in the SLP used by students, teachers, and the principal at this school.

The questions in the questionnaire were almost the same as those used in the semi-constructed interviews. The main questions were:

1. Based on your experience of using the SLP platform, can you tell us your overall experience of using the SLP platform?
2. How did the platform support you in your study (or in your work)? Please give some concrete examples.

Table 1 Participants in this study

Participants	Gender	User information	How many years was the SLP platform used by the user? (Until 1st of June 2021)
Students (interview)	2 girls	7th–9th grade students	1–3 years
Students (Questionnaire in paper)	5 boys, 10 girls		
Teachers (interview)	2 female	Math teacher, physics teacher	Over 3 years
Principal (interview)	1 female	Principal	Over 3 years

3. Did the platform introduce any changes in your study (or in your work)? If the answer is yes, please share with us the kind of changes you had after you have used the platform.
4. What were the best experiences that you had when using the platform? Please share some concrete examples.
5. What challenges did you have when using the platform? Please share some concrete examples.
6. Would you like to share your wishes or suggestions for further improvement of the platform to better help you in your studies (or in your work)?
7. Based on our discussion, is there anything else you would like to share with us concerning your experience of using the platform? (This question was only used in the interview).

All participants in the interviews and questionnaire responses participated voluntarily. The participants were informed about their confidentiality and the possibility of withdrawing from the study at any time. All their personal information was removed, and it was not possible to identify the participants. All the interview data were voice recorded and transcribed.

The qualitative data analysis used content analysis to identify the key information. Two experienced researchers analyzed the qualitative data using content analysis. They also discussed the data analysis to achieve a synthesis in the data interpretation. The data analysis revealed the major ways in which the SLP platform assisted teaching and learning at the school, such as diagnosis assessments, student learning analytical reports, and accessing micro lecture resources and learning enhancement. We strove to identify how these aspects assisted in teaching and learning at the school by looking into the students', teachers', and the principal's self-reported experiences. Additionally, we aimed to identify the major challenges and further improvements in these kinds of learning platforms.

5 Findings

In this section, we present the main findings based on the multiple users' perspectives of students, teachers, and the principal and their experiences in using this platform.

5.1 Students' Self-Reported Experiences

The majority of student participants stated that the *main functions* they used in the SLP platform were self-assessment, checking the reports of their learning, and studying the micro lectures (online teaching videos). These functions helped them in the following ways: providing new ways and possibilities for learning and additional

learning resources; recognizing the weak parts and mistakes they made in their study and specific areas which needed to be improved; receiving recommendations and suggestions from the platform or from teachers; and consolidating student learning. Several students stated the following in their answers:

(It) provides more learning resources, such as the micro lectures (online videos) . . .
(Students 1, 2, 9, 10, 12, 16, 17)

(It) helps me to see which parts I am not good at, and provides suggestions for making improvements. (Students 2, 3, 4, 7, 9, 13, 15, 16, 17)

(It) helps me to reinforce and consolidate my learning . . . (Students 5, 6, 7, 8, 11, 12, 16, 17)

Later, we asked the students *what kinds of changes* they had experienced since using the SLP platform. Almost all the students stated that using the SLP platform changed their ways of learning in the following ways: broadening their thinking, building habits of self-assessment, becoming more active in learning, becoming more self-disciplined, making their own study plans and being able to follow the plans, finding their own ways of learning, improving their study, etc. Some students also stated that their learning motivation increased after they had used the platform to assist their learning:

(Using this platform) changed my way of learning . . . (Students 1, 2, 3, 4, 5, 7, 8, 9, 10, 11, 12, 13, 14, 15, 1, 17)

I became more active in my learning, my interests in studying increased, my thinking ability increased, I found my own ways of learning which are better than before . . . (Student 8)

My learning motivation has increased (Student 6)

When we asked the students what their *best experiences* were when using the SLP platform, the most mentioned was receiving feedback/suggestions/reports, especially instant feedback. Several students also mentioned that they liked the visualized diagram reports. The second most mentioned best experience was having the possibility of watching the online micro lectures and online videos at any time. They stated the following:

Receiving *instant* feedback/results (Students 4, 11, 8, 13, 16)

Receiving feedback/suggestions/reports (Students 1, 2, 4, 7, 8, 9, 10, 11, 12, 13, 16)

I can receive instant feedback after the exams, and I also get an analytical report of my learning, for example my weak parts, and I also received suggestions from the platform and teachers. (Student 11)

. . . watching the micro lecture (online videos) . . . (Students 3, 4, 5, 6, 13, 16)

When we asked the students what the *challenges* were when using the platform and invited their suggestions for further improvements, they stated that the challenges were a slow network or long response times from the platform or not being familiar with how to use some of the SLP functions. The students wished to have better connections and shorter response times from the platform, more user-friendly interface in the platform, and more resources in the platform, such as more

Table 2 Students' self-reported experiences of using the SLP platform

Key analysis categories	Responses from students
<i>Functions in SLP that helped the students' learning</i>	Self-diagnosis Feedbacks Suggestions and recommendations for further development Learning analysis reports Resources (e.g., micro lectures)
<i>Changes after using SLP</i>	New ways of learning More active Increased motivation
<i>Best experiences</i>	Instant feedback Learning analytical reports Micro lectures
<i>Challenges</i>	Slow internet connection Long connecting time from platform Some functions in SLP are not so clear
<i>Suggestions for further development</i>	Better connection Faster connecting time from platform More learning resources (e.g., micro lectures) User-friendly interface Online interaction and studying together with student peers

micro lectures and learning materials. Two students also suggested the possibility of having pairs or groups studying with peer students or chat functions with peer students.

The overall student self-reported experiences of using the SLP platform are summarized in Table 2. Based on the students' self-reported experiences of using the SLP platform, the overall experiences are very positive. Almost all students have stated that this platform assisted their learning and even changed their ways of learning. Students appreciated the feedback, reports, and suggestions and the available platform learning resources. Having said that, platform improvements were still needed, such as optimizing response time, user-friendly interface, and more learning resources. One interviewed student summarized her overall opinion about the platform:

This platform is like another teacher who can help me in my learning. (Student 16)

5.2 Teachers' and Principal's Self-Reported Experiences

We interviewed one math teacher, one physics teacher, and the principal from our case study school. In this section, we will discuss the self-reported experiences from teachers and the principal concerning *how this platform assisted in their work*.

Both subject teachers described this platform as a tool that assisted their teaching. The teachers used diagnosis assessment and generated reports to identify the students' weak points, key points, and individual needs in the students' learning. Based on the information and on analyzing reports, the teachers were able to provide individualized teaching to the students as well as adjust the teaching progress and pedagogical methods based on students' different needs. The teachers often used two major functions in the platform. One was the examination/test with auto-marking for diagnosis as well as for homework. The other was the platform-generated analysis reports of the individual students' learning as well as the overall situation of the whole class students' learning. As the math teacher stated in the interview:

(this platform) can provide accurate students' learning analytical reports as well as recommendations for students' improvement. This helps to provide individualized teaching based on students' needs . . . I like very much the 'instant' feedback/report from the system. I can see the students' learning reports right away after the exams . . .

When discussing *what changes* were experienced by the teachers after using the platform in their work, the teachers said that they could have a clearer picture of the students' learning needs and the overall learning situation in the class. One teacher also stated that their work became easier since they did not need to do marking when students took exams from the platform. Another teacher added that she felt the changes in her students, whether they were more advanced students or those who had learning difficulties, started when they did self-assessment and self-analysis of their own learning. Teachers also felt that their role had changed as more facilitators and students became more active learners when using the resources (micro lectures and tests) in the platform. The math teacher summarized that he was greatly impressed by:

. . . the platform's strong analysis capabilities which generated the students' learning reports. And this helps a great deal by providing individualized teaching and learning.

In the conversations that followed, we also discussed the challenges the teachers had encountered and their wishes for further platform improvements. In some comments from the teachers interviewed, one desired platform feature was that the teachers wished to add their teaching materials or teaching videos to the platform:

There are difficulties in adding the special math symbols to the platform.(Math teacher)

It is difficult to add some course information and contents in paper format to the platform.
(Physics teacher)

One teacher stated that the current functions were enough to assist her teaching. However, another teacher indicated that it would be good if some PowerPoint slides in the micro lectures could be downloaded so that she could modify and use them in her course.

The principal from this case study school participated in the project from the very beginning in order to implement the SLP platform at the school. She stated that the best function she used was the reports of the students' learning. From these reports she could see the overall picture of the teachers' teaching as well as the students'

learning. It became easier to identify which class was doing better, which teachers taught more effectively, and what were the improvement areas needed in teaching and learning. As the principal stated in her interview:

As a principal, I need to know the overall situation of teaching and learning at my school. From the reports generated from the SLP platform, I can have a good picture of how well the students have learned, which is also reflected in how well the teachers have taught in that class. And I can also see the changes over time. This helped me to pay attention to which parts needed to be improved . . .

The principal also felt that using the platform had brought some changes to the school:

The teachers' ICT competency has increased. There was more collaboration between ICT teachers and other teachers. Teachers were using the resources in the platform to improve their teaching . . . I also noticed that students became more active in their learning, and it changed their way of learning and thinking, as the students started to carry out more self-assessment.

Based on the teachers' self-reported experiences, the platform assisted their work in providing individualized teaching based on their students' needs. Overall, they were satisfied with the functions in the platform, although further development could be explored for specific subjects or needs. The principal also felt that the platform was useful in her work and would improve the teaching and learning in her school. The diagnosis assessment, platform generated reports, and the micro-lecture resources were beneficial, though the principal wished to have more varied tests and examinations and more micro-lecture resources in the platform. The overall student self-reported experiences of using the SLP platform is summarized in Table 3.

6 Discussion and Learning from This Case Study

In this section, we discuss the main findings and what we learned from this case study.

Based on the self-reported experience from students, teachers, and the principal, the findings demonstrate that this SLP platform can provide additional assistance for teaching and learning at schools (Lu et al. 2018). The following five major forms of learning were found.

6.1 Major Functions Favored by Students and Teachers

The following functions in the AI-aided SLP platform are important for teaching and learning: assessment tools, analytical reports, recommendations for further learning, and educational resources.

Table 3 Teachers' and the principal's self-reported experiences of using the SLP platform

Key analysis categories	Responses from teachers and the principal
<i>Functions in SLP that helped the teachers' and the principal's work</i>	Diagnosis assessment and generated students' learning analysis reports Suggestions and recommendations to improve students' learning Assessment tools for students' homework with platform auto-marking Using the micro lectures in teaching
<i>Changes after using SLP</i>	Teachers' work became easier, assisted by auto-marking, assessment tools, and micro lecture resources from the platform Teachers' role gradually became facilitators when students became more active in their learning using the resources from the platform Teachers were better able to provide individualized teaching based on the students' learning analysis reports from the platform Teachers' ICT competencies were increased
<i>Best experiences</i>	Instant feedback Analytical reports of students' learning Micro lectures
<i>Challenges</i>	Teachers had some difficulties in introducing materials or documents or teaching materials into the system
<i>Suggestions for further development</i>	Teachers felt that overall, they were satisfied with the current functions in the platform. It would be good if the teachers could download some teaching materials from the system so that they could make modifications in their teaching

From the teachers' point of view, teachers can better support the students' learning by obtaining more teaching resources and analytical reports of students' learning from big data, LA, and EDM in the platform. The resources from the platform for teachers include diagnostic assessment tools, homework assignments, micro lectures, etc. The teachers can see students' learning analytical reports with instant results and learning progress over a specific period. This function enables teachers to provide individualized teaching and learning for students. It also supports teachers in adjusting their teaching through making pedagogical decisions according to the students' needs.

From the students' point of view, students have more opportunities to be active in their learning. They can carry out self-diagnosis assessment of their own learning. Based on their learning analytical reports, they can gain recommendations for further development or actively seek resources for their learning.

6.2 New Ways and Possibilities in Learning and Teaching

AI-aided educational applications can provide new ways and possibilities in teaching and learning. In this case study, we can see that teachers can provide better personalized teaching based on the students' learning reports using diagnosis assessment. Additionally, teachers can use and select ready-made assessment items for students' homework or for both formative and summative assessments. Teachers can also use the micro lectures or other educational resources, such as teaching materials, in their courses.

This case study shows that students became more active in their learning. They can carry out self-assessment to become more aware of their strong and weak points in their learning. Moreover, they receive recommendations from the platform about how to enhance their learning and make improvements in their areas of difficulty. Students can also freely choose their interest areas to study from the large number of micro lectures available in the platform.

6.3 Positive Experiences and Changes

All students, teachers, and the principal in this study indicated that they have benefited from using the AI-aided SLP educational platform. The platform also introduced changes in teaching and learning at schools. Teachers said that their work became easier. The teachers' role gradually changed to become facilitators, and students became more active in their learning. Students' learning motivation also increased, and they stated that they found new ways and methods for their learning. Finally, the principal found that she could optimize and increase the level of proficiency in school planning and resource allocation.

6.4 The Importance of Learning Theories Applied in AIED Applications

Many current learning theories could be implemented in the design of the AI-aided educational platform. All the assessment tools were based on Bloom's learning pyramid (Bloom 1956) at various levels. The learning reports showed the students' learning capabilities in remembering, understanding, applying information, analyzing, evaluating, and creating. This matches our current expected learning outcomes from students. An adaptive learning cognitive map model was used in the SLP (Wan and Yu 2020).

6.5 Continuous Improvements and the Social Nature of Learning

This AI-aided SLP platform is a dynamic progressive system design. It continuously collects data from students as well as from teachers, which contribute to the big data, ML, and EDM. The platform becomes more intelligent with the continuous input data, and it also becomes more adaptive to the users' needs. It represents a continuously progressive improvement in the interactions between the SLP's AI and its human users. However, another important concern is the social aspects, which refer to the interactions among peer students through the platform. It is critical to know how to build a social supportive and collaborative learning community as well as to use the AI-aided platform to support students' learning. Students in this study wished to study together with other peer students and have more social interaction when using the SLP platform. Designers and developers for the educational platform need to think and rethink how to satisfy the students' social needs in the platform.

7 Conclusion and Recommendation

To conclude, the AI-aided SLP educational platform can be used as another tool to support better learning and teaching at school. Students can have more resources and options in their learning and become more active learners when they have various choices and receive instant feedback. Teachers have additional ready-made expert contents and assessment tools for their teaching. Teachers and principals can receive an instant view of all the students as well as individual students' learning situation and learning progress. This enables them to provide better teaching and individualized support for students. The teachers' role is also gradually shifting more to facilitators. This case study demonstrates that the AI-aided SLP educational platform did lead to positive effects and changes and assisted in teaching and learning at school.

Based on this case study, practical implications and recommendations were drawn concerning further development of this kind of AI-aided educational platform. First, to support better learning and teaching, the most favored functions were identified as the assessment tools, learning reports, online resources, and recommendations. Second, it was found that learning theories should be combined with AI technology. This enables positive experiences in teaching and learning. The students became more motivated and active in their learning. Teachers had more time and information to provide individualized learning, and the teachers' role gradually shifted to a more facilitative role. However, there were calls to expand the system to incorporate the social aspects of learning and make continuous improvement. Students wished to have social interactions with their student peers. Providing group learning and/or peer support and a learning community could become extremely valuable. The future new design features should respond to the

“social nature” of learning and consider how to synthesize the AI technology with social human learning needs to enhance its usefulness. Both students and teachers expressed the need for an easier-to-use and faster user interface. These factors can be taken into consideration in improved designs for AI-aided educational platforms. Additionally, rethinking is needed concerning ways in which the platform can help teachers save time and make their work easier. The platform should focus on refining the individualized services for each student, for example, more and better choices concerning micro lectures and online off-school support for their homework or when students face difficulties.

This case study was based on multiple users’ self-reported experiences in one school. Conducting further studies in a wider school population is suggested for future research. It is also worthwhile comparing the study results of this AI-aided SLP platform with other similar kinds of AI-aided educational platforms. More studies are needed concerning specific new design features to meet the needs from users and to enhance the usefulness of AI-aided educational platforms.

References

- Alexander, B., Ashford-Rowe, K., Barajas-Murphy, N., Dobbin, G., Knott, J., McCormack, M., Pomerantz, J., Seilhamer, R., & Weber, N. (2019). *EDUCAUSE horizon report: 2019 higher education edition*. <https://library.educause.edu/resources/2019/4/2019-horizon-report>. Accessed 11 Oct 2021.
- Baker, R.S., & Inventado, P.S. (2014). Educational Data Mining and Learning Analytics. In: Larusson J., White B. (eds) *Learning Analytics*. New York: Springer.
- Bloom, B.S. (ed.). (1956). *Taxonomy of Educational Objectives: Handbook I, Cognitive Domain*. New York: Longman.
- Chen, P., Lu, Y., Zheng, V., & Bian, Y. (2018). Prerequisite-driven deep knowledge tracing. In *IEEE Conference on Data Mining (ICDM 2018)* (pp. 39–48). IEEE.
- Fischer, C., Pardos, Z. A., Baker, R. S., Williams, J. J., Smyth, P., Yu, R., Slater, S., Baker, R., & Warschauer, M. (2020). Mining big data in education: Affordances and challenges. *Review of Research in Education*, 44(1), 130–160.
- Gartner. (2012). What is big data? Gartner IT glossary—big data. <https://www.gartner.com/it-glossary/big-data/>. Accessed 11 Oct 2021.
- Keleş, A., Ocak, R., Keleş, Ali, & Gülcü, A. (2009). ZOSMAT: Web-based intelligent tutoring system for teaching–learning process. *Expert Systems with Applications*, 36(2), 1229–1239.
- Labarthe, H., Luengo, V., & Bouchet, F. (2018). Analyzing the relationships between learning analytics, educational data mining and AI for education. *14th International Conference on Intelligent Tutoring Systems (ITS): Workshop Learning Analytics* (pp. 10–19), June 2018, Montreal, Canada.
- Lorenz, P., & Saslow, K. (2019). Demystifying AI & AI companies—What foreign policy makers need to know about the global AI industry. Stiftung Neue Verantwortung. <https://www.stiftung-nv.de/de/publikation/demystifying-ai-ai-companies-whatforeign-policy-makers-need-know-about-global-ai>. Accessed 31 July 2019.
- Lu Y., Chen, C., Chen P., Chen X., & Zhuang Z. (2018). Smart learning partner: An interactive robot for education. In: Penstein Rosé C. et al. (Eds.). *Artificial Intelligence in Education* (pp. 447–451). Cham: Springer.

- Lu, Y., Wang, D., Meng, Q., & Chen, P. (2020). Towards interpretable deep learning models for knowledge tracing. *International Conference on Artificial Intelligence in education* (pp. 185–190). Cham: Springer.
- Minsky, M. L., & Papert, S. A. (1968). *Perceptrons: An Introduction to Computational Geometry*. Cambridge, MA: MIT Press.
- Mousavinasab, E., Zarifsanaiey, N., S. R. Niakan Kalhor, Rakhshan, M., Keikha, L., & Ghazi Saeedi, M. (2021). Intelligent tutoring systems: A systematic review of characteristics, applications, and evaluation methods. *Interactive Learning Environments*, 29(1), 142–163.
- Pian, Y., Lu, Y., Chen, P., & Duan, Q. (2019). Coglearn: A cognitive graph-oriented online learning system. *2019 IEEE 35th International Conference on Data Engineering (ICDE)* (pp. 2020–2023). IEEE.
- Renz, A., & Hilbig, R. (2020). Prerequisites for artificial intelligence in further education: Identification of drivers, barriers, and business models of educational technology companies. *International Journal of Educational Technology in Higher Education*, 17(1), 1–21. <https://doi.org/10.1186/s41239-020-00193-3>.
- Renz, A., Krishnaraja, S., & Gronau, E. (2020). Demystification of artificial intelligence in education—how much AI is really in the educational technology? *International Journal of Learning Analytics and Artificial Intelligence for Education*, 2(1), 4–30.
- Stone, P., Brooks, R., Brynjolfsson, E., Calo, R., Etzioni, O., Hager, G., Hirschberg, J., Kalyanakrishnan, S., Kamar, E., Kraus, S., Leyton-Brown, K., Parkes, D., Press, W., Saxenian, A., Shah, J., Tambe, M., & Teller, A. (2016). Artificial Intelligence and Life in 2030. *One Hundred Year Study on Artificial Intelligence: Report of the 2015-2016 Study Panel*, Stanford University, Stanford, CA, September 2016. Doc: <http://ai100.stanford.edu/2016-report>
- Tan, H., Wang, C., Duan, Q.L., Lu, Y., Zhang, H., Li, R. (2020). Automatic short answer grading by encoding student responses via a graph convolutional network. *Interactive Learning Environments*, 29, 1–15.
- VanLehn, K. (2006). The behavior of tutoring systems. *International Journal of Artificial Intelligence in Education*, 16(3), 227–265.
- Wan, H. & Yu, S. (2020). A recommendation system based on an adaptive learning cognitive map model and its effects. *Interactive Learning Environments*, <https://doi.org/10.1080/10494820.2020.1858115>

Open Access This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.



Deep Learning in Automatic Math Word Problem Solvers



Dongxiang Zhang

Contents

1	Introduction	233
2	Methodology and Analysis	236
2.1	Rule-Based Methods	236
2.2	Statistic-Based Methods	236
2.3	Tree-Based Methods	238
2.4	Deep Learning Models	239
2.5	Geometry Problem Solving	242
3	Conclusions	243
	References	244

1 Introduction

Designing an automatic solver for mathematical word problems (MWPs) has a long history dating back to the 1960s and continues to attract intensive attention as a frontier research topic. The problem is challenging because there remains a wide semantic gap to parse the human-readable words into machine-understandable logics to conduct quantitative reasoning. Various attempts have been made to bridge the gap, from rule-based pattern matching to semantic parsing with statistical machine learning, and to the recent end-to-end deep learning models that are considered as the state-of-the-art performers. To a certain extent, the problem has been recognized as a good test bed to evaluate the intelligence level of agents, as it requires semantic understanding of natural languages and capabilities of automatic reasoning. Hence, the successful solving of MWPs would constitute a milestone toward general AI.

A large body of research works start from solving arithmetic word problems for elementary school students. Its input is the text description for the math problem, represented in the form of a sequence of tokens. There are multiple quantities

D. Zhang (✉)
Zhejiang University, Hangzhou, China
e-mail: zhangdongxiang@zju.edu.cn

Fig. 1 An example of an arithmetic word problem

Word Problem
Oceanside Bike Rental Shop charges 17 dollars plus 7 dollars an hour for renting a bike. Tom paid 80 dollars to rent a bike. How many hours did he pay to have the bike checked out?
Equation
$17 + (7 * x) = 80$
Solution
$x = 9$

mentioned in the text and an unknown variable in the question whose value is to be resolved. The problem solver's objective is to extract the relevant quantities and map this problem into an arithmetic expression whose evaluation value provides the solution to the problem. For simplicity, there are only four types of fundamental operators $O = \{+, -, \times, \div\}$ involved in the math expression.

An example of an arithmetic word problem is illustrated in Fig. 1. The relevant quantities to be extracted from the text include 17, 7, and 80. The number of hours spent on the bike is the unknown variable x . To solve the problem, we need to identify the correct operators between the quantities and their operation order such that we can obtain the final equation $17 + 7x = 80$ or expression $x = (80 - 17) \div 7$ and return 9 as the solution to this problem.

The early approaches mainly relied on rule-based reasoning. They heavily count upon human interventions to manually craft rules and schemas for pattern matching. Each rule consists of a set of conditions that must be satisfied and the actions to be carried out. For example, as a system published in 1985, WORDPRO, predefines a collection of rules to handle simple math problems. If the problem text matches the “HAVE-MORE-THAN” proposition, the agent will identify the two operands and use the “ $-$ ” operator to derive the answer. It is evident that the usefulness of these rule-based solvers is doubtful because they can only resolve a limited number of scenarios that are defined in advance.

To improve the generality, subsequent efforts have been devoted to making use of semantic parsing to map the sentences from problem statements into structured logic representations so as to facilitate quantitative reasoning. It has regained considerable interests from the academic community, and a booming number of methods have been proposed in the past years. These methods leverage various strategies of feature engineering and statistical learning for performance boosting. For instance, if two quantities have the same dependent verbs, as in a problem like “in the first round she scored 40 points and in the second round she scored 50 points,” it is likely that “ $+$ ” would be the operator for these two numbers. Despite the promising results

claimed in some small datasets, these approaches are not completely automatic and still require human knowledge to help extract semantic features.

To further reduce human intervention and enable the automatic extraction of discriminative features, applying deep learning (DL) models in MWPs has become a promising research direction. In 2017, Wang et al. proposed DNS as the first end-to-end DL-based framework that directly converts the input of question text into the output of math expression. It is then a natural idea to apply an existing sequence-to-sequence (seq2seq) learning model to encode the text input and decode the hidden features into a math expression. The drawback is that the seq2seq model is a black box that lacks interpretability, and it cannot guarantee the output is in valid math format and normally requires a post-processing step. Nonetheless, this work still occupies an important position in the literature of MWP solving because it opened up a new research direction to apply end-to-end DL models to solve MWPs and attracted a good number of followers to contribute to this research area.

Following the research line of seq2seq models, various optimization techniques have been proposed to further improve accuracy. A recent breakthrough is that since the resulting math expression can be naturally represented as a tree structure, this finding allows us to leverage more informative context for decoding. For example, a math expression $2 + 3$ can be converted to a tree structure in which the root is operator $+$, and there are two child nodes with operands 2 and 3. The decoder can recursively generate an expression tree in a top-down manner and take into account the encodings of parent node and sibling nodes as the more informative context. Following the idea, we have witnessed the success of seq2tree models which have exhibited clear superiority over seq2seq models. There have also emerged several incremental works on top of seq2tree models. The general idea is to replace the encoder or decoder with more effective graph-based embedding since sequences and trees can be viewed as two special cases of graphs.

At the end of the chapter, we will cover geometry problem solvers that require both textual and visual understanding. The problem is even more challenging because the input needs to be mapped into a logical representation that is compatible with both the problem text and the accompanying diagram. Common strategies to solve geometry word problems constitute three key components, including diagram understanding to capture visual clues, text parsing to capture semantic information, and deductive reasoning via a knowledge base with geometry axioms and theorems. We will introduce representative systems such as GEOS and Inter-GPS. They parse the problem text and geometry diagram into formal language and then perform symbolic reasoning step by step to derive the solution. The readers can try the demos of GEOS published by the University of Washington.¹

¹ <https://geometry.allenai.org/>.

2 Methodology and Analysis

In the following, we present the general design principles of rule-based methods, statistic-based methods, tree-based methods, as well as recent advances with deep learning models.

2.1 Rule-Based Methods

The early approaches to math word problems are rule-based systems based on hand engineering. Published in 1985, WORDPRO (Fletcher 1985) can solve three types of simple one-step arithmetic problems, including value *change*, *combine*, and *compare*. A collection of rules is predefined for pattern matching. For example, given a problem text “*Dan has six books. Jill has two books. How many books does Dan have more than Jill?*”, it matches the predefined “HAVE-MORE-THAN” proposition. The agent will identify the two operands and use the “−” operator to derive the answer. Another system ROBUST, developed by (Bakman 2007), expanded the rule base and could better understand free-format multistep arithmetic word problems. It further extends the *change* schema of WORDPRO into six distinct categories. The multistep problem is solved by splitting the problem text into sentences and each sentence is mapped to a proposition. Yun et al. also proposed to use schema for multistep math problem solving (Yun et al. 2010). However, their implementation details were not explicitly revealed. Since these systems are out of date, we only provide such a brief overview for representativeness. The readers can refer to Mukherjee and Garain (2008) for a comprehensive survey of early rule-driven systems for automatic understanding of natural language math problems. Since these systems heavily rely upon human interventions to manually craft rules and schemas for pattern matching, it is evidently that the usefulness of these rule-based solvers is doubtful because they can only resolve a limited number of scenarios defined in advance.

2.2 Statistic-Based Methods

The statistic-based methods leverage traditional machine learning models to identify the entities, quantities, and operators from the problem text and yield the numeric answer with simple logic inference procedure. The scheme of quantity entailment proposed in (Roy et al. 2015) can be used to solve one-step arithmetic problems. It involves three types of classifiers to detect different properties of the word problem. The *quantity pair classifier* is trained to determine which pair of quantities would be used to derive the answer. The *operator classifier* picks the operator $op \in \{+, -, \times, \div\}$ with the highest probability. The *order classifier* is relevant only

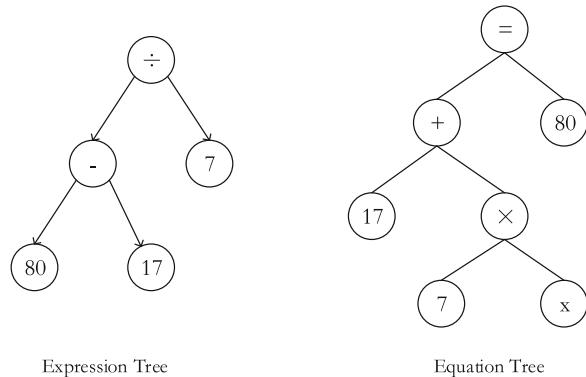
for problems involving subtraction or division because the order of operands matters for these two types of operators. With the inferred expression, it is straightforward to calculate the numeric answer for the simple math problem.

To solve math problems with multistep arithmetic expression, the statistic-based methods require more advanced logic templates. This usually incurs additional preparatory overhead to annotate the text problems and associate them with the introduced template. As an early attempt, ARIS (Hosseini et al. 2014) defined a logic template named *state* that consists of a set of entities, their containers, attributes, quantities, and relations. For example, “*Liz has 9 black kittens*” initializes the number of *kitten* (referring to an entity) with *black* color (referring to an attribute) and belonging to *Liz* (referring to a container). The solution splits the problem text into fragments and tracks the update of the states by verb categorization. More specifically, the verbs are classified into seven categories: *observation*, *positive*, *negative*, *positive transfer*, *negative transfer*, *construct*, and *destroy*. To train such a classifier, we need to annotate each split fragment in the training dataset with the associated verb category. Another drawback of ARIS is that it only supports addition and subtraction. Sundaram and Khemani (2015) followed a similar processing logic to ARIS. They predefined a corpus of logic representation named *schema*, inspired by Bakman (2007). The sentences in the text problem are examined sequentially until the sentence matches a schema, triggering an update operation to modify the number associated with the entities.

Mitra and Baral (2016) proposed a new logic template named *formula*. Three types of formulas are defined, including *part whole*, *change*, and *comparison*, to solve problems with addition and subtraction operators. For example, the text problem “*Dan grew 42 turnips and 38 cantelopes. Jessica grew 47 turnips. How many turnips did they grow in total?*” is annotated with the part-whole template: $\langle \text{whole} : x, \text{parts} : \{42, 47\} \rangle$. To solve a math problem, the first step connects the assertions to the formulas. In the second step, the most probable formula is identified using the log-linear model with learned parameters and converted into an algebraic equation. Another type of annotation is introduced by Liang and colleagues (Liang et al. 2016a,b) to facilitate solving a math word problem. A group of *logic forms* is predefined and the problem text is converted into the logic form representation by certain mapping rules. For instance, the sentence “*Fred picks 36 limes*” will be transformed into $\text{verb}(v_1, \text{pick}) \& \text{nsubj}(v_1, \text{Fred}) \& \text{dobj}(v_1, n_1) \& \text{head}(n_1, \text{lime}) \& \text{nummod}(n_1, 36)$. Finally, logic inference is performed on the derived logic statements to obtain the answer.

To sum up, these statistical-based methods have two drawbacks that limit their usability. First, it requires additional annotation overhead that prevents them from handling large-scale datasets. Second, these methods are essentially based on a set of predefined templates, which are brittle and rigid. It will take great efforts to extend the templates to support other operators like multiplication and division.

Fig. 2 Examples of expression tree and equation tree for Fig. 1



2.3 Tree-Based Methods

The arithmetic expression can be naturally represented as a binary tree structure such that the operators with higher priority are placed in the lower level and the root of the tree contains the operator with the lowest priority. The idea of tree-based approaches is to transform the derivation of the arithmetic expression to constructing an equivalent tree structure step by step in a bottom-up manner. One of the advantages is that there is no need for additional annotations such as equation template, tags, or logic forms. Figure 2 shows two tree examples derived from the math word problem in Fig. 1. One is called an *expression tree* that is used in (Roy and Roth 2015, 2017; Wang et al. 2018b), and the other is called an *equation tree* (Koncel-Kedziorski et al. 2015). These two types of trees are essentially equivalent and result in the same solution, except that equation tree contains a node for the unknown variable x .

The overall algorithmic framework common to the tree-based approaches consists of two processing stages. In the first stage, the quantities are extracted from the text and form the bottom level of the tree. The candidate trees that are syntactically valid, but with different structures and internal nodes, are enumerated. In the second stage, a scoring function is defined to pick the best matching candidate tree, which will be used to derive the final solution. A common strategy among these algorithms is to build a local classifier to determine the likelihood of an operator being selected as the internal node. The input of the classifier consists of the contextual embeddings for its two child nodes and the output is a label in the operator set $\{+, -, *, \div\}$. Such local likelihood is taken into account in the global scoring function to determine the likelihood of the entire tree.

Roy and Roth (2015) proposed the first algorithmic approach that leverages the concept of an expression tree to solve arithmetic word problems. Its first strategy to reduce the search space is training a binary classifier to determine whether an extracted quantity is relevant or not. Only the relevant ones are used for tree construction and placed in the bottom level. The irrelevant quantities are discarded. The tree construction procedure is mapped to a collection of simple prediction

problems, each determining the lowest common ancestor operation between a pair of quantities mentioned in the problem. The global scoring function for an enumerated tree takes into account two terms. The first one is the likelihood of quantity being irrelevant, i.e., the quantity is not used in creating the expression tree. The other term is the likelihood of selecting an operator in one of the internal tree nodes. The service is also published as a web tool (Roy and Roth 2016), and it can respond promptly to a math word problem.

ALGES (Koncel-Kedziorski et al. 2015) differs from (Roy and Roth 2015) in two major ways. First, it adopts a more brute-force manner to exploit all the possible equation trees. More specifically, ALGES does not discard irrelevant quantities but enumerates all the syntactically valid trees. Second, its scoring function is different. There is no need to measure quantity relevance because ALGES does not build such a quantity classifier. The goal of (Roy et al. 2016) is also to build an equation tree by parsing the problem text. It makes two assumptions that can simplify the tree construction, but sacrifice its applicability. First, the final output equation form is restricted to have at most two variables. Second, each quantity mentioned in the sentence can be used at most once in the final equation. The tree construction procedure consists of a pipeline of predictors that identify irrelevant quantities, recognize grounded variables, and generate the final equation tree. With customized feature selection and SVM (support vector machine)-based classifier, the relevant quantities and variables are extracted and used as the leaf nodes of the equation tree. Finally, the tree is built in a bottom-up manner.

UnitDep (Roy and Roth 2017) can be viewed as an extension of work by the same authors (Roy and Roth 2015). An important concept, named Unit Dependency Graph (UDG), is proposed to enhance the scoring function. The vertices in UDG consist of the extracted quantities. If the quantity corresponds to a rate (e.g., 8 dollars per hour), the vertex is marked as RATE. There are six types of edge relations to be considered, such as whether two quantities are associated with the same unit. Building the UDG requires additional annotation overhead as we need to train two classifiers for the nodes and edges. The node classifier determines whether a node is associated with a rate. The edge classifier predicts the type of relationship between any pair of quantity nodes. This facilitates the processing of operators “*” and “/.”

2.4 Deep Learning Models

In recent years, deep learning (DL) has witnessed great success in a wide spectrum of “smart” applications. The main advantage is that with enough training data, DL is able to learn an effective feature representation in a data-driven manner without human intervention. It is not surprising that several efforts have sought to apply DL for math word problem solving. Deep Neural Solver (DNS) (Wang et al. 2017) is the first deep learning-based algorithm that does not rely on hand-crafted features. This is a milestone contribution because all the previous methods required human intelligence to help extract features that are effective. The deep model used in DNS

is a typical sequence to sequence (seq2seq) model (Sutskever et al. 2014). The readers without deep learning background can view it as a black box to magically encode the input sequence, which refers to the problem text, and generate a math expression as the output. To ensure that the output equations by the model are syntactically correct, five rules are predefined as validity constraints. For example, if the i th character in the output sequence is an operator in $\{+, -, \times, \div\}$, then the model cannot result in $c \in \{+, -, \times, \div, , =\}$ for the $(i + 1)$ th character.

Following DNS, there have emerged multiple DL-based solvers for arithmetic word problems. Seq2SeqET (Wang et al. 2018a) extended the idea of DNS by using expression tree as the output sequence. In other words, it applied seq2seq model to convert the problem text into an expression tree (as depicted in Fig. 2). Given the output of an expression tree, we can easily infer the numeric answer. T-RNN (Wang et al. 2019) can be viewed as an improvement of Seq2SeqET, in terms of quantity encoding, template representation, and tree construction. First, an effective embedding network (with Bi-LSTM and self-attention) is used to vectorize the quantities. Second, the detailed operators in the templates are encapsulated to further reduce the number of template space. For example, $n_1 + n_2$, $n_1 - n_2$, $n_1 \times n_2$, and $n_1 \div n_2$ are mapped to the same template $n_1 \langle op \rangle n_2$. Third, they are the first to adopt recursive neural network (Goller and Kuchler 1996) to infer the unknown variables in the expression tree in a recursive manner.

Wang et al. made the first attempt of applying deep reinforcement learning to solve arithmetic word problems (Wang et al. 2018b). The motivation is that deep Q-network has witnessed success in solving various problems with large search space. To fit the math problem scenario, they formulate the expression tree construction as a Markov Decision Process and propose the MathDQN that is customized from the general deep reinforcement learning framework. Technically, they tailor the definitions of states, actions, and reward functions which are key components in the reinforcement learning framework. The framework learns model parameters from the reward feedback of the environment and iteratively picks the best operator for two selected quantities.

A recent breakthrough comes from the observation that tree structures (e.g., the expression trees in Fig. 2) provide a more informative data structure than sequential expression (e.g., $17 + (7 * x) = 80$) to leverage. Following the idea, the sequence-to-sequence generation model can be replaced by sequence-to-tree model to improve performance. GTS (Xie and Sun 2019) is a representative sequence-to-tree model and is still considered as a competitive method in solving MWPs. Its decoder recursively generates an expression tree in a top-down manner. During the decoding process, it takes into account the encodings of parent node and sibling nodes as more informative context. There have also emerged several incremental works on top of seq2seq or seq2tree models, either by replacing the encoder with graph-based embedding or using a graph as a more general structure than trees to represent math expressions. For example, Graph2Tree (Zhang et al. 2020) replaces the sequential model with graph-based embedding to better capture the relationships and order information among the quantities. Seq2DAG (Cao et al. 2021) works by extracting the equation as a Direct Acyclic Graph (DAG) structure upon problem description.

Table 1 Performance of deep learning models on benchmark datasets

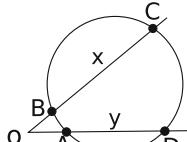
	Math23K	Math23K*	MAWPS
DNS (Wang et al. 2017)	–	58.1	59.5
Seq2SeqET (Wang et al. 2018a)	66.7	–	69.2
T-RNN (Wang et al. 2019)	66.9	–	66.8
MathDQN (Wang et al. 2018b)			60.25
GTS (Xie and Sun 2019)	75.6	74.3	82.6
Graph2Tree (Zhang et al. 2020)	77.4	75.5	83.7
Seq2DAG (Cao et al. 2021)	–	77.1	–
MWP-BERT (Liang et al. 2021)	84.4	82.3	–
TM-generation (Lee et al. 2021)	85.3	–	85.2

In Table 1, we summarize the performance of these models in benchmark datasets. There are three datasets commonly used, including Math23K, Math23K*, and MAWPS.

1. **Math23K** (Wang et al. 2017). The dataset contains Chinese math word problems for elementary school students and is created by web crawling from multiple online education websites. Initially, 60,000 problems with only one unknown variable are collected. The equation templates are extracted in a rule-based manner. To ensure high precision, a large number of problems that do not fit the rules are discarded. Finally, 23,162 math problems remained. Since the test set in Math23K is predefined, some researchers use its modified version called **Math23K***. This dataset applies fivefold cross-validation and is considered to be more challenging than Math23K.
2. **MAWPS** (Koncel-Kedziorski et al. 2016) is another testbed in English language for arithmetic word problems with one unknown variable in the question. Its objective is to compile a dataset of varying complexity from different websites. Operationally, it combines the published word problem datasets used in the previous literature. There are 2373 questions in the harvested dataset.

From the results, we can see that accuracy continues to improve as a more complex encoder or decoder is applied. Seq2DAG achieves state-of-the-art performance in Math23K*. It is worth noting that there is a recent trend to leverage the power of pretrained language models, such as BERT (Devlin et al. 2019) or its variants (Clark et al. 2020; Lewis et al. 2020), to further boost the accuracy. For instance, MWP-BERT (Liang et al. 2021) incorporates BERT and TM-generation model (Lee et al. 2021) adopts ELECTRA (Clark et al. 2020) as the pretraining model. These models are pretrained using a very large number of documents with billions of words in total. The training of BERT and ELECTRA consumes enormous hardware resources and computation time and the trained model contains hundreds of millions of parameters (110M for BERT-Base and 340M for BERT-Large). When they are applied to solve MWPs, we can observe significant performance improvement.

Fig. 3 An example of geometric problem

Geometric Problem	
In the figure below triangle OAB has an area of 72 and triangle ODC has an area of 288. Find x and y.	
	
Equation	area of OAB = $72 = (1/2) \sin(\angle AOB) * OA * OB$ solve the above for $\sin(\angle AOB)$ to find $\sin(\angle AOB) = 1/2$ area of ODC = $288 = (1/2) \sin(\angle DOC) * OD * OD$ Note that $\sin(\angle DOC) = \sin(\angle AOB) = 1/2$, $OD = 18 + y$ and $OC = 16 + x$ and substitute in the above to obtain the first equation in x and y $1152 = (18 + y)(16 + x)$ We now use the theorem of the intersecting lines outside a circle to write a second equation in x and y $16 * (16 + x) = 14 * (14 + y)$
Solution	$x=20, y=14$

2.5 Geometry Problem Solving

Geometry problem solving is more challenging because they require considering visual diagram and textual expressions simultaneously. As illustrated in Fig. 3, a typical geometry word problem contains text descriptions or attribute values of geometric objects. The visual diagram may contain essential information that is absent from the text. For instance, points O , B , and C are located on the same line segment, and there is a circle passing points A , B , C , and D . To well solve geometry word problems, three main challenges need to be tackled: (1) diagram parsing requires the detection of visual mentions, geometric characteristics, the spatial information, and the co-reference with text, (2) deriving visual semantics that refer to the textual information related to the visual analogue involves assigning semantic and syntactic interpretation to the text, and (3) the inherent ambiguities lie in the task of mapping visual mentions in the diagram to the concepts in real world.

G-ALINGER (Seo et al. 2014) is an algorithmic work that addresses the geometry diagram understanding and text understanding simultaneously. To detect primitives from a geometric diagram, the Hough transform (Shapiro and Stockman 2001) is first applied to initialize lines and circles segments. An objective function, which incorporates pixel coverage, visual coherence, and textual–visual alignment,

is applied. The function is sub-modular, and a greedy algorithm is designed to pick the primitive with the maximum gain in each iteration. The algorithm stops when no positive gain can be obtained according to the objective function. GEOS (Seo et al. 2015) can be considered as the first work to tackle a complete geometric word problem as shown in Fig. 3. Its method consists of two main steps: (1) parsing text and diagram, respectively, by generating a piece of logical expression to represent the key information of the text and diagram as well as the confidence scores, and (2) addressing the optimization problem by aligning the satisfiability of the derived logical expression in a numerical method that requires manually defining the indicator function for each predicate. It is noticeable that G-ALINGER is applied in GEOS (Seo et al. 2014) for primitive detection. Despite the superiority of automated solving process, the performance of the system would be undermined if the answer choices are unavailable in a geometry problem and the deductive reasoning based on geometric axioms is not used in this method. Inter-GPS (Lu et al. 2021) adopts a similar strategy to parse the problem text and diagram into formal language automatically via rule-based text parsing and neural object detecting, respectively. It incorporates theorem knowledge as conditional rules and performs symbolic reasoning in a stepwise manner. A subsequent improver of GEOS is presented in Sachan et al. (2017). It harvests axiomatic knowledge from 20 publicly available math textbooks and builds a more powerful reasoning engine that leverages the structured axiomatic knowledge for logical inference.

GeoShader (Alvin et al. 2017) is the first tool to automatically handle geometry problems with shaded area, presenting an interesting reasoning technique based on an analysis hypergraph. The nodes in the graph represent intermediate facts extracted from the diagram and the directed edges indicate the relationship of deductibility between two facts. The calculation of the shaded area is represented as the target node in the graph and the problem is formulated as finding a path in the hypergraph that can reach the target node.

3 Conclusions

In summary, despite the great success achieved by applying DL models to solve MWPs, the current status in this research domain still has room for improvement. We now consider a number of possible future directions that may be of interest to the AI in education community.

First, aligning visual understanding with text mention is an emerging direction that is particularly important for solving geometry word problems. However, this challenging problem has only been evaluated in self-collected and small-scale datasets, similar to those early efforts on evaluating the accuracy of solving algebra word problem. There is a chance that these proposed aligning methods fail to work well in a large and diversified dataset. Hence, it calls for a new round of evaluation for generality and robustness with a better benchmark dataset yet to be developed for geometry problems.

Second, interpretability plays a key role in measuring the usability of MWP solvers in the application of online tutoring but may pose new challenges for the deep learning-based solvers. For instance, AlphaGo (Silver et al. 2016) and AlphaZero (Silver et al. 2017) have achieved astonishing superiority over human players, but their near-optimal actions could be difficult for human to interpret. Similarly, for MWP solvers, domain knowledge and reasoning capability are useful and they are easy to interpret and understandable for human beings. It may be interesting to combine the merits of DL models, domain knowledge, and reasoning capability to develop more powerful MWP solvers.

Last but not the least, solving math word problems in English plays a dominating role in the literature. We only observed a very rare number of math solvers proposed to cope with other languages. This research topic may grow into a direction with significant impact. To our knowledge, many companies in China have harvested an enormous number of word problems in K12 education. As reported in 2015,² Zuoyebang, a spin off from Baidu, has collected 950 million questions and solutions in its database. When coupled with deep learning models, this is an area ripe for investigatory imagination and exciting achievements can be expected.

References

- Alvin, C., Gulwani, S., Majumdar, R., & Mukhopadhyay, S., (2017). Synthesis of problems for shaded area geometry reasoning. In *Aied* (pp. 455–458).
- Bakman, Y., (2007). Robust Understanding of Word Problems with Extraneous Information. *ArXiv Mathematics e-prints*.
- Cao, Y., Hong, F., Li, H., & Luo, P., (2021). A bottom-up DAG structure extraction model for math word problems. In *Aaaai* (pp. 39–46). AAAI Press.
- Clark, K., Luong, M., Le, Q. V., & Manning, C. D., (2020). ELECTRA: pre-training text encoders as discriminators rather than generators. In *Iclr*. OpenReview. net.
- Devlin, J., Chang, M., Lee, K., & Toutanova, K., (2019). BERT: pre-training of deep bidirectional transformers for language understanding. In J. Burstein,C. Doran, & T. Solorio (Eds.), *Naacl-hlt* (pp. 4171–4186). Association for Computational Linguistics.
- Fletcher, C. R. (1985, Sep 01). Understanding and solving arithmetic word problems: A computer simulation. *Behavior Research Methods, Instruments, & Computers*, 17(5), 565–571.
- Goller, C., & Kuchler, A., (1996). Learning task-dependent distributed representations by back-propagation through structure. *Neural Networks*, 1, 347–352.
- Hosseini, M. J., Hajishirzi, H., Etzioni, O., & Kushman, N., (2014). Learning to solve arithmetic word problems with verb categorization. In *Emnlp* (pp. 523–533).
- Koncel-Kedziorski, R., Hajishirzi, H., Sabharwal, A., Etzioni, O., & Ang, S. D., (2015). Parsing algebraic word problems into equations. *TACL*, 3, 585–597.
- Koncel-Kedziorski, R., Roy, S., Amini, A., Kushman, N., & Hajishirzi, H., (2016). MAWPS: A math word problem repository. In *Naacl* (pp. 1152–1157).
- Lee, D., Ki, K. S., Kim, B., & Gweon, G. (2021, May). TM-generation model: a template-based method for automatically solving mathematical word problems. *The Journal of Supercomputing*. Retrieved 2021-05-19, from <https://doi.org/10.1007/s11227-021-03855-9>

² <http://www.marketing-interactive.com/baidus-zuoyebang-attracts-outside-investors/>.

- Lewis, M., Liu, Y., Goyal, N., Ghazvininejad, M., Mohamed, A., Levy, O., ... Zettlemoyer, L., (2020). BART: denoising sequence-to-sequence pre-training for natural language generation, translation, and comprehension. In D. Jurafsky, J. Chai, N. Schluter, & J. R. Tetreault (Eds.), *Acl* (pp. 7871–7880). Association for Computational Linguistics.
- Liang, C., Hsu, K., Huang, C., Li, C., Miao, S., & Su, K., (2016a). A tag-based statistical English math word problem solver with understanding, reasoning and explanation. In *Ijcai* (pp. 4254–4255).
- Liang, C.-C., Hsu, K.-Y., Huang, C.-T., Li, C.-M., Miao, S.-Y., & Su, K.-Y., (2016b). A tag-based English math word problem solver with understanding, reasoning and explanation. In *Naacl* (pp. 67–71).
- Liang, Z., Zhang, J., Shao, J., & Zhang, X., (2021). MWP-BERT: A strong baseline for math word problems. *CoRR, abs/2107.13435*. Retrieved from <https://arxiv.org/abs/2107.13435>
- Lu, P., Gong, R., Jiang, S., Qiu, L., Huang, S., Liang, X., & Zhu, S. (2021). Inter-GPS: Interpretable geometry problem solving with formal language and symbolic reasoning. In C. Zong, F. Xia, W. Li, & R. Navigli (Eds.).
- Mitra, A., & Baral, C., (2016). Learning to use formulas to solve simple arithmetic problems. In *Acl*.
- Mukherjee, A., & Garain, U. (2008, April). A review of methods for automatic understanding of natural language mathematical problems. *Artif. Intell. Rev.*, 29(2), 93–122.
- Roy, S., & Roth, D., (2015). Solving general arithmetic word problems. In *Emnlp* (pp. 1743–1752).
- Roy, S., & Roth, D., (2016). Illinois math solver: Math reasoning on the web. In *Naacl*.
- Roy, S., & Roth, D., (2017). Unit dependency graph and its application to arithmetic word problem solving. In *AAAI* (pp. 3082–3088). AAAI Press.
- Roy, S., Upadhyay, S., & Roth, D., (2016). Equation parsing : Mapping sentences to grounded equations. In *EMNLP* (pp. 1088–1097).
- Roy, S., Vieira, T., & Roth, D., (2015). Reasoning about quantities in natural language. *TACL*, 3, 1–13.
- Sachan, M., Dubey, A., & Xing, E. P., (2017). From textbooks to knowledge: A case study in harvesting axiomatic knowledge from textbooks to solve geometry problems. In *Emnlp* (pp. 784–795).
- Seo, M. J., Hajishirzi, H., Farhadi, A., & Etzioni, O., (2014). Diagram understanding in geometry questions. In *Aaaa* (pp. 2831–2838).
- Seo, M. J., Hajishirzi, H., Farhadi, A., Etzioni, O., & Malcolm, C., (2015). Solving geometry problems: Combining text and diagram interpretation. In *Emnlp* (pp. 1466–1476).
- Shapiro, L. G., & Stockman, G. C., (2001). *Computer vision*. Prentice Hall.
- Silver, D., Huang, A., Maddison, C. J., Guez, A., Sifre, L., van den Driessche, G., ... Hassabis, D., (2016). Mastering the game of go with deep neural networks and tree search.
- Silver, D., Schrittwieser, J., Simonyan, K., Antonoglou, I., Huang, A., Guez, A., ... Bolton, A., (2017). Mastering the game of go without human knowledge. *Nature*, 550(7676), 354.
- Sundaram, S. S., & Khemani, D. (2015, December). Natural language processing for solving simple word problems. In *Proceedings of the 12th international conference on natural language processing* (pp. 394–402). Trivandrum, India: NLP Association of India.
- Sutskever, I., Vinyals, O., & Le, Q. V., (2014). Sequence to sequence learning with neural networks. In *Nips* (pp. 3104–3112).
- Wang, L., Wang, Y., Cai, D., Zhang, D., & Liu, X., (2018a). Translating math word problem to expression tree. In *EMNLP* (pp. 1064–1069). Association for Computational Linguistics.
- Wang, L., Zhang, D., Gao, L., Song, J., Guo, L., & Shen, H. T., (2018b). MathDQN: Solving arithmetic word problems via deep reinforcement learning. In *AAAI* (pp. 5545–5552). AAAI Press.
- Wang, L., Zhang, D., Zhang, J., Xu, X., Gao, L., Dai, B., & Shen, H. T., (2019). Template-based math word problem solvers with recursive neural networks. In *AAAI*. AAAI Press.
- Wang, Y., Liu, X., & Shi, S., (2017). Deep neural solver for math word problems. In *Emnlp* (pp. 845–854).
- Xie, Z., & Sun, S., (2019). A goal-driven tree-structured neural model for math word problems. In S. Kraus (Ed.), *Ijcai* (pp. 5299–5305).

- Yun, R., Yuhui, M., Guangzuo, C., Ronghuai, H., & Ying, Z. (2010, 03). Frame-based calculus of solving arithmetic multi-step addition and subtraction word problems. In *Education technology and computer science, international workshop on(ETCS)* (Vol. 02, p. 476-479). Retrieved from doi.ieeecomputersociety.org/10.1109/ETCS.2010.316 doi:10.1109/ETCS.2010.316
- Zhang, J., Wang, L., Lee, R. K., Bin, Y., Wang, Y., Shao, J., & Lim, E., (2020). Graph-to-tree learning for solving math word problems. In D. Jurafsky, J. Chai, N. Schluter, & J. R. Tetreault (Eds.), *Acl* (pp. 3928–3937). Association for Computational Linguistics.

Open Access This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.



Recent Advances in Intelligent Textbooks for Better Learning



Bo Jiang, Meijun Gu, and Ying Du

Contents

1 Introduction	247
2 The Development of Intelligent Textbooks	248
3 Intelligent Tutoring Technologies of Intelligent Textbooks	249
3.1 Domain Modeling Technologies in the Intelligent Textbook	249
3.2 Student Modeling Technologies	252
3.3 Instructional Technologies	253
4 Evaluation of Intelligent Textbooks	255
4.1 Students' Comments on Intelligent Textbooks	255
4.2 The Effectiveness of Intelligent Textbooks	255
5 Discussions and Conclusions	256
References	257

1 Introduction

Intelligent textbooks embed digital textbooks with intelligent tutoring technologies to provide intelligent reading support to students. Intelligent textbooks not only provide interactions that traditional digital textbooks have, such as highlighting, underlining, and note-taking, but also attempt to understand why readers interact with the textbooks and then build scaffoldings to enhance reading experiences. For example, the intelligent textbook *Inquire Biology* could actively ask the reader a question to promote deep thinking when the reader highlights a sentence. Also, the reader could raise questions to the textbook, which would respond to them using the reasoning technologies (Chaudhri et al. 2013). Over the last 30 years, intelligent textbooks have been used in many schools. Some recent empirical studies about the usage of intelligent textbooks have demonstrated their abilities to improve students'

B. Jiang (✉) · Y. Du

East China Normal University, Shanghai, China
e-mail: bjiang@deit.ecnu.edu.cn

M. Gu
Zhejiang University of Technology, Hangzhou, China

learning gain (Chaudhri et al. 2014; Ericson 2019; Kim et al. 2020; Koć-Januchta et al. 2020).

This chapter offers a state-of-the-art overview of intelligent textbooks. The overview is divided into three parts. The first part focuses on the history of intelligent textbooks and attempts to answer the question: *What are the intelligent textbooks and which authoring tools can be used to create the intelligent textbooks?* The second part focuses on the technologies behind the intelligent textbooks and attempts to answer the question: *What mechanism makes a textbook intelligent?* The third part focuses on the usage of intelligent textbooks and attempts to answer the question: *What is the effect of intelligent textbooks on students' learning?* The last section discusses the future and challenges of intelligent textbooks.

2 The Development of Intelligent Textbooks

The emerging of intelligent textbooks was driven by the idea of combining adaptive hypermedia systems and intelligent tutoring systems (ITS). An earlier attempt at intelligent textbook named *ELM-ART* was proposed by (Brusilovsky et al. 1996a, b) to develop an interactive and adaptive Web-based programming textbook with problem-solving support. The *ELM-ART* enables students to explore program examples by running them with different parameters, interactively solving problems, and receiving instant feedback. It also provides individual curriculum sequencing based on students' learning status on the previously visited pages to suggest the next best pages to work on. Although *ELM-ART* can only offer adaptive multimedia, text presentation, as well as navigation support, it provides a design paradigm of the intelligent textbook that inspired many other studies of this area in the first decade of the twenty-first century.

With the rapid development of artificial intelligence (AI), the recent intelligent textbooks provide more sophisticated learning services, such as automatic resource matching, automatic question answering, personalized learning evaluation, and planning. For example, *Interlingua* is an intelligent platform where students can study textbooks in a foreign language supported by on-demand access to relevant reading material in their mother language (Alpizar-Chacon and Sosnovsky 2019). *FlexBooks* is a math & science textbook platform designed to suit learners' learning styles, regions, languages, or skill levels and allows learners to customize content (Lindshield and Adhikari 2013). *OpenDSA* is an interactive textbook for data structures and algorithms courses involving the use of many algorithm visualizations and a wide range of automatic exercises assessment (Shaffer et al. 2011). Another tool for studying computer science is *Runestone*. It incorporates code visualizations and customizes interactive course materials (Miller and Ranum 2014). *Reading Mirror* is an online reading system that permits students to track their reading progress and compare with peers through a mirrored icicle plot visualization (Barria-Pineda et al. 2019). *PASTEL* is an online courseware authoring platform that applies embedded skill model and cognitive tutors to divide assessment items

into clusters with similar semantic meanings and perform on-demand hints on how to perform the next step (Matsuda and Shimmei 2019). Other intelligent textbook authoring platforms are shown in (Table 1).

3 Intelligent Tutoring Technologies of Intelligent Textbooks

The intelligent tutoring system Brusilovsky et al. (1996b) is formalized by three models: domain, student, and instruction. While it is designed to make use of students' answering questions or testing data to intervene and regulate students' learning in real-time, intelligent textbooks combine AI technologies with electronic textbooks; in addition to collecting the result data generated by the exercises and tests in the textbook, it also mines and analyzes the data generated during the process of using textbooks. Developing intelligent textbooks are based on the idea of ITS (Boulanger and Kumar 2019). The domain model is a knowledge base and ontology that stores and codifies a vast amount of knowledge of specific subjects via taxonomies, examples, exercises, and so on. The student model identifies a student's knowledge state and how it evolves during learning. The instruction model specifies a policy for administering automated instructional actions that are conditioned on the student.

3.1 *Domain Modeling Technologies in the Intelligent Textbook*

The domain model provides the knowledge base of an intelligent textbook. Usually, an authoring tool or platform is required for instructors to manually create learning content, build scaffoldings, and link resources. This process is incredibly time-consuming and expensive, and some recent efforts are invested to develop automated modeling technologies to save expert effort. Domain knowledge is complicated and currently, we cannot expect technologies to generate delicate domain knowledge, but they can replace or assist humans in knowledge annotation. Knowledge annotation is a fundamental but critical component of intelligent textbooks as automated algorithms like machine learning algorithms need well-labeled data as the training samples. Without high-quality annotated data, intelligent linking, matching, and recommendation services could not be implemented. Current efforts in automatic knowledge annotation can be simply categorized into the following three categories.

The first approach is the automatic concept extraction that extracts concepts and knowledge from text automatically. Although a wide range of concept extraction methods has been developed, few have been applied in intelligent textbooks context. According to what features are used, three popular approaches for concept extraction are the pure word-based method (bag-of-words), chapter-based method (coarse-grained semantic-based), and latent topic-based method (fine-grained semantic-based). Huang et al. (2016) compared the three approaches and found that the latent topic-based method outperformed the others on predicting

Table 1 Some intelligent textbook authoring platforms

Platform name	Creation date	Target subject(s)	Introduction
CS Circles	2010	Python programming	It provides students with instant feedback on code exercises and tracks their programming progress (Pritchard and Vasiga 2013)
<i>Inquire biology</i>	2012	Biology	It is a prototype of an intelligent textbook. It provides unique capabilities through a knowledge representation that captures conceptual knowledge from the textbook and uses automated question-answering technology to answer students' questions (Chaudhri et al. 2013)
nyBooks	2012	STEM, computer science, etc.	It uses interactive animations to demonstrate the concepts and formulas in the textbook to provide instant feedback and analysis; it also provides personalized questions with gradients of difficulty (Liberatore 2017)
SmartBook	2013	Reading	Based on question-answering technology and interactive functions, it tracks students' behavior, estimates their knowledge level, and plans a suitable learning path for them. It highlights key learning objects and provides links to external learning materials (Koychev et al. 2009)
BBookX	2015	General	It automatically and collaboratively builds free open personalized online books using publicly available educational resources such as Wikipedia (Liang et al. 2015)
IntDynGeoBook	2017	Geometry	As an intelligent textbook, when readers ask closed or open-ended questions, the textbook can provide interactive exercises with hints and automatic corrections (Quaresma 2017)
iRead	2017	Reading	It detects students' reading skills by analyzing their profile and reading history logs and supports the creation of interactive e-textbooks based on personalized learning content and activities (Deligiannis et al. 2019)
CyberBook	2019	Mathematics, science	It provides cognitive tutor support based on cognitive models, applies knowledge tracing to dynamically control the number of test items, and supplies auxiliary learning materials (Matsuda and Shimmei 2019)
MoFaCTS	2020	Anatomy and physiology	It automatically creates cloze sentence practice content, generates refutational feedback for incorrect cloze responses, and a dialog system, which will trigger a short conversation by a tutor to correct conceptual misunderstandings (Pavlik et al. 2020)
LiveHint	2020	Mathematics	It is a dialog-driven textbook using chatbots, which support intelligent hints (Fisher et al. 2020)
Intextbooks	2020	General	With domain modeling and adaptive navigation technology, it automatically converts e-textbooks in PDF format into interactive intelligent Web resources (Alpiraz-Chacon and Sosnovsky 2020)

students' knowledge acquisition state after reading textbooks. To extract concepts from text automatically, Chau et al. (2021) proposed a supervised feature-based machine learning method that uses multi-view features, including linguistic-based, statistics-based, title-based, and external resources-based features. The proposed method outperformed several state-of-the-art concept extraction approaches. Furthermore, some concept extraction technologies focus on using formatting rules and internal structures of textbooks (Alpizar-Chacon and Sosnovsky 2020) or discourse and text layout features of textbooks (Sachan et al. 2019). Although several new features and technologies can be used for concept extraction, their performances are still very low, which makes them not effective enough to use in real-world tasks. Human extraction is still the most reliable approach. Most recently, Wang et al. (2021) proposed a team-based systematic knowledge engineering approach for fine-grained concept annotation of textbooks.

The second approach is the automatic concept relationship extraction, including internal relationships (hierarchy concepts or prerequisite concepts) as well as external relationships. (Guerra et al. 2013) proposed a latent Dirichlet allocation (LDA)-based method to generate intelligent links among textbooks sections that presented a similar topic based on the LDA model. Wang et al. (2015) argued the concept hierarchy in textbooks is not only decided by the relatedness between the concept and the subchapter but also by the coherence between this concept and the concepts in the same/different subchapter(s). They furtherly formalized the concept extraction from the textbook as an optimization problem and combined local features and global features to train a support vector machine to extract concept hierarchies. Labutov et al. (2017) proposed two probabilistic graphical models to identify outcome and prerequisite concepts on six textbooks and demonstrated improvements over several baselines of automatic concept linking. Meng et al. (2017) explored multiple knowledge-based contents linking algorithms for connecting online resources with textbooks, and this algorithm reported its value for improving textbook subsection linking performance. Alpizar-Chacon and Sosnovsky (2021) presented an extensible linking model to enrich textbook contents connected with internal or external resources with the help of DBpedia.

A third strategy is to extract concepts and relationships among concepts simultaneously. For example, Lu et al. (2019) created a learning graph by classifying semantically similar chapters via an unsupervised clustering method, then extracted the structural relationship, and built the metro map by applying an integer linear programming-based technique. Wang et al. (2016) proposed a concepts extraction and concept relationship-building framework using the knowledge maps of textbooks. Sastry et al. (2017) extracted concept relationships through an elegant algorithm of the idea of transitive closure and visualized the concept relationship as a network graph. The *Interlingua* is an intelligent tool that links textbooks in different languages covering the same topic (Alpizar-Chacon and Sosnovsky 2019). The *Interlingua* first extracts index terms and pages referenced by the terms from the textbook and then uses them as semantic anchors to link pages and sections of the textbook to the concepts and through them to other textbooks available in the repository.

3.2 Student Modeling Technologies

An important feature of distinguishing an intelligent textbook from a normal digital textbook is whether it provides personalized learning services. Student modeling aims to understand students' learning using their interaction data as they work on problems in the text. The student model drives the learning system to adapt to the needs and knowledge of students. Generally, a completed student model contains students' knowledge state, behavior patterns, learning emotional state, as well as some domain-independent traits such as cognitive ability, learning style, motivation, and attitude.

One of the most popular student modeling approaches in ITS is "knowledge tracing," which aims to predict students' knowledge acquisition state using their performance data. Three popular knowledge tracing methods are Bayesian Knowledge Tracing (Corbett and Anderson 1995), logistic model (Pelánek 2017), and deep knowledge tracing (Piech et al. 2015). The Bayesian Knowledge Tracing uses a hidden Markov chain to estimate knowledge mastery probabilities, and the logistic model combines multiple factors that affect learning into a logistic regression model to make predictions; the deep knowledge tracing applies a long short-term memory neural network to model student learning. However, these well-explored approaches could not be directly used in intelligent textbooks, as these methods require students' response data that is generated in solving problems, yet the most frequent learning activity in textbook-based learning is reading.

Recently, Mouri et al. (2016) analyzed the relationship between students' e-book reading time and their final grade using the Bayesian network based on association analysis with social network analysis. They found that more time devoted to reading the e-book before the class was associated with a higher final grade. Meanwhile, Huang et al. (2016) incorporated the reading time variable into a Bayesian Knowledge Tracing model and two logistic models to predict students' acquisition state on the concepts covered by a textbook. This study serves as the first step to construct a dynamic knowledge tracing model in intelligent textbooks. However, only considering reading time is not robust as students' reading logs are noisy. For example, we cannot identify whether a student read a specific page even if he or she opened the page and kept it open there for a long time. Thaker et al. (2018) incorporated both the reading data and the performance data in an improved Bayesian Knowledge Tracing model. The comparison results show that the model using two-view data significantly outperformed the model that only uses reading data and the model that only considers quiz performance data. Furthermore, Thaker et al. (2019) presented a logistic model that also takes into account students' previous performances and reading behaviors to predict their success rate for a given question. Okubo et al. (2018) also used students' reading time in an e-book system and previous quiz scores to predict their final grades. Besides the reading time, other reading behaviors such as underlining and highlighting can also be used to predict students' performance Okubo et al. (2017). Kim et al. (2020) investigated whether students' comprehension and knowledge retention could be predicted by

their highlighting behavior. The data analysis suggests that when students choose to highlight, the specific pattern of highlights can explain about 13% of the variance in observed quiz grades.

Students' reading behavior also helps us to understand students' preferences and cognitive features. For example, recent studies used clustering algorithms and lag sequence analysis to explore students' reading behavior patterns in using an e-book. They found a very interesting phenomenon that students always use the memos and bookmarkers function rather than underlines and highlights (Yin et al. 2019; Yin and Hwang 2018). With students' reading behavior data, Gu et al. (2020) applied multiple classification models, including logistic regression, support vector machine, and decision tree to predict students' learning styles. The results show that the decision tree achieves promising performance in the prediction of learning style.

The domain-independent traits describe student profiles of cognitive ability, learning style, motivation, attitudes, working memory capacity, and emotions when using cognitive processing skills and strategies, such as induction and reasoning in the process of selecting and acquiring knowledge. A variety of technologies in cognitive science and psychometrics are being used to measure learners' traits. For example, *ELM-ART* intelligent textbook platform can diagnose learners' cognitive abilities changes of programming process based on example-based and constraint-based model (Weber and Brusilovsky 2016). A new didactical model for modern online textbooks was applied for developing student self-regulated competence (Railean 2010). A personalized recommendation mechanism was presented through some information about the individual cognitive levels and learning styles (Sun et al. 2013). Besides, some recent studies also used wearable smart devices like eye tracking (Ishimaru et al. 2016) and Kinect (Lin et al. 2017) to track students' attention and emotional state.

3.3 Instructional Technologies

An instructional model takes the domain and student model as input and determines what next information to present to the student. This section summarizes several instructional technologies utilized in intelligent textbooks, including hyperlink annotation and direct navigation support, error-sensitive feedback, tutoring dialog instruction, and content presentation orders.

Hyperlink annotation and direct navigation support are the most frequently used instructional techniques in intelligent textbooks. Online textbooks contain several types of instructional resources, such as graphics, audio, videos, and plain texts. Hyperlink annotation instruction is used to create a nonlinear medium among these multimedia. The navigation support instruction is to guide learners through hyperspace by making direct next-link suggestions. Nowadays, these instruction techniques extend to intelligent links, semantic relationships, concept mapping, knowledge graphs, and so on. For example, *KBS-HyperBook* created intelligent links to external Web learning resources to satisfy learners' knowledge, goals, and

preferences on Java programming (Henze and Nejdl 2001). *Wikibooks* provided intelligent links instruction to the course concepts in the collaborative textbook (O’Shea 2011). *Interlingua* connected automated semantic relationships of sections and subsections across textbooks with on-demand access to relevant reading material in their mother tongue (Alpizar-Chacon and Sosnovsky 2019). *MM4Books* automatically build metro knowledge graphs among massive electronic textbooks (Lu et al. 2019). Another study proposed a concept mapping instruction method that allows students to link words in the textbook (Wang et al. 2017).

Error-sensitive feedback is an instruction technique to be given when learners answer a question incorrectly, are unsure of a correct answer, or repeatedly request help. This technology can not only judge whether an answer is correct or not but also mainly aim to fix students’ misunderstandings. For example, *CS Circle* tracked their programming progress and gave instant feedback on code exercises (Pritchard and Vasiga 2013). *IntDynGeo Book* offered hints and automatic corrections about geometry knowledge (Billingsley and Robinson 2005). *Intextbooks* developed interactive assessment question components to fix students’ knowledge concepts (Alpizar-Chacon and Sosnovsky 2020).

Tutoring dialog is an instructional technique that uses natural language processing to engage students in interactive dialogs. These tutoring dialogs often supply guidance for during problem-solving and motivational supports. For example, the intelligent textbook, *Inquiry* used *inquiry-based instruction* through a question-asking dialog to ask the student a question if they highlight a word or sentence (Chaudhri et al. 2014). Another intelligent textbook, *MoFaCTS*, provided a dialog system to correct student conceptual misunderstandings of cloze sentence practice contents (Pavlik et al. 2020). *LiveHint* is a dialog-driven textbook via a chatterbot with access to thousands of context-sensitive hints (Fisher et al. 2020).

Personalized content sequencing is another instruction technique that has the function of organizing sequential KCs and then presenting students with learning paths. One example is *SmartBook*, which implemented a tailor-made courseware solution for learners (Koychev et al. 2009). Another textbook is *iRead* that provided personalized learning content and activities by analyzing their profiles and reading history logs (Deligiannis et al. 2019).

Furthermore, there are other instruction techniques rarely used in intelligent textbooks. For example, in the intelligent textbook, *Runstone* applied the *learning-by-doing* strategy that encourages students to experiment with examples as they are reading (Ericson 2019). *Runstone* also provides a visualization tool to demonstrate and control the *step-by-step execution* of a program. Like *Runstone*, *FlexBooks* also provides an interactive simulation tool that supports *learning by playing* (Lindshield and Adhikari 2013).

4 Evaluation of Intelligent Textbooks

Reviewing the development in the past 10 years, researchers have carried out many empirical studies in schools, demonstrating the effectiveness of intelligent textbooks. According to these findings, intelligent textbooks were exceptional in facilitating students' reading and learning. Meanwhile, combined with the users' reflections of intelligent textbooks, the promising prospects of this new form of the digital textbook could be expected.

4.1 Students' Comments on Intelligent Textbooks

It was gratifying that most students made positive evaluations of intelligent textbooks. Users' evaluation of the popular intelligent textbook *ELM-ART* proved that students had high satisfaction with intelligent textbooks and expressed a strong willingness to continue to use them (Weber and Brusilovsky 2001). Another investigation shows that when students were faced with static PDF textbooks and interactive intelligent textbooks (their content was the same), students were more inclined to use intelligent textbooks (Pollari-Malmi et al. 2017). Most students believed that intelligent textbooks altered their learning patterns (Barria-Pineda et al. 2019).

Pursel et al. (2019) present an intelligent textbook authoring tool that can retrieve open educational resources from Wikipedia for users to create their books. The responses from the student survey indicated generally favorable reactions when asked questions about this intelligent textbook compared to a traditional textbook. Most recently, Feng and Li (2019) developed an offline-to-online intelligent textbook that grade and correct students' calculation in a paper-based workbook automatically by cell phone's camera and then use it to provide adaptive tutor service to students. An investigation showed that more than 30% have become active users and more than 20% of active users have recommended it to others.

4.2 The Effectiveness of Intelligent Textbooks

Inspired by the positive influence of social learning, the intelligent textbook *Reading Mirror* extended social navigation with social comparison. It enabled students to visually track their reading and test progress through icicle plots and compared them with their peers. Researchers have performed a series of classroom studies in three different courses. They proved that the *Reading Mirror* could help students ($N = 200$) focus on the most important pages and increase their reading engagement. The social comparison would encourage students to work harder and achieve higher achievement in quizzes (Barria-Pineda et al. 2019). Researchers have used

Runestone to create several free intelligent textbooks for introductory computing courses. By analyzing the log files, they reported that owing to various interactive components, the intelligent textbooks created by Runestone improved students' learning gains and motivation in programming (Ericson 2019). The results of a large-scale study ($N > 600$) showed that in programming courses, interactive intelligent textbooks were more conducive to enhancing students' learning motivation, gains, and feedback on learning resources than static PDF format textbooks (Pollari-Malmi et al. 2017).

Intelligent textbooks also exerted unexpected benefits for teachers. In a small simple study of high school teachers ($N = 10$), they used the intelligent textbooks developed by Runestone, which helped them improve their professional knowledge and teaching confidence (Ericson et al. 2015). Some studies also showed the positive effect of intelligent textbooks in improving students' academic performance. For example, *Inquire Biology* significantly improved students' homework quiz scores ($p = 0.02$) and quiz scores ($p = 0.05$) (Chaudhri et al. 2013). The intelligent textbook created based on *ELM-ART* significantly improved the test scores of those students with weak programming skills ($p = 0.011$) (Weber and Brusilovsky 2001).

It was worth noting that not all intelligent textbooks could help students achieve expected learning gains. Just like the *Math CyberBook* (Matsuda and Shimmei 2019), it did not achieve a significant impact on students' academic performance ($p = 0.63$). The reason for this phenomenon needs further analysis. Moreover, students did not achieve the expected learning progress in the first 3 weeks of using the intelligent textbook created by *Reading Mirror* (Barria-Pineda et al. 2019). After comparative analysis, researchers believed that one of the reasons that could explain this issue was that students needed time to adapt to the social comparison feature. Maybe it proved that some external conditions should be satisfied for its desired functions to work.

5 Discussions and Conclusions

Intelligent textbooks have attracted much attention in the past decade, with increasing evidence demonstrating their positive influences on improving students' reading and learning. A short review of tools, adaptation technologies, and evaluations provided in this chapter could serve as a collection of useful information for the researchers and developers of the next generation of intelligent textbooks. Although intelligent textbook research has made big progress in the past decade, many crucial technical and usage problems remain unsolved. For example, current technologies cannot understand the mathematical language within the textbooks very well, which seriously hinders the development of mathematical intelligent textbooks. Also, authoring a new intelligent textbook is expensive, so while making the huge quantity of existing PDF-based digital textbooks intelligent is very necessary, it is challenging (Alpizar-Chacon et al. 2021). Another area of future work is

interconnecting intelligent textbooks, learning management systems, practices, and exams to construct a closed intelligent learning loop.

References

- Alpizar-Chacon, I., & Sosnovsky, S. (2019). Interlingua: Linking textbooks across different languages. In: Proceedings of the first workshop on intelligent textbooks, international conference on artificial intelligence in education, Chicago, IL, USA, 25–29 June 2019. <https://dspace.library.uu.nl/handle/1874/385766>.
- Alpizar-Chacon, I., & Sosnovsky, S. (2020). Order out of chaos: Construction of knowledge models from pdf textbooks. In: Proceedings of the ACM Symposium on Document Engineering, CA, United States, 29 September - 1 October 2020, p 1–10. <https://doi.org/10.1145/3395027.3419585>.
- Alpizar-Chacon, I., & Sosnovsky, S. (2021). Knowledge models from PDF textbooks. *New Review of Hypermedia and Multimedia*, 1–49 (2021). <https://doi.org/10.1080/13614568.2021.1889692>.
- Alpizar-Chacon, I., van der Hart, M., Wiersma, Z. S., Theunissen, L., Sosnovsky, S. (2021). Transformation of PDF textbooks into intelligent educational resources. In: Proceedings of the third workshop on intelligent textbooks, international conference on artificial intelligence in education, Utrecht University, Utrecht, Netherlands, 14–18 June 2021, p 60–73. <http://ceur-ws.org/Vol-2674/paper01.pdf>.
- Barria-Pineda, J., Brusilovsky, P., He, D. (2019). Reading mirror: Social navigation and social comparison for electronic textbooks. In: Proceedings of the first workshop on intelligent textbooks, international conference on artificial intelligence in education, Chicago, IL, USA, 25–29 June 2019. <http://ceur-ws.org/Vol-2384/paper03.pdf>.
- Billingsley, W., Robinson, P. (2005). Towards an intelligent online textbook for discrete mathematics. In: Proceedings of the 2005 International Conference on Active Media Technology, Kagawa, Japan, 19–21 May 2005, p 291–296. <https://doi.org/10.1109/AMT.2005.1505353>.
- Boulanger, D., & Kumar, V. (2019). An Overview of Recent Developments in Intelligent e-Textbooks and Reading Analytics. In: Proceedings of the first workshop on intelligent textbooks, international conference on artificial intelligence in education, Chicago, IL, USA, 25–29 June 2019. <http://ceur-ws.org/Vol-2384/paper05.pdf>.
- Brusilovsky, P., Schwarz, E., Weber, G. (1996a). A tool for developing adaptive electronic textbooks on WWW. In: Proceedings of WebNet'96, World Conference of the Web Society, San Francisco, CA, 1996. <https://eric.ed.gov/?id=ED427653>.
- Brusilovsky, P., Schwarz, E., Weber, G. (1996b). ELM-ART: An intelligent tutoring system on World Wide Web. In C. Frasson, G. Gauthier, A. Lesgold (Ed.), *Intelligent Tutoring Systems*. ITS 1996. Lecture Notes in Computer Science (vol 1086, pp. 261–269). Springer, Berlin, Heidelberg. https://doi.org/10.1007/3-540-61327-7_123.
- Chau, H., Labutov, I., Thaker, K., He, D., Brusilovsky, P. (2021). Automatic concept extraction for domain and student modeling in adaptive textbooks. *International Journal of Artificial Intelligence in Education*, 31, 820–846 (2021). <https://doi.org/10.1007/s40593-020-00207-1>.
- Chaudhri, V.K., Cheng, B., Overholtzer, A., Roschelle, J., Spaulding, A., Clark, P., Greaves, M., Gunning, D. (2013). Inquire biology: A textbook that answers questions. *AI Magazine*, 34(3), 55–72. <https://doi.org/10.1609/aimag.v34i3.2486>.
- Chaudhri, V.K., Overholtzer, A., Spaulding, A. (2014). An intelligent textbook that answers questions. In: P. Lambrix et al. (Ed.), *Knowledge Engineering and Knowledge Management*. EKAW 2014. Lecture Notes in Computer Science (vol 8982, pp. 131–135). Springer, Cham. https://doi.org/10.1007/978-3-319-17966-7_16.

- Corbett, A.T., & Anderson, J.R. (1995). Knowledge tracing: Modeling the acquisition of procedural knowledge. *User modeling and user-adapted interaction*, 4(4), 253–278. <https://doi.org/10.1007/BF01099821>.
- Deligiannis, N., Panagiotopoulos, D., Patsilinakos, P., Raftopoulou, C., Symvouli, A. (2019). Interactive and Personalized Activity eBooks for Learning to Read: The iRead Case. In: Proceedings of the first workshop on intelligent textbooks, international conference on artificial intelligence in education, Chicago, IL, USA, 25–29 June 2019. <http://ceur-ws.org/Vol-2384/paper06.pdf>.
- Ericson, B.J. (2019). An analysis of interactive feature use in two ebooks. In: Proceedings of the first workshop on intelligent textbooks, international conference on artificial intelligence in education, Chicago, IL, USA, 25–29 June 2019. <http://ceur-ws.org/Vol-2384/paper01.pdf>.
- Ericson, B.J., Guzdial M.J., Morrison B.B. (2015). Analysis of interactive features designed to enhance learning in an ebook. In: Proceedings of the eleventh annual International Conference on International Computing Education Research, Omaha, Nebraska, USA, 9–13 July 2015, p 169–178. <https://doi.org/10.1145/2787622.2787731>.
- Feng, J.C., & Li, K. (2019). Tobbits calculation workbook: An offline-to-online intelligent textbook. In: Proceedings of the first workshop on intelligent textbooks, international conference on artificial intelligence in education, Chicago, IL, USA, 25–29 June 2019. <http://ceur-ws.org/Vol-2384/paper12.pdf>.
- Fisher, J., Fancsali, S., Lewis, A., Fisher, V., Hausmann, R., Pavelko, M., Finocchi, S.B., Ritter, S. (2020). LiveHint: Intelligent Digital Support for Analog Learning Experiences. In: Proceedings of the second workshop on intelligent textbooks, international conference on artificial intelligence in education, Ifrane, Morocco, 6–10 July 2020. <http://ceur-ws.org/Vol-2674/paper08.pdf>.
- Gu, M., Jiang, B., Yin, C. (2020). Learning style prediction using students' E-textbook reading behaviors data. In: Proceedings of the 28th International Conference on Computers in Education. Virtual, Online, 23–27 November 2020, v 2, p 332–340. https://apsce.net/icce/icce2020/proceedings/W1-13/W7/ICCE2020-Proceedings-Vol2-W7_7.pdf.
- Guerra, J., Sosnovsky, S., Brusilovsky, P. (2013). When one textbook is not enough: Linking multiple textbooks using probabilistic topic models. In: Scaling up Learning for Sustained Impact. EC-TEL 2013. Lecture Notes in Computer Science (vol 8095, pp. 125–138). Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-40814-4_11.
- Henze, N., & Nejdl, W. (2001). Adaptation in open corpus hypermedia. International Journal of Artificial Intelligence in Education, 12(4), 325–350. <https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.454.3576&rep=rep1&type=pdf>.
- Huang, Y., Yudelson, M.V., Han, S., He, D., Brusilovsky, P. (2016). A framework for dynamic knowledge modeling in textbook-based learning. In: Proceedings of the 2016 conference on user modeling adaptation and personalization, Halifax, Nova Scotia, Canada, 13–17 July 2016, p 141–150. <https://doi.org/10.1145/2930238.2930258>.
- Ishimaru, S., Bukhari, S.S., Heisel, C., Kuhn, J., Dengel, A.R. (2016). Towards an intelligent textbook: eye gaze based attention extraction on materials for learning and instruction in physics. In: Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct, Heidelberg, Germany, 12–16 September 2016, p 1041–1045. <https://doi.org/10.1145/2968219.2968566>.
- Kim, D., Winchell, A., Waters, A., Grimaldi, P., Baraniuk, R., Mozer, M. (2020). Inferring student comprehension from highlighting patterns in digital textbooks: An exploration in an authentic learning platform. In: Proceedings of the second workshop on intelligent textbooks, international conference on artificial intelligence in education, Ifrane, Morocco, 6–10 July 2020. <https://intextbooks.science.uu.nl/workshop2020/>.
- Koć-Januchta, M.M., Schonborn, K.J., Tibell, L.A.E., Chaudhri, V.K., Heller, H.C. (2020). Engaging With Biology by Asking Questions: Investigating Students' Interaction and Learning With an Artificial Intelligence-Enriched Textbook. *Journal of Educational Computing Research*, 58(6), 1190–1224. <https://doi.org/10.1177/0735633120921581>.

- Koychev, I., Nikolov, R., Dicheva, D. (2009). SmartBook: The new generation e-book. In: Proceedings of Books Online, 9, 60–67. <https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.605.9558&rep=rep1&type=pdf>.
- Labutov, I., Huang, Y., Brusilovsky, P., He, D. (2017). Semi-supervised techniques for mining learning outcomes and prerequisites. In: Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Halifax, NS, Canada, 13–17 August 2017, p 907–915. <https://doi.org/10.1145/3097983.3098187>.
- Liang, C., Wang, S., Wu, Z., Williams, K., Pursel, B.K., Brautigam, B., Saul, S., Williams, H., Bowen, K., Giles, C.L. (2015). Bbookx: An automatic book creation framework. In: Proceedings of the 2015 ACM Symposium on Document Engineering, Lausanne, Switzerland, 8–11 September 2015, p 121–124. <https://doi.org/10.1145/2682571.2797094>.
- Liberatore, M.W. (2017). Reading analytics and student performance when using an interactive textbook for a material and energy balances course. In: Proceedings of 2017 ASEE Annual Conference & Exposition, Columbus, Ohio, 24–28 June 2017. <https://doi.org/10.18260/1-2%2D%2D28780>.
- Lin, J.M., Hsu, J.H., Hong, Z.W. (2017). Implementation of an eBook Reader System with the Features of Emotion Sensing and Robot Control. In: J.S. Pan, P.W. Tsai, J. Watada, L. Jain (Ed.), *Advances in Intelligent Information Hiding and Multimedia Signal Processing*. IIH-MSP 2017. Smart Innovation, Systems and Technologies (vol 81, pp. 91–106). Springer, Cham. https://doi.org/10.1007/978-3-319-63856-0_12.
- Lindshield, B., & Adhikari, K. (2013). Online and campus college students like using an open educational resource instead of a traditional textbook. *Journal of Online Learning and Teaching*, 9(1), 26–38. https://jolt.merlot.org/vol9no1/lindshield_0313.htm.
- Lu, W., Ma, P., Yu, J. (2019). Metro maps for efficient knowledge learning by summarizing massive electronic textbooks. *International Journal on Document Analysis and Recognition*, 22, 99–111. <https://doi.org/10.1007/s10032-019-00319-y>.
- Matsuda, N., & Shimmei, M. (2019). PASTEL: Evidence-based learning engineering method to create intelligent online textbook at scale. In: Proceedings of the first workshop on intelligent textbooks, international conference on artificial intelligence in education, Chicago, IL, USA, 25–29 June 2019. <http://ceur-ws.org/Vol-2384/paper07.pdf>.
- Meng, R., Han, S., Huang, Y., He, D., Brusilovsky, P. (2017). Knowledge-Based Content Linking for Online Textbooks. In: Proceedings of 2016 IEEE/WIC/ACM International Conference on Web Intelligence (WI), Omaha, NE, USA, 13–16 October 2016, p 18–25. <https://doi.org/10.1109/WI.2016.0014>.
- Miller, B., & Ranum, D. (2014). Runestone interactive: tools for creating interactive course materials. In: Proceedings of the first ACM conference on Learning, Atlanta, Georgia, USA, 4–5 March 2014, p 213–214. <https://doi.org/10.1145/2556325.2567887>.
- Mouri, K., Okubo, F., Shimada, A., Ogata, H. (2016). Bayesian network for predicting students' final grade using e-book logs in university education. In: Proceedings of the 16th International Conference on Advanced Learning Technologies (ICALT), Austin, TX, USA, 25–28 July 2016, p 85–89. <https://doi.org/10.1109/ICALT.2016.27>.
- Okubo, F., Atsushi, S., Yuta, T. (2017). A visualization system for predicting learning activities using state transition graphs. In: Proceedings of the 14th International Conference on Cognition and Exploratory Learning in the Digital Age, Vilamoura, Algarve, Portugal, 18–20 October 2017. <https://eric.ed.gov/?id=ED579452>.
- Okubo, F., Yamashita, T., Shimada, A., Taniguchi1, Y., Konomi1, S. (2018). On the Prediction of Students' Quiz Score by Recurrent Neural Network. In: Proceedings of the 8th International Conference on Learning Analytics & Knowledge (LAK18), Sydney, Australia, 6 March 2018. <http://ceur-ws.org/Vol-2163/paper3.pdf>.
- O'Shea, P.M. (2011). A technological reinvention of the textbook: A wikibooks project. *Journal of Digital Learning in Teacher Education*, 27(3), 109–114. <https://doi.org/10.1080/21532974.2011.1078465>.
- Pavlik, P.I., Olney, A.M., Banker, A., Eglington, L., Yarbro J. (2020). The mobile fact and concept textbook system (mofaCTS). In: Proceedings of the second workshop on intelligent textbooks,

- international conference on artificial intelligence in education, Ifrane, Morocco, 6–10 July 2020. <https://intextbooks.science.uu.nl/workshop2020/>.
- Pelánek, R. (2017). Bayesian knowledge tracing, logistic models, and beyond: an overview of learner modeling techniques. *User Modeling and User-Adapted Interaction*, 27(3), 313–350. <https://doi.org/10.1007/s11257-017-9193-2>.
- Piech, C., Spencer, J., Huang, J., Ganguli, S., Sahami, M., Guibas, L., Sohl-Dickstein, J. (2015). Deep knowledge tracing. In C. Cortes, N. D. Lawrence, D. D. Lee, M. Sugiyama, and R. Garnett (Ed.), *Advances in Neural Information Processing Systems* (pp. 505–513). Curran Associates, Inc. <https://arxiv.53yu.com/abs/1506.05908>.
- Pollari-Malmi, K., Guerra, J., Brusilovsky, P., Malmi, L., Sirkiä, T. (2017). On the value of using an interactive electronic textbook in an introductory programming course. In: the 17th Koli Calling International Conference on Computing Education Research, Koli, Finland, 16–19 November 2017, p 168–172. <https://doi.org/10.1145/3141880.3141890>.
- Pritchard, D., & Vasiga, T. (2013). CS circles: an in-browser python course for beginners. In: Proceeding of the 44th ACM technical symposium on Computer science education, Denver, Colorado, USA, 6–9 March 2013, p 591–596. <https://doi.org/10.1145/2445196.2445370>.
- Pursel, B., Ramsay, C., Dave, N., Liang C., Giles, C.L. (2019). BBookX: Creating semi-automated textbooks to support student learning and decrease student costs. In: Proceedings of the first workshop on intelligent textbooks, international conference on artificial intelligence in education, Chicago, IL, USA, 25–29 June 2019. <http://ceur-ws.org/Vol-2384/paper08.pdf>.
- Quaresma, P. (2017). Towards an intelligent and dynamic geometry book. *Mathematics in Computer Science*, 11(3), 427–437. <https://doi.org/10.1007/s11786-017-0302-8>.
- Railean, E. (2010). A new didactical model for modern electronic textbook elaboration. In: Proceedings of the 5th International Conference on Virtual Learning, Romania, Europe, 29–31 October 2010. http://www.icvl.eu/2010/disc/icvl/documente/pdf/met/ICVL_ModelsAndMethodologies_paper16.pdf.
- Sachan, M., Dubey, A., Hovy, E.H., Mitchell, T.M., Roth, D., Xing, E.P. (2019). Discourse in multimedia: A case study in extracting geometry knowledge from textbooks. *Computational Linguistics*, 45(4), 627–665. <https://doi.org/10.1162/COLIa00360>.
- Sastray, C.S., Jagaluru, D.S., Mahesh, K. (2017). Visualizing textbook concepts: Beyond word co-occurrences. In: A. Gelbukh (Ed.), *Computational Linguistics and Intelligent Text Processing*. CICLing 2017. Lecture Notes in Computer Science (vol 10761, pp. 363–376). Springer, Cham. https://doi.org/10.1007/978-3-319-77113-7_29.
- Shaffer, C.A., Karavirta, V., Korhonen, A., Naps, T.L. (2011). OpenDSA: beginning a community active-ebook project. In: Proceedings of the 11th Koli Calling International Conference on computing education research, Koli, Finland, 17–20 November 2011, p 112–117. <https://doi.org/10.1145/2094131.2094154>.
- Sun, Z., Xiaomeng, T., Haijiao, S. (2013). A research about personalized recommendations in e-textbook. In: Proceedings of 2013 Fourth International Conference on Intelligent Systems Design and Engineering Applications, Zhangjiajie, China, 6–7 November 2013, p 44–47. <https://doi.org/10.1109/ISDEA.2013.417>.
- Thaker, K., Huang, Y., Brusilovsky, P., He, D. (2018). Dynamic Knowledge Modeling with Heterogeneous Activities for Adaptive Textbooks. In: Proceedings of the 11th International Conference on Educational Data Mining, Buffalo NY, 15–18 July 2018, p 592–595. <http://dscholarship.pitt.edu/id/eprint/34939>.
- Thaker, K., Brusilovsky, P., He, D. (2019). Student Modeling with Automatic Knowledge Component Extraction for Adaptive Textbooks. In: Proceedings of the first workshop on intelligent textbooks, international conference on artificial intelligence in education, Chicago, IL, USA, 25–29 June 2019. <http://ceur-ws.org/Vol-2384/paper10.pdf>.
- Wang, M., Chau, H., Thaker, K. (2021). Knowledge Annotation for Intelligent Textbooks. *Technology, Knowledge and Learning*. <https://doi.org/10.1007/s10758-021-09544-z>.
- Wang, S., Walker, E., Wylie, R. (2017). What matters in concept mapping? Maps learners create or how they create them. In: André E., Baker, R., Hu, X., Rodrigo, M., Du, B.B. (eds) Artificial

- Intelligence in Education. AIED 2017. Lecture Notes in Computer Science (vol 10331, pp. 406–417). Springer, Cham. https://doi.org/10.1007/978-3-319-61425-0_34.
- Wang, S., Liang, C., Wu, Z., Williams, K., Pursel, B.K., Brautigam, B., Saul, S., Williams, H., Bowen, K., Giles, C.L. (2015). Concept hierarchy extraction from textbooks. In: Proceedings of the 2015 ACM Symposium on Document Engineering, Lausanne, Switzerland, 8–11 September 2015, p 147–156. <https://doi.org/10.1145/2682571.2797062>.
- Wang, S., Ororbia, A.G., Wu, Z., Williams, K., Liang, C., Pursel, B.K., Giles, C.L. (2016). Using prerequisites to extract concept maps from textbooks. In: Proceedings of the 25th ACM International Conference on Information and Knowledge Management, Indianapolis, Indiana, USA, 24–28 October 2016, p 317–326. <https://doi.org/10.1145/2983323.2983725>.
- Weber, G., & Brusilovsky, P. (2001). ELM-ART: An Adaptive Versatile System for Web-based Instruction. *International Journal of Artificial Intelligence in Education (IJAIED)*, 12, 351–384. <https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.332.3900&rep=rep1&type=pdf>.
- Weber, G., & Brusilovsky, P. (2016). ELM-ART-An Interactive and Intelligent Web-Based Electronic Textbook. *International Journal of Artificial Intelligence in Education*, 26(1), 72–81. <https://doi.org/10.1007/s40593-015-0066-8>.
- Yin, C., Ren, Z., Polyzou, A., Wang, Y. (2019). Learning Behavioral Pattern Analysis Based on Digital Textbook Reading Logs. In: N. Streitz, S. Konomi (Ed.), *Distributed, Ambient and Pervasive Interactions*. HCII 2019. Lecture Notes in Computer Science (vol 11587, pp. 471–480). Springer, Cham. https://doi.org/10.1007/978-3-030-21935-2_36.
- Yin, C., & Hwang, G.J. (2018). Roles and strategies of learning analytics in the e-publication era. *Knowledge Management & E-Learning: An International Journal*, 10(4), 455–468. <https://doi.org/10.34105/j.kmel.2018.10.028>.

Open Access This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.



Part IV

**AI and Ethical Challenges in New
Learning Environments**

Ethical Guidelines for Artificial Intelligence-Based Learning: A Transnational Study Between China and Finland



Ge Wei and Hannele Niemi

Contents

1 Introduction	265
2 AI-Based Learning Needs an Ethical Basis.....	266
3 Ethics as a Theoretical Concept	268
4 Research Design	269
5 Chinese and Finnish Contexts	270
5.1 AI in Learning and Education in the Chinese Context.....	270
5.2 AI in Learning and Education in the Finnish Context	270
6 Ethical Guidelines for AI-Based Learning.....	271
6.1 Inclusion and Personalization	271
6.2 Justice and Safety.....	273
6.3 Transparency and Responsibility	274
6.4 Autonomy and Sustainability	276
7 Discussion and Conclusions	277
References	280

1 Introduction

As artificial intelligence (AI) has been a core technology in national power and global competition, it has received much attention and support from global states (Roll and Wylie 2016). During the past 5 years, both the Chinese and Finnish governments have initiated programmatic policies to promote AI development in society. Thus, AI has not merely been a technological or engineering issue but is profoundly associated with ethics. The Global Technology Governance Report (World Economic Forum 2021) urgently demands ethical guidelines for technology development in the Fourth Industrial Revolution. The World Artificial

G. Wei (✉)
Capital Normal University, Beijing, China
e-mail: ge.wei@cnu.edu.cn

H. Niemi
Faculty of Educational Sciences, University of Helsinki, Helsinki, Finland
e-mail: hannele.niemi@helsinki.fi

Intelligence Conference held in July 2021 in Shanghai, China organized a forum about “trustworthy AI,” which implies a series of ethical considerations in AI technology, such as robustness of algorithm, explicability of analytic results, privacy protection in big data, and equality among different user groups (Tao 2021). As Aizenberg and van den Hoven (2020) argued, machine learning and deep learning in AI concern both the accuracy of technological analytics and serving social justice and human rights. However, the fields of learning and education have not paid sufficient attention to ethical issues when AI technologies are applied. Our study fills a gap in reflections on the ethical guidelines of AI-based learning. Through our transnational comparative research, we propose a human-centered stance for a better understanding of AI in education.

We used China and Finland as two contextual cases to conduct our comparative research. We adopted an inductive analytical approach to review the most relevant and latest policy documents in the past decade that have initiated and facilitated AI within and beyond the educational field. Thus, four major themes in the national policies about AI ethics, both in China and Finland as part of Europe, were distilled: (1) inclusion and personalization, (2) justice and safety, (3) transparency and responsibility, and (4) autonomy and sustainability. Our transnational dialogue has implications for a wide range of audiences, including learners, teachers, AI-technology developers, and policy makers. It provides insights for international research and practice in AI-based learning about how to protect human rights, reduce the risks related to technology, and activate human beings’ autonomy and subjectivity in the age of intelligence.

2 AI-Based Learning Needs an Ethical Basis

Recent effective technology and advancements in programming in computer sciences have opened new doors for AI-based teaching and learning (Niemi 2021), and AI has been increasingly adopted into education. It has been widely discussed how AI can increase students’ engagement, leading to improved learning outcomes, integrating technologies involving interactivity, dialogue, automated question generation, and learning analytics (Bozkurt et al. 2021). We have much promising evidence from previous studies, for example, showing that AI has already been utilized in predicting students’ academic achievement, identifying at-risk students in earlier stages, conducting formative assessment, providing descriptive information about teaching and contributing to teacher development, creating flexible and effective learning tools, and implementing adaptive learning environments (e.g., Almohammadi et al. 2017; Baneres et al. 2019; Baradwaj and Pal 2011; Kay 2012; Vinuesa et al. 2020). In a review study, Goksel and Bozkurt (2019) identified three broad themes in AI-based learning: adaptive learning, personalization, and learning styles; expert systems and intelligent tutoring systems; and AI as a future component of educational processes. While the systematic review on AI-based learning offers

great potential if AI is integrated into the educational process, it also raises open questions to be resolved, such as ethical guidelines on the use of AI-based learning tools.

As aforementioned, there is a lack of literature dealing with the ethics of AI-based learning. Nye (2016) claimed that ethics for data sharing are still being revised to accommodate an increasingly connected educational world. They further stressed that no common ethical guidelines exist for processing educational data and this issue has persisted for years (Coeckelbergh 2020). Recently, Niemi (2020, 2021) stated that, although AI in learning has high potential, it also has many limitations. Many worries are linked to ethical issues, such as biases in algorithms, privacy, transparency, and data ownership. To explain the emergent need for ethical guidelines, Mouta et al. (2019) provided some examples demonstrating the lack of “explainability in terms of educational decisions, for example, relating to students allowance or rejection in entering some educational institutions” and “personalized learning by avoiding the personal right to boredom” (p. 2). Thus, educational systems powered by AI, without accounting for ethical considerations, can be seen as black boxes. Another problem that arises in AI is that data is not immune to bias. AI algorithms are designed by programmers and developed by companies or governments; they can include their own agendas or biases in their development stages (Crawford 2021). Such examples enforce the need to increase research on AI ethics in education.

Some studies have begun on the ethical assessment of AI-based learning by international organizations. In 2019, the Beijing Consensus on AI and Education published its document to offer guidance and recommendations on how global states can respond to the opportunities and challenges brought by AI (UNESCO 2019). The consensus reaffirms a humanistic approach to deploying AI technologies in education for augmenting human intelligence, protecting human rights, and promoting sustainable development through effective human–machine collaboration in life, learning, and work. It also elaborates recommendations corresponding to four crosscutting issues: (1) promoting equitable and inclusive use of AI in education; (2) gender-equitable AI; (3) ensuring ethical, transparent, and auditable use of education data and algorithms; and (4) monitoring, evaluation, and research.

In addition, the European Commission, through the High-Level Expert Group on Artificial Intelligence (HLEG 2019), recently released the ethics guidelines for trustworthy AI, and the European Union Parliament (EP 2020) published the European framework on ethical issues of AI. Both reports emphasize European fundamental values of human dignity, freedom, equality, and solidarity and are based on the principles of democracy and the rule of law. The approach “places the individual at the heart of its activities” (EP 2020, p. 5). When speaking about AI and technology, the moral core is freedom, security, and justice. “Trustworthy AI” can be realized by ensuring that the development, deployment, and use of AI systems meet seven key requirements: (1) human agency and oversight; (2) technical robustness and safety; (3) privacy and data governance; (4) transparency; (5) diversity, nondiscrimination, and fairness; (6) environmental and societal well-being; and (7) accountability (HLEG 2019).

In China, the Ministry of Science and Technology (MST 2019) published the principles of AI governance for the next generation. To promote the healthy development of the new generation of AI, the safety, reliability, and controllability of AI need to be ensured, and all parties involved in the development should follow the principles of harmony, friendship, fairness, and justice. Promoting sustainable development in AI-based learning has eight principles: inclusiveness, sharing, respect for privacy, security, control, shared responsibility, open cooperation, and agile governance.

Nevertheless, more studies are urgently required to provide answers about better living with AI in a learning society and what AI means in education. Most strategies are general, covering all domains in which AI can be applied, and we have several guidelines for AI use that cover different sectors of society. However, the ethical principles for AI in education remain largely unexplored and undiscussed in national and global guidelines. This chapter explores the ethical guidelines for AI-based learning from a transnational approach by comparing the national policies of China and Finland. The Chinese and Finnish governments have each emphasized the significant stance of AI in social and economic development, while education as a unique sector requires special ethical guidelines. A comparative policy analysis on AI-driven education between China and Finland can inspire more countries and areas to recognize and reflect on the ethical issues in AI-based learning.

3 Ethics as a Theoretical Concept

Ethics is a starting point to determine what values we want to uphold in the development, design, and deployment of AI. The extension, enhancement, and replacement of human agency and reasoning in AI serve as the loci of many of the ethical issues that arise in its use, sometimes presenting us with vivid versions of classical questions (Boddington 2017).

Tegmark (2017) summarized that “Aristotle emphasized virtues, Immanuel Kant emphasized duties, and utilitarianism emphasized the greatest happiness for the greatest number” (p. 269). There are also deontological theories that emphasize “doing the right thing” and consequentialist theories claiming that the best action is the one that drives the best consequences (Boddington 2017). According to Rawls’s ethical theory, justice is the criterion according to which goods and services are distributed among people (Rawls 1999). Rawls used two principles of reasoning to set out and encapsulate his theory of justice. First, “each person is to have an equal right to the most extensive scheme of equal basic liberties compatible with a similar scheme of liberties for others.” Second, “social and economic inequalities are to be arranged so that they are both (a) reasonably expected to be to everyone’s advantage and (b) attached to positions and offices open to all” (Rawls 1999, p. 53). These two principles raise the questions of how AI can be made available to all so that it does not re-inscribe inequality in power, wealth, income, and other resources. Rawlsian philosophy of ethics enlightened for us that AI is not only a public good

to be equally distributed in society but also a means to promote a better society with equity and justice.

Whittaker et al. (2018) noted that “ethics can only help close the AI accountability gap if they are truly built into the processes of AI development and are backed by enforceable mechanisms of responsibility that are accountable to the public interest.” (p. 9) Characteristic ethical questions regarding AI are typical enhancements or replacements of human agency; crucially, questions of agency and subjectivity are at the heart of how we see ethics (Biesta 2017). Floridi et al. (2018) reviewed several guidelines for ethically sustainable AI policies that lay the foundations for a “good AI society.” They present a synthesis of five ethical principles that should undergird its development and adoption and offer 20 concrete recommendations for national or supranational policy makers and other stakeholders, including beneficence, non-maleficence, autonomy, justice, and explicability, which also serve as a foundation for our discussion of the ethical guidelines in AI-based learning.

In this chapter, we analyze how AI can advance justice and fairness in education and learning and make AI safe for its users. To further focus, our discussion centers on existing policies about how these AI technologies are impacting our lives and reshaping education.

4 Research Design

This chapter is a transnational study of China and Finland as part of the EU. The reason why we chose China and Finland as two contextual cases is that they, respectively, represent an eastern and a western country, a developing and a developed country. Finland plays a double role in the study. It is a nation, but it is also a member state of the EU. Many AI-related issues have been developed in the context of the EU, but Finland also has its own specific national mission. Although many sociocultural differences exist between the two nations, we found the possibility of conducting transnational research due to the similar attention paid by the two governments to AI in learning.

In terms of the collection of policy documents, we separately collected native documents about AI at the national policy level in the past decade (2011–2021). In our analysis, key documents at the official policy-making level were selected (e.g., European Commission 2019, 2020, 2021a, 2021b; Ministry of Economic Affairs and Employment, Finland 2017, 2019; Ministry of Education, China 2020; State Council, China 2021). All policy documents were downloaded or could be fully accessed online.

Then, we used the thematic analysis method (Flick 2006) to distill the major themes about AI ethics amidst the policy documents. The ethical issues have not been particularly clearly answered for the education field, but the governments have been aware of the importance or implied directions for ethical reflections on AI-based learning. Thus, we tried to dig out the ethical principles hidden in

political discourses. Finally, we built four pillars for a trustworthy AI ecosystem for achieving a more equal and democratic education system, which involves (1) inclusion and personalization, (2) justice and safety, (3) transparency and responsibility, and (4) autonomy and sustainability.

The credibility and reliability of our analysis was achieved by constant comparison and triangulation checking between the two authors. Additionally, as international experts in the fields of education and AI-based education, our professional vision on the ethical guidelines for AI in learning can also be regarded as a Delphi demonstration. Nevertheless, more perspectives from other sociocultural contexts are required in further research.

5 Chinese and Finnish Contexts

In this section, we briefly introduce the contexts of China and Finland and the national-level policies related to AI-based learning within and beyond the education sectors during the past decade, both in the two countries.

5.1 *AI in Learning and Education in the Chinese Context*

China is a developing country located in the east of Asia. Since the 2010s, Chinese national policies related to AI and the subsidiary principles of national decision-making in education have been much produced and issued. In March 2017, Prime Minister Keqiang Li mentioned that AI technology should be researched and developed rapidly. The goal is that, by 2030, China's AI technology and application should reach a leading level globally (State Council, China 2017).

In this context, the Ministry of Education (MOE 2020a) has attached great importance to the improvement in teachers' and students' digital literacy by employing AI in learning and education to build high-quality education systems. As AI has been a core technology in education reform and innovation, it has received much attention and support from the Chinese government. During the past 5 years, the Chinese government has initiated several policies to push AI development in teacher training, student learning, and schooling. However, the ethical principles of AI in education at China's national level are not very clear, which needs to be constructed for further analysis of the policy documents of the future.

5.2 *AI in Learning and Education in the Finnish Context*

Finland is a member state in the EU and has been in active interaction with the working groups preparing documents for AI. In addition, according to the Ministry

of Economic Affairs and Employment (MEAE 2017), it has also been a forerunner in Europe, publishing its national AI strategy as the first in the EU, and updated in 2019 (MEAE 2019). Before that, Finland had several national digitization programs since the 1990s, promoting digital competences to all people and specific programs for schools and teacher education (e.g., Niemi et al. 2014). The most important value has been equity, which means supporting everyone in using their equal rights. This value is also a leading principle in AI strategies.

Ethical principles have been discussed at the EU level in several documents (e.g., European Commission 2020, 2021a, b; European Parliament 2020; HLEG 2019). In 2021, the Commission published a proposal for a regulation of the European parliament and that of the council for harmonizing rules on AI. This approach is applicable to the entire AI development, not specifically for education. The EU is established for economic and social purposes. Therefore, the recommendations for AI are mainly related to business, technology, and commerce. The EU sees AI as a powerful tool for innovation and productivity. However, AI is also seen for social development, covering a wide spectrum of efforts to promote inclusion, tolerance, justice, solidarity, and nondiscrimination based on the EU's fundamental values of democracy, human dignity and freedom, and human rights. Education and its regulations are beyond the EU mandate and are each nation's own responsibility. However, the EU can provide recommendations and guidelines for advancing educational actions for the well-being of society and citizens.

6 Ethical Guidelines for AI-Based Learning

In this section, we introduce the national-level policies related to AI-based learning within and beyond the education sectors during the past decade, both in China and Finland. We focus on the aforementioned themes: (1) inclusion and personalization, (2) justice and safety, (3) transparency and responsibility, and (4) autonomy and sustainability.

6.1 *Inclusion and Personalization*

When AI permeates learning and education, it first concerns equal access by all learners. Meanwhile, AI should supply learners with differentiated alternatives in education. In China and Finland, there are different contents and strategies for promoting the inclusion and personalization of AI in learning.

6.1.1 Chinese Context

In China, AI has been regarded as a tool to achieve personalized and differentiated learning through cultivating qualified teachers who have the ability to use AI in teaching. China's MOE (2018a) issued the "action plan for revitalization of teacher education," which encourages full use of cloud computing, big data, virtual reality, and AI to promote information-based teaching. Moreover, China's MOE (2020b) issued a policy about reforming teacher training programs in rural areas. By integrating 5G and AI into the teacher education curriculum, it optimizes pre-service teachers' capability of digitized teaching. These policies focus on integrating AI into teacher education and then suggest that the teacher is a primary agent who ensures AI ethics in education.

Considering the disparate development of the east and west regions of China, AI has become an effective technology to support disadvantaged and vulnerable groups in access to high-quality education resources. One example is building an intelligent schooling platform, where students' learning can be recorded and diagnosed accurately. In terms of personalized learning, China's MOE has planned future tasks to strengthen the construction of the platform for educational resources, especially to achieve balanced development of urban and rural schools with the help of AI (MOE 2018b).

6.1.2 Finnish Context

Inclusion is a leading principle throughout the Finnish educational system, and equity has been the highest priority over 40 years (e.g., Niemi et al. 2016). Equity and inclusion have also been a core in all national digitization programs since the 1990s. At present, AI provides new tools for supporting students' learning and keeping all students active in their learning paths. Personalization has already been included in the national core curricula 2014 for basic education, and now AI will provide new tools to pursue that aim. So far, the main ethical considerations in AI strategies have focused on equal opportunities for life-course learning.

Both in European and Finnish national AI strategy documents, AI's connection with education originates primarily from working life's perspectives and how work will change radically with AI. Lifelong learning and people's capacity to understand and use AI in their lives are central concerns. The focus is on people's capacity to use digital tools in their daily life: "all Europeans need digital skills to study, work, communicate, access online public services and find trustworthy information" (EC 2021b). The Finnish AI strategy (MEAE 2019) also states that, in lifelong learning, society should meet substantial continuing education needs. This aim requires reforms in the education system and the division of responsibilities arising from the updating of professional skills, and "AI and digitalization should be extensively incorporated into a broad range of different educational programs" (MEAE 2019, p. 13).

The Ministry of Education and Culture (MEC 2020) in Finland sees AI's connection to education as wider than only lifelong learning. The MEC published 2020 education policy foresight until 2040, which set the aims that "new methods introduced by science and AI can be utilized in many ways in the guidance and evaluation of the entire education and research system" (p. 83). In addition to system-level data, AI can help identify and eliminate learning problems. Inclusion can be achieved by evaluating how the system and educational services work and how to help individual students. AI can promote inclusion and support learners through personalization, but it should be integrated with other services, and the development must be based on contributions from many partners. However, the report also warns that "data utilization also has its inherent risks in the absence of clear ethical, legislative, and data management guidelines, in the development of which the public administration plays a major role" (MEC 2020, p. 83).

6.2 Justice and Safety

Justice and safety means that people can trust AI solutions and have the skills and procedures required to influence AI use and AI-based decisions. Meanwhile, in this process, personal data and our privacy will be protected. Both China and Finland have been aware of the challenges of AI in learning when it comes to algorithmic unfairness and risks.

6.2.1 Chinese Context

In 2016, China National Commission of Development and Reform (NCDR 2016) issued the "3-year action plan for Internet + AI," which proposed cultivating many global leading AI backbone enterprises in key fields, initially to build a solid foundation, active innovation, open cooperation, green and safe AI ecology, and a 100-billion scale of AI market application. In terms of education, AI is primarily regarded as an ought-to-be safe technology for all children.

In 2017, one document entitled "development plan of new generation AI" issued by the State Council, China, stated that AI has become a new engine of social development. On one hand, AI has brought new opportunities; on the other hand, its development has also brought some uncertainties and new challenges (State Council, China 2017). One of the major challenges is network security. China's MOE discussed this issue in 2019, stating that educators should be aware of the potential security risks in big data. Schools and teachers should strengthen forward-looking prevention and minimize the possible risks in some AI platforms and ensure the safety, controllability, and reliability of AI in education (MOE 2019).

6.2.2 Finnish Context

The EU has expressed very strongly that “AI systems must not undermine democratic processes, human deliberation, or democratic voting systems” nor “the foundational commitments upon which the rule of law is founded” (European Commission 2019, p. 11). The Finnish AI strategy reinforces that approach and claims that, in Finland and Europe, AI-related systems and use must respect the principles of Western democracy and freedom and AI should be seen as “a way of reinventing society and increasing citizens’ participation in decision-making and democratic processes” (pp. 38–39).

As an indicator of justice, the EU has enacted in 2018 (regulation already accepted in 2016), the General Data Protection Regulation (GDPR), which applies across the EU (EC 2018). The GDPR sets out principles for the lawful processing of personal data. Personal data are any information that is related to an identified or identifiable natural person. The GDPR’s primary aim is to enhance individuals’ control and rights over their personal data.

The purpose of the GDPR is to provide a set of standardized data protection laws across all member countries. It covers all phases of collecting, using, and storing data. This should make it easier for EU citizens to understand how their data are being used and raise any complaints. Finland is also committed to following this regulation. Schools and educational institutions must follow GDPR principles and have organizational and technical measures and policies in place to keep personal data safe and secure. GDPR sets high demands for data collection and restoration with AI-based applications. It must be applied, for example, with big data and learning analytics when being used for profiling students or when schools use other AI-based tools, such as massive online courses or intelligent tutoring systems that collect data from students. All data collections must be accurate and require permission from a person. A consent must be a specific, freely given, plainly worded, and unambiguous affirmation given by the data subject, and data subjects must be allowed to withdraw this consent at any time. Consent for children, defined in the regulation as being less than 16 years old, must be given by the child’s parent or custodian and should be verifiable. Schools should also ensure that external organizations from whom they have contracts (e.g., AI services) meet the GDPR requirements. The aim of the GDPR is to ensure the right for safety and privacy to citizens in the use and contexts of AI.

6.3 Transparency and Responsibility

Another two ethical questions that need to be addressed are how a decision is made and who is responsible or accountable in AI-based learning. These questions relate to issues of transparency and responsibility. Both educators and learners should be able to see and understand how the algorithmic process works and what possible results can be achieved.

6.3.1 Chinese Context

The State Council, China (2016) published the national plan for technology innovation in the 13th Five-Year Plan. To build a modern industrial technology system with international competitiveness, it is proposed to vigorously develop a new generation of information technology with ubiquitous integration, green broadband, security, and intelligence; develop a new generation of Internet technology; ensure the security of cyberspace; and promote the wide penetration and deep integration of information technology into various industries. In this ambition, accountability is about a clear acknowledgement and assumption of responsibility and answerability for actions, decisions, products, and policies in the national plan.

In the 2018 forum of AI standardization, the white paper of AI standardization was published, which makes AI technology more transparent within social audibility (AI Standardization Commission 2018). Recently, the China Ministry of Science and Technology (MST 2021) published the regulations of new generation AI ethics, which emphasized the data transparency and audible outcomes of AI technology.

In 2021, the China MOE actively explored the use of AI technology to enhance interactive communication, intelligent question answering, and personalized learning resources to push the functions of platforms at all levels. The China MOE encouraged K-12 schools to strengthen the collection and analysis of students' learning data and information through AI platforms (MOE 2021a). The prior action is to educate teachers and students to know the algorithmic process of AI platforms in schooling so that teachers can carry out concise guidance for students. It can be seen as the responsibility of AI in education when learners get enrolled in the technological environment. Educators also have the responsibility to create a learning environment with a sense of safety.

6.3.2 Finnish Context

The European Commission has opted for a human-centric approach, meaning that AI applications must comply with the fundamental rights of European citizens. At this moment, the focus in the AI debate in Finland and elsewhere in Europe is on ethical issues: "protection of privacy, accountability for the errors made by AI systems, and the traceability and transparency of algorithm-based decision-making" (EC 2020, p. 35).

In terms of transparency and responsibility, the GDPR provides a general framework and contains specific obligations and rights for the processing of personal data (e.g., the right not to be subjected to solely automated decision-making, except in certain situations). It also includes specific transparency requirements on the use of automated decision-making (e.g., to inform about the existence of such decisions) to provide meaningful information and explain its significance and the envisaged consequences of the processing for the individual (EC 2018, 2020).

The Finnish AI strategy (MEAE 2019, p. 106) critically observes that one ethical challenge is that AI is produced in ecosystems and ensuring compliance with ethical

practices can seldom be controlled by an individual organization. Services with the help of AI applications require complex global value chains. The report also claims that we need multidisciplinary discussion and research data to understand and interpret the broad societal impacts of AI. It also emphasizes that AI ethics must not be seen as a factor posing limitations on the activities only but also as a factor that creates something new and provides increasing opportunities (p. 106).

6.4 Autonomy and Sustainability

Although AI technologies are still a work in progress, it is not inconceivable that such AI machines, assuming other outward forms, will interact with humans holistically in the future. Autonomy and sustainability concern the ethical considerations of human–AI relations.

6.4.1 Chinese Context

Deep learning and machine learning are the core concepts for AI in educational data mining. Due to the significant progress in theory and practice derived from the application of AI in educational data mining and learning analytics, it is further argued that learners' autonomy in the AI-based learning environment has become increasingly important.

China MST (2021) issued the new ethical regulations of AI, which implies a prudent consideration of human–AI relations. In the learning and educational field, AI has increased the effectiveness of teaching and decreased teachers' workload (e.g., homework checking). However, it is another issue to preserve educators' and learners' autonomy, which means that some humans' abilities cannot be replaced by AI (e.g., socio-emotional literacy and agentive decision-making) (MOE 2021b).

In terms of the sustainability of human development using AI, the China MOE declared to promote the teaching contingent by AI, which explores new ways through AI to promote teacher management, educational innovation, and accurately helping the alleviation of poverty (MOE 2018b). Similarly, in April 2021, China MOE (2021c) proclaimed to make full use of the advantages of AI to cultivate highly qualified teachers with new pedagogical ideas. In China, the sustainable development of education starts with an excellent teacher contingent. If AI serves as a tool for constructing a high-quality education system, then it means that education would gain sustaining development.

6.4.2 Finnish Context

Autonomy “is a quality that can be attributed only to human beings. It is expressed in the human abilities to be self-aware, self-conscious, and a self-author, meaning

being able to set own rules and standards and choose own goals and purposes in life. Autonomy is a central aspect of human dignity and agency" (EP 2020, p. 12). There is ongoing discussion on the explicability of algorithms. Although deep learning has the so-called black boxes that are difficult to explain, human beings are still responsible for the decisions made by AI and consequences in society. Therefore, respect for human autonomy requires that there is meaningful human intervention and participation in AI and that AI systems are not to "subordinate, coerce, deceive, manipulate, condition, or herd humans" (HLEG 2019, p. 12; EP 2020, p. 52).

AI provides multichannel and multimodal data collection. Big data has the capability to combine data from different sources, to segment, and to profile students. In education, digital traces start from early childhood throughout the course of life. Finland has clear thought leadership in the development of the principles, operating models, information architectures, and technical solutions of a human-centered data economy (MEAE 2019, pp. 58–60). Here, the MyData approach also plays a key role. The model is derived from healthcare, and the most essential part of MyData is the consent for the use of patients' data and the safe transfer of information. In an international comparison, the Finnish MyData work is advanced because of the development of interoperability between operators and data ecosystems functioning in a fair manner. The European Commission has highlighted it as part of the preparation work for its data economy communication (MEAE 2019, p. 60). This type of MyData can also be useful in the education sector.

The Finnish strategy (MEAE 2019) makes critical remarks and indicates that the recent debate is very expert centric and claims that civil society should be allowed to participate in the discussion about AI ethics and its societal impacts to an increasing extent. AI-based solutions should be seen as a way of reinventing society and increasing citizens' participation in decision-making and democratic processes. For a sustainable use of AI, the Finnish strategies set as the future aim that everyone has sufficient understanding of AI and has this as a new civic skill. The strategies propose interdisciplinary, long-term research on the interaction of AI and society, supporting the autonomy of research and the critical voice (MEAE 2019). This sets new demands for the entire education system.

7 Discussion and Conclusions

Achieving the global benefits of AI requires transnational dialogues on many areas of governance and ethical standards while allowing for diverse cultural perspectives and priorities. This chapter is an endeavor to reimagine our learning and education, by exploring a global contract about AI ethics (UNESCO 2021). In the final section, we discuss the commonalities and differences in AI ethics between China and Finland. The results of our analyses of Chinese and Finnish policies of AI in learning and education can provide insights into constructing AI ethics in education internationally.

Table 1 Differences in AI ethics in education between China and Finland

		China	Finland
<i>Approaches</i>		Top-down initiatives Governments direct and guide	Bottom-up explorations The third parties and professional and expert organizations involved
<i>Properties</i>		Most of the policies are regulations AI ethics in education depend on other social sectors	Some policies have been legitimated AI ethics in education are unique but need interaction with other sectors and stakeholders
<i>Strategies</i>	<i>Inclusion and personalization</i>	Teacher education as a focus AI-based schooling has been explored	Lifelong learning as a focus Informal education sectors are included
	<i>Justice and safety</i>	Data safety and data protection are emphasized Forward-looking ability to prevent risks Personal data protection relates to social harmony	Human-centered approach to all AI-based activities Strong claims of personal data protection and privacy Human dignity and freedom, human rights, and democracy as AI grounds
	<i>Transparency and responsibility</i>	Make the algorithmic process public to learners Building audible AI platforms for learning outcomes	All stakeholders have responsibilities of transparency Specific requirements for the whole process of decision-making
	<i>Autonomy and sustainability</i>	Learners have irreplaceable subjectivity Teacher-driven reform is a vital step in AI's sustainable development	Autonomy is a quality that can be attributed only to human beings AI knowledge as a civic skill

In terms of the differences in expressing ethical principles in AI-based learning, Table 1 depicts the variations from three aspects (i.e., policy approaches, properties, and strategies), which can work as a theoretical framework for further comparative studies among other countries and regions.

Due to the differences between the sociocultural context and political regime, China mainly takes a top-down approach to initiate AI ethics. In contrast, Finland, as part of Europe, has more third parties and professional communities to contribute to the exploration of AI ethics in education. Interestingly, in Finland, some policies about AI ethics are legislations approved by local governments or the EU. Yet, Chinese policies about AI ethics in education depend much on other social sectors'

coordination. In terms of the specific strategies, Finland has a strong emphasis on the value basis of equity, nondiscrimination, human rights, and democracy and how all citizens can be made capable of using AI in their lives. Influenced by the Western value of citizenship, Finland emphasizes that all people should be able to understand the basics of AI and how it influences their lives and give their consent for the safe use of their personal data. China, as a communist country, pays much more attention to cross-sectional cooperation between education and other social sectors in developing AI and a shared ethical basis in a harmonious society.

Despite the differences, in this chapter, we call for a human-centered or humanist stance upon AI. Through our comparative research between China and Finland, the deployment of AI technologies in learning should be proposed to enhance human capacities and to protect human rights for effective human–machine collaboration in life, learning in and out of formal sectors, and lifelong sustainable development. In terms of inclusion, justice, and equity, the promise of “AI for all” (UNESCO 2020) must be that everyone can take advantage of the technological revolution underway and access its fruits, notably in terms of innovation and knowledge.

It is not difficult to conclude that both China and Finland assume that AI has the potential to address some of the challenges in education today, to innovate teaching and learning practices and, ultimately, to accelerate progress toward sustainable development goals (United Nations 2020). As two of the earliest nations proclaiming digital national programs, China and Finland are active partners in international discussions on the ethical principles of AI in society. All national strategies in the two countries for AI have a long-term view on social development. However, as noted in the beginning of this chapter, both countries hold the perspective of tending to industry and business for international competitiveness, not much specifically on education. This finding implies that educational sectors should not simply be the customers of AI technologies but should maintain relative independence or even lead the change in AI. This transnational study suggests that we should be aware of the uniqueness of learning and education compared to industry and business, which needs localized ethical guidelines for AI-based learning in the future.

Bridging these socio-technical gaps and the deep divide between the abstract value language and design requirements is essential to facilitate nuanced, context-dependent design choices that can support moral and social values (Aizenberg and van den Hoven 2020). In addition, we need more ethical reflections for education and learning at different levels: (1) society-level impacts and consequences on justice, equality, and inequality; (2) ethical guidelines for technological developers' and users' for making intelligent tools safe and explicable; and (3) ensure individuals' capacity and rights when using AI. By benefiting from AI technologies alongside ethical guidelines, we can reimagine a better future in which teaching and learning can shape the future of humanity and the world.

Acknowledgements The writing of this article is sponsored by Business Finland (Grant 7818/31/2018), University of Helsinki, and the Funding of Landmark Academic Achievement at the Faculty of Education, Capital Normal University (Grant No. 21530420006).

References

- Aizenberg, E., & Hoven, J. (2020). Designing for human rights in AI. *Big Data & Society*, 7(2), 1–14.
- AI Standardization Commission. (2018). *White paper of AI standardisation*. Retrieved on 01-05-2021. http://www.cesi.cn/201801/3545.html?ivk_sa=1024320u
- Almohammadi, A., Hagras, H., Alghazzawi, D., & Aldabbagh, G. (2017). A survey of artificial intelligence techniques employed for adaptive educational systems within e-learning platforms. *Journal of Artificial Intelligence and Soft Computer Research*, 7, 47–64.
- Baneres, D., Rodríguez-Gonzalez, M. E., & Serra, M. (2019). An early feedback prediction system for learners at-risk within a first-year higher education course. *IEEE Trans Learning Technology*, 12, 249–263.
- Baradwaj, B. K., & Pal, S. (2011). Mining educational data to analyze students' performance. *International Journal of Advancing Computer Science Application*, 2, 63–69.
- Biesta, G. (2017). *The rediscovery of teaching*. Routledge.
- Boddington, P. (2017). Towards a Code of Ethics for Artificial Intelligence. Springer.
- Bozkurt, A., Karadeniz, A., & Baneres, D. (2021). Artificial intelligence and reflections from educational landscape: A review of AI Studies in half a century. *Sustainability*, 13(2), 1–16.
- Coeckelbergh, M. (2020). Artificial intelligence, responsibility attribution, and a relational justification of explainability. *Science and Engineering Ethics*, 26, 2051–2068.
- Crawford, K. (2021). *Atlas of AI: Power, politics, and the planetary costs of artificial intelligence*. Yale University Press.
- EC. (2018) (CEPEJ). *European ethical charter on the use of Artificial Intelligence in judicial systems and their environment*. Strasbourg, European Commission for the Efficiency of Justice. <https://rm.coe.int/ethical-charter-en-for-publication-4-december-2018/16808f699c>
- EC. (2019). *Ethics guidelines for trustworthy AI*. <https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai>
- EC. (2020). *White paper. On artificial intelligence – A European approach to excellence and trust*. https://ec.europa.eu/info/publications/white-paper-artificial-intelligence-european-approach-excellence-and-trust_en
- EC. (2021a). Fostering a European approach to artificial intelligence. Brussels, 21.4.2021
- EC. (2021b). Digital skills and jobs. <https://digital-strategy.ec.europa.eu/policies/digital-skills-and-jobs>
- European Parliament (EP). (2020). *European framework on ethical aspects of Artificial Intelligence, robotics and related technologies*. European Added Value Assessment (EAVA). Retrieved from: [https://www.europarl.europa.eu/RegData/etudes/STUD/2020/654179/EPRS_STU\(2020\)654179_EN.pdf](https://www.europarl.europa.eu/RegData/etudes/STUD/2020/654179/EPRS_STU(2020)654179_EN.pdf)
- Flick, U. (2006). *An introduction to qualitative research*. Sage Publications.
- Floridi, L., Cowls, J., Beltrametti, M., Chatila, R., Chazerand, P., Dignum, V., Luetge, C., Madelin, R., Pagallo, U., Rossi, F., Schafer, B., Valcke, P., & Vayena E. (2018). AI for people—An ethical framework for a good AI society: Opportunities, risks, principles, and recommendations. *Minds and Machines*, 28, 689–707.
- Goksel, N., & Bozkurt, A. (2019). Artificial intelligence in education: Current insights and future perspectives. In Sisman-Ugur, S., & Kurubacak, G., (Eds.). *Handbook of research on learning in the age of transhumanism*(pp. 224–236). IGI Global.
- High-Level Expert Group, EU (HELG). (2019). *Ethics guidelines for trustworthy AI*. Retrieved on 23-05-2020. <https://ec.europa.eu/digital-single-market/en/news/ethics-guidelines-trustworthy-ai>.
- Kay, J. (2012). AI and education: Grand challenges. *IEEE Intelligent System*, 27, 66–69.
- MEAE. (2017). *Turning Finland into a leading country in the application of artificial intelligence objective and recommendations for measures*. Retrieved from: https://julkaisut.valtioneuvosto.fi/bitstream/handle/10024/160391/TEMrap_47_2017_verkkojulkaisu.pdf?sequence=1&isAllowed=true

- Ministry of Economic Affairs and Employment, Finland (MEAE). (2019). *Leading the way into the age of artificial intelligence: Final report of Finland's Artificial Intelligence Programme 2019*. Retrieved from: <http://urn.fi/URN:ISBN:978-952-327-437-2>
- Ministry of Education and Culture, Finland (MEC). (2020). *Education Policy Report of the Finnish Government, Finland*. <http://urn.fi/URN:ISBN:978-952-383-927-4>
- Ministry of Education, China (MOE). (2018b). *Notice on carrying out the pilot work of artificial intelligence to boost the construction of teachers*. Retrieved on 01-05-2021. http://www.moe.gov.cn/srcsite/A10/s7034/201808/20180815_345323.html
- Ministry of Education, China (MOE). (2018a). *Action plan for revitalization of teacher education*. Retrieved on 01-05-2021. http://www.moe.gov.cn/srcsite/A10/s7034/201803/t20180323_331063.html
- MOE. (2019). *Key points of education informatization and network security in 2019*. Retrieved on 01-05-2021. http://www.moe.gov.cn/srcsite/A16/s3342/201903/t20190312_373147.html
- MOE. (2020b). *Opinions on strengthening the construction of rural teachers in the new era*. Retrieved on 01-05-2021. http://www.moe.gov.cn/srcsite/A10/s3735/202009/t20200903_484941.html
- MOE. (2020a). *The proposal of steadily promoting the integration of programming education into China's basic education system and focusing on the cultivation of digital talents*. Retrieved on 01-05-2021. Retrieved from: http://www.moe.gov.cn/jyb_xxgk/xxgk_jyta/jyta_jiaocaiju/202012/20201209_504364.html
- MOE. (2021c). *Notice on the recommendation and selection of the second batch of pilot projects of artificial intelligence boosting teacher team construction*. Retrieved on 01-05-2021. http://www.moe.gov.cn/srcsite/A10/s7034/202104/t20210423_527853.html
- MOE. (2021b). *Opinions on strengthening the construction and application of online teaching resources in primary and secondary schools*. Retrieved on 01-05-2021. http://www.moe.gov.cn/srcsite/A06/s3325/202102/t20210207_512888.html
- MOE. (2021a). *Notice on the implementation of the national training plan for kindergarten teachers in primary and secondary schools (2021-2025)*. Retrieved on 01-05-2021. http://www.moe.gov.cn/srcsite/A10/s7034/202105/t20210519_532221.html
- Mouta, A., Torrecilla, E., & Pinto, A. (2019). Blending machines, learning, and ethics. In *Proceedings of the Seventh International Conference on Technological Ecosystems for Enhancing Multiculturality* (TEEM 2019). León, Spain, 16–18 October 2019; ACM.
- Ministry of Science and Technology, China (MST). (2019). *Principles of AI Governance for the next generation*. Retrieved on 01-05-2021. http://www.most.gov.cn/kjbgz/201906/t20190617_147107.htm
- MST, China (2021). Regulations of new generation AI ethics. Retrieved on 30-09-2021. https://www.most.gov.cn/kjbgz/202109/t20210926_177063.html
- National Commission of Development and Reform, China (NCDR). (2016). *Three-year action plan for Internet + AI*. Retrieved on 01-05-2021. https://wap.miit.gov.cn/xwdt/szyw/art/2020/art_bf006f97f8b5433cb6f5d89e70991aae.html
- Niemi, H. (2020). Artificial intelligence for the common good in educational ecosystems. In *Humanistic Futures of Learning: Perspectives from UNESCO Chairs and UNITWIN Networks* (pp. 148–152). Ediciones UNESCO. Retrieved from <https://researchportal.helsinki.fi/en/publications/artificial-intelligence-for-the-common-good-in-educational-ecosystems>
- Niemi, H. (2021). AI in learning: Preparing grounds for future learning. *Journal of Pacific Rim Psychology*, 15, 1–12. <https://doi.org/10.1177/18344909211038105>
- Niemi, H., Multisilta, J., Lipponen, L., & Vivitsou, M. (Eds.) (2014). *Finnish innovations and technologies in schools. A guide towards new ecosystems of learning*. Brill-Sense.
- Niemi, H., Toom, A., & Kallioniemi, A. (2016). *Miracle of Education: The Principles and Practices of Teaching and Learning in Finnish Schools*. (Second Revised Edition) Sense publishers.
- Nye, B. D. (2016). The end of the world as we know it: Transitioning AIED into a service-oriented ecosystem. *International Journal of Artificial Intelligence in Education*, 26, 756–770.
- Rawls, J. (1999). *A theory of justice: Revised edition*. Harvard University Press.

- Roll, I., & Wylie, R. (2016). Evolution and revolution in artificial intelligence in education. *International Journal of Artificial Intelligence in Education*, 26(2), 582–599.
- State Council, China (2016). The 13th five-year national plan for technology innovation. Retrieved on 01-05-2021. http://www.gov.cn/zhengce/content/2016-08/08/content_5098072.htm
- State Council, China (2017). *Development plan of new generation artificial intelligence*. Retrieved on 01-05-2021. http://www.gov.cn/zhengce/content/2017-07/20/content_5211996.htm
- State Council, China. (2021). *National schedule for the 14th five-year plan*. Retrieved on 01-05-2021. Retrieved from: http://www.gov.cn/xinwen/2021-03/13/content_5592681.htm
- Tao, D. (2021). *Trustworthy AI*. Speech at the WAIC 2021. Shanghai, China.
- Tegmark, M. (2017). *Life 3.0: Being human in the age of artificial intelligence*. New York: First Vintage Books.
- United Nations. (2020). *Take action for the sustainable development goals*. Retrieved on 10-01-2021. <https://www.un.org/sustainabledevelopment/sustainable-development-goals>
- UNESCO. (2019). *Beijing consensus on AI and education*. Retrieved on 01-05-2021. <https://unesdoc.unesco.org/ark:/48223/pf0000376709>
- UNESCO. (2020). *AI and education: Guidance for policymakers*. Retrieved on 01-05-2021. <https://unesdoc.unesco.org/ark:/48223/pf0000368303>
- UNESCO. (2021). *Reimagining our future together: A new social contract for education*. Retrieved on 29-11-2021. <http://www.unesco.org/open-access/terms-use-ccbysa-en>
- Vinuesa, R., Azizpour, H., Leite, I., Balaam, M., Dignum, V., Domisch, S., & Nerini, F. F. (2020). The role of artificial intelligence in achieving the sustainable development goals. *Nature and Communication*, 11, 1–10.
- Whittaker, M., Crawford, K., Dobbe, R., Fried, G., Kaziunas, E., Mathur, V., West, S.M., Richardson, R., Schultz, J., Schwartz, O. (2018). AI Now Report 2018. Retrieved on 01-05-2021. https://ainowinstitute.org/AI_Now_2018_Report.pdf
- World Economic Forum. (2021). *Global technology governance report 2021: Harnessing fourth industrial revolution technologies in a Covid-19 world*. <http://www.weforum.org/reports/global-technology-governance-report-2021>

Open Access This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.



Artificial Intelligence Ethics from the Perspective of Educational Technology Companies and Schools



Päivi Kousa and Hannele Niemi

Contents

1 Introduction	283
2 AI in Education and Learning	284
3 Many Ethical Guidelines and Principles for AI	285
4 Case I: Finnish EdTech Companies' Views on Ethical Challenges	287
5 Case II: Finnish School's Ethical Challenges and Practical Viewpoints on Explicability	289
6 Discussion	291
6.1 Beneficence	292
6.2 Non-maleficence and Justice	292
6.3 Autonomy	293
6.4 Explicability	293
7 Conclusions	294
References	294

1 Introduction

In this chapter, AI will be presented in contemporary educational contexts. The aim is to understand what kind of ethical challenges EdTech companies and schools have and how those challenges affect their daily work. As technology evolves at an accelerating pace and the education sector seeks to keep up, rapid actions are needed to avoid the ever-growing gap between EdTech companies and schools. First, companies' and schools' reflections during interviews are presented inductively with their own concepts based on two Finnish case studies. Thereafter their thoughts are contextualised in terms of five ethical principles by Morley et al. (2020).

Artificial intelligence (AI) has become part of the global discussion and our everyday lives more than ever, although AI and machine learning have been among us for decades (Turing 1950/2009). AI is influencing almost all levels of our economy and society. For example, it enables people to use new tools and

P. Kousa (✉) · H. Niemi

Faculty of Educational Sciences, University of Helsinki, Helsinki, Finland

e-mail: paiivi.m.kousa@jyu.fi; hannele.niemi@helsinki.fi

applications, e.g. transportation, services, healthcare, education, public safety and security, employment and workplace, and entertainment (Stone et al. 2016; Littman et al. 2021). All these changes have fundamental influences on organisations which establish new demands which then need to be fulfilled by their staff developing new competences. New technology and advanced methods in computing with AI applications are increasingly used also in education. Globally, there are several common AI-related practices and tools for education and learning, such as teaching robots, intelligent tutoring systems (ITS), online learning, and learning analytics. Augmented and virtual realities are interactive systems often used for competence training, especially in many areas of life-long learning (e.g. Grover and Pea, 2018).

Although there is a global consensus that AI should be ethical, many problems exist in defining the values embodied in ethical guidelines. Ethical guidelines are not conceptually congruent but are rather open to a wide range of interpretations (e.g. Jobin et al. 2019). Many companies find general guidelines useless and prepare guidelines of their own instead (Hagendorff 2020). Cath (2018) suggests that universities and other organisations (e.g. policymakers and schools) could offer a leading, research-based, and objective role in the development of ethical guidelines since industry-produced guidelines may be too subjective. There are also fears that companies are too involved in drafting legislation and guidelines which serve to pursue their own interests (Cath 2018). However, cooperation is needed since so many parties are involved in the AI ecosystem, such as policymakers, universities, schools, and industries. Yet discussions between developers and researchers have lasted decades without sufficient outcomes (Bostrom and Yudkowsky 2011). Secondly, products and services based on AI are difficult or, in many cases, almost impossible to explain (Goebel et al. 2018), although their explainability and interpretability would enhance the fairness, transparency, and accountability needed for those who use AI products and services (Cath 2018). Thirdly, schools need education and guidelines which can be implemented during their daily work. Nnaji (2019) discusses how ethical conflicts in schools have more to do with how the technology is used than in the technology itself. He states that different applications are simply tools to help students and teachers in their work but should not be blindly trusted or allowed to guide school activities without critical considerations. AI in education presents serious challenges in relation to the issues of student privacy, accuracy, data ownership, accessibility, and integrity which need to be addressed (Nnaji 2019).

2 AI in Education and Learning

The increased use of AI for education and learning has promoted many opportunities as well as major challenges (Torresen 2018). According to United Nations Educational, Scientific, and Cultural Organization, UNESCO (2019), there are six major challenges related to AI in education (AIED): (1) lack of comprehensive public policy on AI, (2) unequal opportunities to use AIED, (3) lack of adequate teacher education, (4) lack of development of quality and inclusive data systems,

(5) lack of significant AI-related research, and (6) lack of ethics and transparency in data collection, use, and dissemination. Concerns with data privacy and ownership issues, and the safety of public/private interfaces, have raised questions especially in educational fields (e.g. Dignum 2018). Many researchers and international organisations claim that AI should be trustworthy—lawful, ethical, and socially as well as technically robust (High-Level Expert Group on Artificial Intelligence, AI HLEG 2019a). In education and learning, ethical challenges have grown in tandem with technological development, as AI trustworthiness has become increasingly important (e.g. Stanford Institute for Human-Centered AI, HAI 2020). Although AI has many benefits for learning, the educational field has faced many challenges in relation to equity, data management, decision-making, and human and machine learning (e.g. Stone et al. 2016). When AI is implemented in educational contexts, education stakeholders must be able to trust that the entire design processes of AI-based solutions are ethical and that the algorithms are designed in accordance with ethical principles that suit the values of the school world.

Yet Holmes et al. (2021) emphasise that ethics is not a straightforward concept in the context of education. They urge distinguishing between ‘doing ethical things’ and ‘doing things in an ethical way’. They suggest that AIED technologies should include specific ‘ingredient lists’ like in food or medicine products. This proclamation in labelling would increase the understandability and transparency of the AI-based solutions. In practice, this could mean that the user (e.g. a teacher or a student) would be informed of the limitations or benefits of the product beforehand. Goebel et al. (2018) remind us that efforts have been made to explain complex AI systems for decades. It can be concluded that many ethical challenges are present when designing AI-based tools and services for education. In addition, ethical factors are always present in education product design (e.g. in schools and workplaces), since the purpose is to exert influences on peoples’ minds, behaviours, and lives. This pervasive influence of education makes educational AI solutions even more challenging to develop. Although AI can provide many beneficial solutions to existing educational challenges, there are many new problems that need to be solved between EdTech companies and schools who use the solutions that are developed. The recent coronavirus disease 2019 (COVID-19) increased distance learning and thus the urgent need for teachers and students to use digital applications and understand how they work (e.g. Niemi and Kousa 2020).

3 Many Ethical Guidelines and Principles for AI

Numerous international, national, governmental, organisational, and company-based guidelines exist for ethical AI. For example, the European Commission’s high-level group on artificial intelligence (AI HLEG) has published four deliverables: ethics guidelines for trustworthy AI with 7 key requirements (AI HLEG 2019a), policy and investment recommendations for trustworthy AI with 33 recommendations (AI HLEG 2019b), assessment list for trustworthy AI which can be

used as a practical aid when implementing requirement into practice (AI HLEG 2020a), and sectoral considerations on the policy and investment recommendations which provide examples concerning how and where regulations can be implemented (AI HLEG 2020b). The guidelines are developed in collaboration with an AI alliance including 4000 stakeholders (e.g. European Union/EU citizens, people from business and industry fields, universities, municipalities, and civil society). Different countries have their own national strategies. For example, Finland published its first AI strategy in 2017 (Ministry of Economic Affairs and Employment in Finland, MEAE 2017) and has provided updates (MEAE 2019). The main goal of the guidelines is to benefit from the opportunities brought by AI in all areas of society but in such a way that ethical aspects are considered and possible risks avoided. The Organisation for Economic Co-operation and Development (OECD) *Recommendation of the Council on Artificial Intelligence* (OECD 2019) has listed more than 70 documents published in the last 3 years which make recommendations about the ethics principles for AI (Spielkamp et al. 2019; Winfield 2019).

It is noteworthy that most of the guidelines developed by companies and other organisations focus on what ethical challenges exist, rather than what actions should be taken to achieve the ethical goals in practice (Cath 2018; Morley et al. 2020). It has been argued that the developers are often aware of the ethical issues, but companies do not provide appropriate tools or support to suitably tackle these issues (Abdul et al. 2018). Ethical guidelines for education as a context of AI application are mainly lacking (Holmes et al. 2021), although the need has been recognised decades ago (Aiken and Epstein 2000). Nonetheless, educational issues are included in general policy-level guidelines (e.g. AI HLEG 2019a). Jobin et al. (2019) analysed 84 regulation documents or guidelines for ethical use of AI, and according to their review, the most important principles are transparency (including explainability and understandability), justice and fairness, non-maleficence, responsibility, and privacy. In addition, Hagendorff (2020) has presented ethical criteria such as accountability, explainability, discrimination-aware data mining, tools for bias mitigation, and fairness in machine learning. Moreover, AI actions should also be predictable and the systems that are based on AI should be robust against manipulation. Clear human accountability for AI actions must also be ensured (Bostrom and Yudkowsky 2011).

According to a literature review by Morley et al. (2020), the five main principles are beneficence, non-maleficence, autonomy, justice, and explicability, which are not only complementary but also partly overlapping. Morley et al. (2020) have combined this typology from the EU's report that lays grounds for trustworthiness (AI HLEG 2019a). The five principles can be summarised as follows:

- *Beneficence* means that the AI-based system is useful and reliable and supports diversity, human well-being, and development. Product development should be justified in alignment with beneficence and not created solely for the sake of the product but for benefiting the user.
- *Non-maleficence* includes many aspects that are related to human and data safety. It is important to be prepared in advance for the possible security threats. Data

security, accuracy, reliability, reproducibility, quality, and integrity must each be guaranteed at all stages of the product's life cycle.

- Autonomy means the freedom to make decisions and choices regarding AI-based systems. Tools that support individual autonomy should be designed and implemented.
- Justice requires that the AI systems operate in a fair manner without obstructing democracy or harming society. The negative effects of AI systems should be minimised. All stages of product development processes should be made more transparent, and they should be able to be evaluated and documented.
- Explicability means that AI systems should be able to be understood and their operations explained and interpreted. This does not mean that everything should be explained, as it is impossible. The level at which AI is explained depends on need and can range from a very simple explanation to a more complex one. For example, basic knowledge could be taught and assessed in schools in ways appropriate to different age levels. This would ensure sufficient civic skills to make ethically sustainable decisions, for example, regarding one's own security. Accountability and responsibility should be clear, transparent, and traceable.

The typology introduced by Morley et al. (2020) shows that many of the ethical principles are very interrelated. Explicability can be seen as both an independent and a unifying factor. In many cases, it is unclear what needs to be explained concerning AI and its applications and how the decision is made (Coeckelbergh 2020) and who makes the decisions (Floridi et al. 2018). Additionally, it is not always clear who should take responsibility if something goes wrong or if AI is to blame in those occasions. In the next section, representatives of EdTech companies and schools will reflect on what their major concerns from an ethical viewpoint are when AI is applied in educational settings.

4 Case I: Finnish EdTech Companies' Views on Ethical Challenges

Seven EdTech company representatives who work in Finland were interviewed in the qualitative study of Kousa and Niemi (2021). The aim was to look for new ideas and solutions on how AI could be utilised in an ethically sustainable way in education. Companies in this study provide AI-based EdTech products and services such as well-being surveys and solutions for schools, tutoring services using VR and AR technology, ethical and safe data management solutions, and game- and simulation-based applications in oil operator training. All companies have extensive international business and more than 10 years of experience in the EdTech field. According to the findings, EdTech companies have faced ethical challenges in their work.

First of all, companies struggle with regulations and guidelines which have been found difficult to understand and implement. Therefore, making their own

guidelines is mostly preferred. The situation is even more complicated in the international marketplace for educational technologies, since other countries are likely to have different cultures, guidelines, and understandings for what is meant by ethical AI in the first place. Additionally, conducting business with schools is challenging as schools' resources, opportunities, and willingness to use AI-assisted solutions vary widely. Negative attitudes or even unrealistic expectations of AI were also seen as problematic. The situation is contradictory when, on the one hand, information is freely provided, for example, on social media, but, on the other hand, there are many kinds of fears. For example, AI solutions are not necessarily trusted in the workplaces or schools, or workers might be afraid that machines will replace them in the future. It was also argued that the bad reputation and negative attitudes of AI is caused by the critical tone with which large companies such as Microsoft, Google, Apple, Facebook, and Amazon have been talked about in the media.

When EdTech companies were asked how they could increase ethical sustainability in their AI solutions, the following issues were raised:

1. Companies argued that there is a serious lack of civic knowledge about what AI is and what it can and cannot do. Therefore, more public education about AI in education is needed. For example, AI is often misunderstood to mean just learning coding in schools. Therefore, AI as well as related ethical issues should be taught at all grade levels. Teachers also need more education on the topic.
2. Schools and workplaces should have equal opportunities to choose AI-based teaching materials and methods that are accessible and understandable. However, customised versions of one-size-fits-all services that are increasingly needed are expensive and often impossible to implement. Therefore, the best solutions could be based on collaboration between the teacher and the AI application, such as an AI tutor.
3. Responsibility issues should be defined and more transparent. It should be clearer to the user when the company is responsible and what responsibilities belong to the user. However, it is typically unclear when a machine or human is responsible when AI is applied in education.
4. One of the companies' main concerns was how to make safe and ethically sustainable products for schools. Problems were seen in data collection, transfer, storage, and modification. Prevention of harm is one of the most important solutions to make more ethically sustainable solutions. That means constant risk analysis and ethical checklists.
5. Universities, companies, and schools should share more of their knowledge and best practices to each other and to the public for a common good about AI in education.
6. Public events about the possibilities, risks, and threats of AI should be held regularly to avoid development of false assumptions about AI and formation of only negative attitudes.

When EdTech companies were asked about the need for support, several issues surfaced. Companies need more understanding about legislation, ethical risks, algorithms, and responsibility issues. They hope that there would be multi-professional

partners such as legal experts, universities, schools, other companies, or decision-makers who could be asked for support and advice on difficult ethical issues. They also wanted to share responsibilities between different stakeholders.

5 Case II: Finnish School's Ethical Challenges and Practical Viewpoints on Explicability

Twenty school principals and/or teachers who work as digital tutors in Finnish schools participated in a qualitative interview study in 2021. The participants were asked about their views on AI, digital applications, and ethical challenges.

As for what constitutes the main challenges related to AI in education, many respondents felt that teachers do not know enough about AI or related applications. According to interviewees, there are usually only a few more dedicated teachers in schools who act as digital educators/tutors. One of the school principals stated that teachers are not motivated to adopt AI tools if there is no guarantee that they will be useful in teaching. In smaller schools, the acquisition and responsibility for digital equipment was generally the responsibility of the principal. AI was seen as a good tool for easy routine tasks and for providing differentiated instructions when needed. However, all teachers did not see AI or digital applications alone as sufficient to guarantee better teaching or learning. One teacher described the scenario as follows: 'So the AI would say to the teacher that Matt is a bit stressed now so you should leave him alone (laughter)? I have to say that I can't imagine what kind of help AI could provide that a teacher cannot. Even though it is AI, someone has coded it. There should also be some kind of control that AI gives the right information before we start doing things based on it'. When asked what kind of additional information teachers would need about AI, one replied: 'We should find out what AI means in practice. If we have an application that collects information about stress, then we need to know well enough about its operating principles and purposes. To see the big picture. And what to think about AI in education'. According to teachers, AI-based applications should be developed in collaboration with schools, companies, and researchers and should be tested long enough before use. One of the future scenarios which teachers are afraid of is that when the use of AI-based solutions increases, their control in the classrooms will diminish. They are concerned that companies are starting to define more and more about what is taught and how. This in turn might reduce objectivity as one of the teachers explained: 'I hope that we would get better AI tools for teaching. This means that our city, which decides what tools are allowed, has to reduce strict restrictions, and make more new contracts with different software houses. Then there is a fear that it will go so that there will be those lobbyists of big companies such as Microsoft, which will forge them. I think our city has a fear that schools would be in breach of EU regulations if they were allowed to decide for themselves which AI tool to use.' In another example, the teacher expressed the concern: 'If AI begins to define what individual students do

in the lesson based on their personal learning profiles, the situation is not controlled by the student or a teacher or the parents, but by some other parties.' Furthermore, teachers did not believe that even the smartest system could replace teachers or make equally good predictions about how to work with diverse students or make decisions for the benefit of students. 'When thinking about an entire school day, it will always be influenced by a terrible number of elements that are related to only one situation. Predicting them and drawing more long-term conclusions would seem to be quite difficult, at least for the time being. For example, we know that in the fall, when it rains and is dark, disruptive behaviour easily increases. In this case, classroom lightning and human factors such as teacher's situational awareness are of great importance. If interpretations, conclusions, and measures come through AI, we will go to the so-called schematic side. That's when we're lost the human side of teaching.'

School representatives also felt that they have unequal opportunities to use AI-based solutions. The situation differs enormously even within a city. Some of the interviewees argued that there are schools that do not even have proper Internet connections and there are teachers and parents who are against digital education since they are afraid of, for example, issues related to privacy or even health.

Information security was an important issue in interview discussion, but there were differences in teachers' opinions on this topic. Some were not worried about sharing information, and others were very precise and also knew about the importance of privacy issues. However, security was seen as a challenge that companies and/or the city needs to address and cannot be an individual teacher's responsibility. New applications and unknown, especially foreign, companies were seen as less trustworthy. Indeed, many believed that larger companies had taken better care of information security. To improve safety, it was proposed that data taken from students should be stored only for a short time and then safely disposed of. Other options they proposed were that students' data should be anonymised and stored in an encrypted form in a secure location so that no one could recognise the student from the data. When asked about the future scenarios, one of the interviewees summarised: 'After all, the school is not out of the community. And AI comes into society on a global scale, whatever was said or done in schools. However, the school should not be the first place to use AI for the industry or business purposes, but vice versa. Schools need to keep up with the development of technology on their own terms. It is challenging because the changes are happening at an increasingly hectic pace. In schools, we need to remember that we are dealing with children or young adults. It seems like we have forgotten the stages of Piaget's cognitive development and so on. It seems that sometimes children are being expected too much these days'.

In order to facilitate the situation in schools where digital skills are becoming more important and a wide range of programs are used and provided by EdTech companies, the interviewees stated that:

1. It has to be explained what kind of data the system collects and what it is used for. In addition, teachers want to know where the data is stored and who owns

- it. This increases the credibility of the system in question and even influences its purchasing decision.
2. Teachers prefer easy-to-use and effective teaching aids. Some of the teachers even argued that privacy or explicability are not as important as usability.
 3. It is difficult to teach with digital applications without understanding how they work. This was noticed especially during the COVID-19 pandemic in distance teaching. Digital applications need to add value to teaching. Teachers have found that there are too many applications where traditional teaching has been digitised without major changes. These applications can make learning and understanding even more difficult.
 4. Many schools have IT support which is usually one person who works at school or in a larger area that includes many schools. According to teachers, sufficient help is mostly difficult to get. One of the problems is that IT supporters rarely know about the pedagogical aspects of the system and therefore cannot explain how it works or what kind of features there are that could help teaching or learning. Therefore, EdTech companies' support was seen as essential.
 5. Although attitudes towards AI are generally positive, most of the teachers have noticed that they have insufficient basic knowledge about it. For example, many teachers did not see any differences between coding and understanding basic things about AI to achieve adequate civic skills for the future.
 6. Teachers interviewed have no resources or desire to learn about AI or to learn numerous new AI systems, although ICT is expected to be used at all stages of education in Finland. Most of the respondents hoped that AIED would be used in teaching only by voluntary teachers who were interested in it.
 7. Decision-makers at schools prefer digital solutions that are as accessible and understandable to as many teachers as possible.
 8. Teachers do not want to be responsible for the privacy issues or functionality of the digital/AI-based solutions but assume that either the companies or the municipality are responsible. According to teachers, liability issues concerning EdTech products should be regulated by law.
 9. Teachers hoped that someone, such as a municipality, would review all EdTech companies and their products and services and ensure that they are ethical and safe to use.

6 Discussion

Ethical issues are strongly present in the daily lives of both schools and companies. These two cases represent a small sample of the situation in Finland, where the technological skills and know-how are at an internationally high level. However, more information is needed on the ethical issues involved and how the gap between businesses and schools could be reduced, *inter alia*, to improve trustworthiness. EdTech companies' and schools' challenges are discussed in the light of five ethical principles (beneficence, non-maleficence, autonomy, justice, and explicability) by Morley et al. (2020).

6.1 Beneficence

In Morley et al's typology, beneficence means that AI brings something positive to users and community and that AI is not a purpose in itself. Teachers strongly emphasised that the use of AI programs would not be an absolute value but would be based on a genuine need, for example, for differentiated instruction or assisting with routine tasks. Since teachers have a constant shortage of time and money, beneficence is an extremely important factor in choosing the right tools for teaching and learning. Companies also see the importance of providing accessible systems that take diversity into account, but they are worried that providing customised versions of one-size-fits-all solutions is challenging. Morley et al. (2020) see that justification belongs to beneficence. The purpose for building the system must be clear and linked to a clear benefit—systems should not be built simply for the sake of AI application or profit only.

6.2 Non-maleficence and Justice

Non-maleficence and justice are very much interrelated in the conceptualization of Morley et al. (2020): it means that AI systems should be protected against vulnerabilities that can allow them to be exploited by adversaries. AI systems should have safeguards that enable a fallback plan in case of problems. AI systems should guarantee privacy and data protection throughout a system's entire life cycle. Justice requires minimising and responding to potential negative impacts of AI systems. Companies on this study want to avoid ethical risks and emphasise that they do not intentionally make AI solutions that would be harmful to an individual or society. However, they need proper guidance, information, and legislation to support their product development processes. On the other hand, schools also need proper guidance on how to safely use digital/AI-based solutions. Recent research of Felderer and Ramler (2021) brings up the importance of quality assurance of AI-based systems. It has been recognised by AI solution developers that the models of machine learning or deep learning are not transparent, intuitive, or understandable. In Europe, General Data Protection Regulation (GDPR) has been developed to understand data management processes and civil rights, for example, how to protect users' personal data (EC 2018). However, identifying the factors that make AI non-maleficent requires considerable understanding of the entire system, from both developers and users. According to EC (2021) people should have basic digital skills and knowledge of AI and the ability to access and use the solutions in their daily lives.

According to one expert group (AI HLEG 2019a), accountability includes 'auditability, minimization and reporting of negative impact, trade-offs and redress' (p. 14). It is related to fairness and responsibility which are extremely necessary in every step of the production development process, both before and after. In this

study, companies emphasised that systems should be preventive of and minimise the risks. Companies complained about the difficulty of legislation and preferred their own guidelines and checklists. Hagendorff (2020) argues that ethical guidelines might not have a sufficient impact on companies' decision-making. They can be interpreted in many different ways because concepts are not clear. It is also easy to slip up on adherence to ethical principles, since there will be no consequences, surfacing policy concerns.

6.3 Autonomy

Autonomy means human agency and human oversight in a typology of Morley et al. (2020). This means that even though machines can intelligently analyse data and make conclusions, human beings are still responsible for the system and its consequences. Teachers in this study admitted that they do not want to be responsible for the privacy issues or functionality of the digital/AI-based solutions. They do not have the capacity to accomplish that. They assumed that either companies or the municipality should be responsible. The situation was twofold in these cases. Companies, on the other hand, understood their responsibilities but also wanted to share them among different stakeholders.

6.4 Explicability

Morley et al. (2020) set us an aim that AI systems should be built in such a way that they are understandable to users. Companies in this study needed more education and knowledge sharing to increase public trustworthiness in AI and its applications. Schools also needed information on both AI and its applications. Coeckelbergh (2020) points out that without explainability and transparency, responsible use of AI technology is problematic. To act in an ethically responsible way means knowing what is being done and being able to explain the system's actions and decisions in a way that others can understand. In addition, it is important to know to whom one is responsible for the creation of AI systems. The issue is complex, because people's need for explanations varies. Most people don't necessarily know that AI is involved in their applications in the first place or what AI does in that application. Even the best software developer may not know all the codes or know how to explain them (Coeckelbergh 2020). It can be concluded that explainability is a very human, content-, and context-dependent issue and, therefore, while extremely complicated, necessary.

7 Conclusions

This chapter has discussed the ethical challenges of EdTech companies and schools. Although EdTech companies and schools share some challenges, it can be said that the gap between companies and schools is in danger of widening as technological development advances. This observation also applies to other parties in the society, including researchers, decision-makers, and legislators. First, in the absence of sufficient legislation in the AIED field (Aiken and Epstein 2000; Holmes et al. 2021), ways should be urgently found for how to develop globally consistent regulations and guidelines, which include practical examples in a sufficiently understandable way to meet educational needs. This topic requires further research and consultation with both parties, as well as legitimate solutions based on consensus. Secondly, it must be recognised that explicability is a broad concept with many levels and needs (e.g. decision-makers, developers, and users) including what needs to be explained and how. In addition to understanding the technical details of individual applications and ‘black boxes,’ more knowledge is needed concerning how to explain AI in general and in the context of everyday life implementations. As stated earlier, it is not necessary to explain everything (Coeckelbergh 2020), but it is, for example, necessary to obtain the necessary civic knowledge and skills to participate in society. That could mean, for example, specifying what added value AI brings to the application used by the teacher or how. In conclusion, a huge amount of work has been done by researchers, companies, policymakers, and schools to increase a common understanding of AI. However, we are still on our journey to a more ethically sustainable future.

Funding Details This work was supported by [Business Finland](#).

Disclosure Statement No potential conflict of interest was reported by the authors.

References

- Abdul, A., Vermeulen, J., Wang, D., Lim, B. Y., & Kankanhalli, M. (2018). Trends and trajectories for explainable, accountable, and intelligible systems: An HCI research agenda. In Proceedings of the 2018 CHI conference on human factors in computing systems—CHI’18 (pp. 1–18). <https://doi.org/10.1145/3173574.3174156>.
- Aiken, R., & Epstein, R. (2000). Ethical Guidelines for AI in Education: Starting a Conversation. *International Journal of Artificial Intelligence in Education*, 11, 163–176.
- Bostrom, N., & Yudkowsky, E. (2011). The ethics of artificial intelligence. In K. Frankish & W. Ramsey (Eds.), *The Cambridge Handbook of Artificial Intelligence* (pp. 316–334). Cambridge: Cambridge University Press. <https://doi.org/10.1017/CBO9781139046855.020>.
- Cath, C. (2018). Governing artificial intelligence: ethical, legal and technical opportunities and challenges. *Phil. Trans. R. Soc. A376*: 20180080. <http://dx.doi.org/10.1098/rsta.2018.0080>.
- Coeckelbergh, M. (2020). Artificial intelligence, responsibility attribution, and a relational justification of explainability. *Science and Engineering Ethics*, 26, 2051–2068. <https://doi.org/10.1007/s11948-019-00146-8>.

- Dignum, V. (2018). Ethics and artificial intelligence: introduction to the special issue. *Ethics and Information Technology*, 20, 1–3. <https://doi.org/10.1007/s10676-018-9450-z>.
- EC (European Commission) (2018). European Commission for the Efficiency of Justice (CEPEJ). European ethical Charter on the use of Artificial Intelligence in judicial systems and their environment. Strasbourg, CEPEJ. Retrieved October 12, 2021 from <https://rm.coe.int/ethical-charter-en-for-publication-4-december-2018/16808f699c>.
- EC (European Commission) (2021). Digital skills and jobs. Retrieved October 12, 2021, from <https://digital-strategy.ec.europa.eu/policies/digital-skills-and-jobs>
- European Commission AI HLEG (High-Level Expert Group on Artificial Intelligence) (2019a). Ethics guidelines for trustworthy AI. <https://ec.europa.eu/futurum/en/ai-alliance-consultation/guidelines#Top>. Accessed April 8, 2019.
- European Commission AI HLEG (High-Level Expert Group on Artificial Intelligence) (2019b). Policy and investment recommendations for trustworthy AI. Accessed June 26, 2019, <https://futurum.ec.europa.eu/en/european-ai-alliance/open-library/policy-and-investment-recommendations-trustworthy-artificial-intelligence>.
- European Commission AI HLEG (High-Level Expert Group on Artificial Intelligence) (2020a). The Assessment List for Trustworthy Artificial Intelligence (ALTAI). https://ec.europa.eu/newsroom/dae/document.cfm?doc_id=68342. Accessed July 17, 2020.
- European Commission AI HLEG (High-Level Expert Group on Artificial Intelligence). (2020b). Sectoral Considerations on the Policy and Investment Recommendations for Trustworthy Artificial Intelligence. <https://futurum.ec.europa.eu/en/european-ai-alliance/document/ai-hleg-sectoral-considerations-policy-and-investment-recommendations-trustworthy-ai>. Accessed July 23, 2020.
- Felderer, M., & Ramler, R. (2021). Quality Assurance for AI-Based Systems: Overview and Challenges (Introduction to Interactive Session). In D. Winkler, S. Biffl, D. Mendez, M. Wimmer, & J. Bergsmann (Eds.), *Software Quality: Future Perspectives on Software Engineering Quality. SWQD 2021*. Lecture Notes in Business Information Processing, vol. 404. Springer. https://doi.org/10.1007/978-3-030-65854-0_3.
- Floridi, L., Cowls, J., Beltrametti, M., Chatila, R., Chazerand, P., Dignum, V., Luetge, C., Madelin, R., Pagallo, U., Rossi, F., Schafer, B., Valcke, P., & Vayena, E. (2018). AI4People—An ethical framework for a good AI society: Opportunities, risks, principles, and recommendations. *Minds and Machines*, 28, 689–707. <https://doi.org/10.1007/s11023-018-9482-5>.
- Goebel, R., Chander, A., Holzinger, K., Lecue, F., Akata, Z., Stumpf, S., Kieseberg, P., & Holzinger, A. (2018). Explainable AI: The new 42? In A. Holzinger, P. Kieseberg, A. Tjoa, E. Weippl (Eds.), *Machine learning and knowledge extraction*. Lecture Notes in Computer Science, vol. 11015 (pp. 295–303). Springer. https://doi.org/10.1007/978-3-319-99740-7_21.
- Grover, S., & Pea, R. (2018). Computational thinking: A competency whose time has come. In S. Sentance, E. Barendsen, & C. Schulte (Eds.) *Computer science education: Perspectives on teaching and learning* (pp. 19–38). London: Bloomsbury Academic.
- Hagendorff, T. (2020). The ethics of AI ethics: An evaluation of guidelines. *Minds and Machines*, 30(1), 99–120. <https://doi.org/10.1007/s11023-020-09517-8>.
- HAI (Human-Centred Artificial Intelligence). (2020). Human Impact Research Mission. Stanford University. Retrieved October 12, 2021, from <https://hai.stanford.edu/research/human-impact>.
- Holmes, W., Porayska-Pomsta, K., Holstein, K., Sutherland, E., Baker, T., Buckingham Shum, S., Santos, O., Rodrigo, M., Cukurova, M., Bittencourt, I., & Koedinger, K. (2021). Ethics of AI in Education: Towards a Community-Wide Framework. <https://doi.org/10.1007/s40593-021-00239-1>.
- Jobin, A., Ienca, M., & Vayena, E. (2019). The global landscape of AI ethics guidelines. *Nature Machine Intelligence*, 1(9), 389–399. <https://doi.org/10.1038/s42256-019-0088-2>
- Kousa, P., & Niemi, H. (2021). AI ethics and learning – educational technology companies' challenges and solutions. *Interactive Learning Environments* (Submitted in final review).
- Littman, M., Ajunwa, I., Berger, G., Boutilier, C., Currie, M., Doshi-Velez, F., Hadfield, G., Horowitz, M., Isbell, C., Kitano, H., Levy, K., Lyons, T., Mitchell, M., Shah, J., Sloman, S., Vallor, S., & Walsh, T. (2021). *Gathering Strength, Gathering Storms: The One Hundred Year*

- Study on Artificial Intelligence (AI100) 2021 Study Panel Report.* Stanford University, Stanford: CA. Retrieved September 16, 2021, from <http://ai100.stanford.edu/2021-report>.
- MEAE (Ministry of Economic Affairs and Employment of Finland) (2019) Leading the way into the age of artificial intelligence: Final report of Finland's Artificial Intelligence Programme 2019. Retrieved from October, 12, 2021, from <https://urn.fi/URN:ISBN:978-952-327-437-2>
- MEAE (Ministry of Economic Affairs and Employment, Finland) (2017). Turning Finland into a leading country in the application of artificial intelligence: Objective and recommendations for measures. Retrieved from October, 12, 2021, from https://julkaisut.valtioneuvosto.fi/bitstream/handle/10024/160391/TEMrap_47_2017_verkkojulkaisu.pdf?sequence=1&isAllowed=y
- Morley, J., Floridi, L., Kinsey, L., & Elhalal, A. (2020). From what to how: An initial review of publicly available AI ethics tools, methods, and research to translate principles into practices. *Science and Engineering Ethics*, 26, 2141–2168. <https://doi.org/10.1007/s11948-019-00165-5>.
- Niemi, H., & Kousa, P. (2020). A case study of students' and teachers' perceptions in a Finnish high school during the COVID pandemic. *International Journal of Technology in Education and Science (IJTES)*, 4(4), 352–369. <https://doi.org/10.46328/ijtes.v4i4.167>.
- Nnaji, J. (2019). Ethical dimensions of the increasing usage of new technologies in virtual education. In A. Blackburn, Chen, L., & Pfeffer, R. (Eds) *Emerging Trends in Cyber Ethics and Education* (pp. 1–21). USA: IGI Global.
- OECD (2019). Recommendation of the Council on Artificial Intelligence. Accessed May 22, 2019, <https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0449>.
- Spielkamp, M., Matzat, L., Penner, K., Thummler, M., Thiel, V., Gießler, S., & Eisenhauer, A. (2019). Algorithm Watch 2019: The AI Ethics Guidelines Global Inventory. Accessed April 9, 2019, <https://algorithmwatch.org/en/project/ai-ethics-guidelines-global-inventory/>
- Stone, P., Brooks, R., Brynjolfsson, E., Calo, R., Etzioni, O., Hager, G., Hirschberg, J., Kalyanakrishnan, S., Kamar, E., Kraus, S., Leyton-Brown, K., Parkes, D., Press, W., Saxenian, A., Shah, J., Tambe, M., & Teller, A. (2016). Artificial Intelligence and Life in 2030. One Hundred Year Study on Artificial Intelligence: Report of the 2015–2016 Study Panel, Stanford University, Stanford, CA. <http://ai100.stanford.edu/2016-report>.
- Torresen, J. (2018). A Review of Future and Ethical Perspectives of Robotics and AI. *Front. Robot. AI*, 4(75). <https://doi.org/10.3389/frobt.2017.00075>.
- Turing, A. M. (1950/2009). Computing machinery and intelligence. *Mind*, LIX (236), 433–460. <https://doi.org/10.1093/mind/LIX.236.433>.
- UNESCO (2019). Artificial Intelligence in Education: Challenges and Opportunities for Sustainable Development. Retrieved from January 5, 2022, from <https://unesdoc.unesco.org/ark:/48223/pf0000366994>.
- Winfield, A. (2019). An updated round up of ethical principles of robotics and AI. Accessed April 18, 2019, <http://alanwinfield.blogspot.com/2019/04/an-updated-round-up-of-ethical.html>.

Open Access This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.



Artificial Intelligence in Education as a Rawlsian Massively Multiplayer Game: A Thought Experiment on AI Ethics



Benjamin Ultan Cowley, Darryl Charles, Gerit Pfuhl,
and Anna-Mari Rusanen

Contents

1	Introduction	298
2	Theoretical Background	299
2.1	Explaining and Evaluating AI	299
2.2	MMOGs, MOOCs and Game-Based Learning	301
2.3	Role of AI in Education	302
2.4	Ethics of AIED	303
3	Methodology and Analysis	303
3.1	AIED-MMOG Schematic Technical Definition	304
3.2	Player Models	304
4	Findings	308
4.1	Rawlsian Justice Game	309
4.2	AIED-MMOG Rawlsian Justice Game	310
4.3	AIED-RJG for AI Evaluation	311
5	Discussion/Synthesis	311
5.1	Implications	312
5.2	Future Outlook	312
6	Conclusions	313
	References	313

B. U. Cowley (✉)

Faculty of Educational Sciences, University of Helsinki, Helsinki, Finland

Cognitive Science, Faculty of Arts, University of Helsinki, Helsinki, Finland

e-mail: ben.cowley@helsinki.fi

D. Charles

School of Computing, Engineering & Intelligent Systems, University of Ulster, Ulster, UK

e-mail: dk.charles@ulster.ac.uk

G. Pfuhl

Department of Psychology, UiT The Arctic University of Norway, Tromsø, Norway

Department of Psychology, Norwegian University of Science and Technology, Norway

e-mail: gerit.pfuhl@uit.no; gerit.pfuhl@ntnu.no

A.-M. Rusanen

Cognitive Science, Faculty of Arts, University of Helsinki, Helsinki, Finland

e-mail: anna-mari.rusanen@helsinki.fi

1 Introduction

Interest in the use of artificial intelligence (AI) within education (AIEd) has grown steadily over the past thirty years, and AI systems are already widely used in a non-teaching manner, e.g. for analytics, monitoring attainment or class planning. Indeed, the OECD firmly advocates the use of AI to measure and improve learning (Kuhl et al. 2019), leading towards digital, data-led governance and AI-based policy-making (Berendt et al. 2017). With the increasing power and ubiquity of computer technology and deep learning algorithms, AI now has the capacity to radically change education. For example, the natural language algorithm GPT-3 (Generative Pre-trained Transformer 3)¹ can help people write code, create websites or apps, co-author stories, summarise legal text and create virtual avatars that chat believably with a person. However, AI-led personalised teaching is in its infancy and carries challenges that must be foreseen and regulated. As yet, there is no methodology to help manage the trade-off between AIEd's possible benefits and challenges (Berendt et al. 2020).

A fundamental problem for AI is the challenge of evaluating AI algorithms in a fair and useful way (Hernández-Orallo 2017b), termed ‘explainable AI’ (XAI). In the education domain, this problem *prefigures* ethical issues because AIEd implies *also* evaluating the humans interacting with AI, who would otherwise (*sans* AI) be operating according to traditional norms. In other words, XAI in AIEd requires evaluating AI algorithms in a context where the performance is traditionally socially constructed (Latour 2005) and done by humans, for humans. Addressing this problem helps to make AIEd more transparent, which can help tackle the ethical quandaries of AIEd, such as distributive fairness (as defined normatively by, e.g. Rawls 1985). Reliable ways to tackle these issues are important for policy makers and other stakeholders involved in curricula development.

In this chapter, we describe a design for a formal setting, where the general concept of AI-enhanced education is simulated as a massively multiplayer online game (MMOG). The aim is to examine the representativeness of given algorithms for classes of individuals and thereby improve AI transparency, independently of which algorithm is examined. Simulations and games have an extensive track record for teaching and learning within the higher education sector (Lean et al. 2021). In response to the inherent problem of satisfying XAI within AIEd, the MMOG simulation provides a way to make benefit–risk comparisons in multi-stakeholder scenarios, including one which we illustrate explicitly as a thought experiment: the Rawlsian justice game (Rawls 1985) applied to the ethics of AI fairness. Rawls' theory and thought experiment game have been important in political philosophy for many decades, and this chapter is an initial attempt by us to integrate it into research on videogame-based learning.

¹ GPT-3 is a deep learning language model, see <https://openai.com/blog/gpt-3-apps/>.

In the rest of the chapter, we first describe the theoretical background in Sect. 2, and then Sect. 3 illustrates how the MMOG simulation is designed. Section 4 shows how the simulation integrates a Rawlsian justice game, and we discuss implications and future directions in Sect. 5.

2 Theoretical Background

Teaching is a dynamic and socially interactive process between at least two individuals (Powell and Kalina 2009) and requires adaptation to novelty, uncertainty and change to ensure efficient learning. AI, we argue, can assist human-guided teaching but requires some scaffolding to do so, and the scope of this requirement ranges from the pragmatic (e.g. XAI) to the epistemic.

2.1 *Explaining and Evaluating AI*

Hernández-Orallo (2017a) describes the crux of the AI evaluation problem: if AI research is the science of making intelligent machines, then algorithms should be evaluated on their intelligence; however, if AI is pragmatically about making machines that perform tasks that would require intelligence if done by humans, their evaluation should be a test of task performance. Thus, the form of the evaluation follows from the scope of the AI: general-purpose AI needs ability-focused evaluation (meaning *cognitive* abilities) and specialised AI needs task-focused evaluation (Hernández-Orallo 2017a).

Most work has been done on task-focused evaluation of specialised AI. Much of this work has had little regard for best practices of human psychometrics (e.g. comparing AI performance to a human reference from a single person) (Cowley et al. 2022). On the other hand, in visual object recognition, for example, (Rajalingham et al. 2018), the best studies are massive and systematic and illustrate great recent progress, as the algorithms become unsupervised and even begin to display biological plausibility (Zhuang et al. 2021). Such work also illustrates one popular method by which algorithms can be judged trustworthy: by human benchmarking. The general approach of benchmarking is central to AI development but has been criticised on grounds that treating a data benchmark as “independent of context, scope and specificity is... a false premise for machine learning evaluation” (Raji et al. 2021).

By contrast, human performance benchmarks are implicitly bound to context. For example, in the specialised AI domain of language models, recent work (Lin et al. 2021) reported a human benchmark designed to show model truthfulness (testing the well-known GPT-3 and variants). Results showed that the largest models made most errors, by learning popular misconceptions from the training data—in other words, the most ‘powerful’ AI was also most prone to learn errors hidden

in the data. Another study (Mohseni et al. 2021) designed a visual recognition benchmark from aggregate human attention data, surpassing benchmarks built on either ground truth image segmentation or human subjective ratings. *These task-focused evaluations illustrate a key issue in AIEd: effective evaluation correlates with ethical evaluation, as both require representative, unbiased, human-grounded training data and/or benchmarks.*

The primacy of task-focused evaluation derives in part from how AI systems typically overspecialise to the task, exemplified by Marcus' (Marcus 2018) list of 10 limitations of so-called 'deep' machine learning²: (1) data hungry, (2) limited transfer, (3) lack of hierarchical structure, (4) poor at open-ended inference, (5) not transparent, (6) not well-integrated to prior knowledge, (7) no causal representation, (8) presumes stability, (9) easily fooled and (10) hard to use for engineering. Any or all of these create serious problems in the domain of AIEd. Of course, other families of algorithms exist, but these also often leverage deep learning in some way, and come with their own challenges for evaluation (Henderson et al. 2019).

Even when task-focused evaluation can be done, there is still the challenge of how to use measured performance in a task to evaluate capability, without error-prone extrapolation. Focusing instead on evaluation of ability is not a silver bullet because abilities are constructs that must be defined, requiring a theoretical framework often derived from behavioural sciences. Bhatnagar et al. (2018) reports some work to map out intelligence in a general manner, and Hernández-Orallo (2017a) proposed a kind of universal psychometrics as a possible future solution. Nevertheless, ability- or intelligence-focused evaluation remains a hard, unsolved problem.

In a constrained context such as education, a hybrid approach might be viable given the wide range of preexisting tasks, and the proliferation of psychometrics or other testing instruments, available there. On the other hand, (as noted above) XAI evaluation requires representative data and benchmarks, and obtaining such presents a particular challenge in the education domain. This domain is replete with contraindications for, e.g. Marcus' list of deep learning vulnerabilities: learning transfer is required, data is hierarchical, learning ablates stability, etc.

The solution we propose, as a thought experiment, is exactly to constrain the domain by setting AIEd within an MMOG. Within such a *simulation* of the classroom, we can experiment with the potential effects of various AI designs. An MMOG-based simulation is a bounded domain with a well-defined application-programmer interface (API), yet nevertheless supports rich, emergent social interaction of players with varied roles. It also does not need to invent novel XAI solutions to individual AIEd problems: rather, the MMOG provides an operating environment where well-structured data and benchmarks can be obtained directly from the game engine.

AI in games has always been a field leader (Laird and Van Lent 2001; Vinyals et al. 2019), and this application domain can be leveraged to illustrate

² Not to be confused with deep human learning in education.

how problems of adaptivity and uncertainty can be dealt with in a well-defined context. For example, adaptive AI in games requires two constraints: to maintain logical consistency of game rules and a coherent ‘Magic Circle’ that preserves player immersion (Huizinga 1949). Games have also been used in XAI, e.g. the Arcade Learning Environment (Bellemare et al. 2013) and the General Video Game Competition (Perez-Liebana et al. 2016), which both consist of collections of game tasks designed to be solved by a single AI agent, and associated evaluations. These works aim to aggregate multiple task-focused evaluations and thereby measure general ability in some sense. Following this approach, the MMOG simulation we propose would ‘wargame’ various scenarios of AIEd.

2.2 MMOGs, MOOCs and Game-Based Learning

Here, we give background on the kinds of game we envision in our thought experiment. Already 40 years ago, Malone (1981) suggested video games can simultaneously deliver learning and motivation. Kirriemuir and others suggest digital games make excellent motivational tools that promote learning and engagement (Kirriemuir and McFarlane 2004), because they intrinsically motivate players to progress in the absence of extrinsic rewards (Malone et al. 1987) and thus engage the player to master a challenge that can be difficult, prolonged and complex (Charles 2010).

Game design also has a lot to offer to learning design, as Gee (2003) outlined with his taxonomy of learning principles in games, which then inspired our own work on learning designs for MMOGs (Cowley et al. 2011). In more recent times, ‘gamification’ and ‘gamefulness’ in learning have become popular topics of applied research. Often the focus of these approaches is using games and theories from cognitive and educational psychology to help support and motivate learning—mirroring the long-established use of games in political philosophy (Rawls 1985).

Game playing can be a very social activity, and some of the most popular recent games are only online, including shooter games like Destiny or Fortnite, real-time strategy games like Dota or StarCraft or roleplay games (MMORPGs) like RuneScape and Final Fantasy XIV. A large part of the appeal of multiplayer games is in the strong social bonds that can be built through co-operation and competition in structured play within an ‘unreal’ environment, each player taking on a role in a fantasy world.

In the early 2000s, MMOGs were a ‘natural laboratory’ to study how individuals interact online, and proposed as a tool for digitising education (Cowley et al. 2011; Sourmelis et al. 2017). MMOGs enable two features valued in education: role-taking (expressing ‘versions’ of oneself in different contexts) and groupwork (important for developing skills transferable to the workplace). Furthermore, a multi-user environment provides a richer context for player choice and a wider psychological basis for behavioural variation than single-player scenarios; for example, explicit

competition and collaboration with others, socialising, philanthropy, disruptive behaviour (e.g. ‘griefing’, ‘trolling’, cheating) etc.

The MMOG is a useful conceptual construct, not least because it has been so well studied, and serves well as the design for a thought experiment simulation. MMOGs also have one distinct advantage over the newer forms of social online platforms: being games, they naturally conform better to the characteristics of formal games, i.e. they describe the behaviour of rational agents (rationality here defined by the rules of the game, entered into knowingly by the players, viz the Magic Circle Huizinga 1949). This allows us to reason about the behaviour of players with confidence.

2.3 *Role of AI in Education*

We consider AIED as incorporating the traditional roles of learners and teachers within a socially constructed educational milieu (Latour 2005). In other words, we start from the assumption that all roles, for human or AI players, for staff or students, are derived from equivalent fundamentals and obtain their unique character through emergence by social construction. This is in line with Actor-Network Theory (ANT) (Latour 2005), which posits that everything in the social and natural worlds exists in constantly shifting networks of relationships. Rather than a predictive theory, ANT provides an empirical ‘form of inquiry’, which we follow by exploiting the bounded structure and complete access to activity data of MMOGs, to track ‘players’ and their interactions.

The roles within the classroom are flexible and mutable. Teacher(s), learner(s), and the social group—e.g. the peer group from the point of view of a given learner—sometimes have more teaching and sometimes more learning motivations. That is, teachers are sometimes in training, and thus also learners. And learners sometimes act as teaching assistants or peer mentors, and are thus also teachers. And this conforms to the socially constructed view, since social constructions are goal oriented. In the general sense, the milieu is not defined by fixed, assigned *roles*, but by shifting relational *goals*.

The future of education must now accommodate another role: AI. How AI-driven roles might perturb the socially constructed equilibrium of the classroom is not known *a priori*: in fact every format of the technology can have a different effect. AI-based learning analytics will play a different role to AI instructional agents or to AI agent-based models of individual learners. How should one anticipate or control the ethical goodness of such unforeseen outcomes?

2.4 *Ethics of AIEd*

From a wider epistemic point of view, AI and other smart technologies change not only the traditional social or physical environments of learning, but also impact the epistemic distribution of labour in classrooms. The role-taking example described above is one example. Thus, AIEd raises a need to evaluate the norms governing the practices of epistemic communities. For example, when cognitive tasks are delegated to machines, it may impact on assessments of ‘trustworthiness’. Trust, or reliance, binds the individual epistemic actors into knowledge communities.

Crucially, in AIEd-based knowledge communities, the individuals need not only extend trust to other individuals but also to instruments and equipment they use. That is, individuals should be able to have reliance that epistemic artefacts—such as computers or data analysis methods—work correctly and generate accurate outcomes.

The opacity of contemporary AI applications threatens this binding of reliance and trust. Many current machine learning systems (such as Deep Neural Networks) are so-called ‘black box’ systems. By definition, we cannot fully explain how such systems work, and thus we cannot fully rely on them as epistemic instruments. This raises a fundamental and deep challenge for the deployment of these technologies as epistemic instruments in knowledge communities (Lo Piano 2020).

There are also many open questions regarding what constitutes transparency or explainability for classroom technologies and what level of transparency is sufficient for different epistemic actors with various positions and roles. For each actor, the interpretation and requirements of ‘transparency’ may vary. While for a teacher (responsible public sector actor), transparency may require a sufficient understanding of the reliability of a student assessment system, for a student, transparency may mean a comprehensible justification for the decision being made. Or, transparency required to analyse legal significance of unjust biases in learning analytics may mean a different thing than explainability in computer science terms.

Thus, there is a need to develop how we analyse and assess the nuanced aspects of explainability for different actors in different classroom situations. The AIEd-MMOG we consider herein aims to address this need.

3 Methodology and Analysis

In this section, we define a complete schema of an AIEd-MMOG, which we use in Sect. 4 to examine the potential ethical problems of AIEd fairness.

First, we define the setting and the population. The thought experiment proposes launching the AIEd-MMOG in teacher training courses sited within several third-level institutions, wherein prospective teachers learn about the uses and challenges of AI in their future career. This setting provides the following features:

- Players are adults only, avoiding issues related to both developmental mutability and legal/ethical issues of child protection.
- This setting maximises the versatility of role-taking, since trainee teachers can meaningfully embody both teachers and students and consider perspectives on a range of subject-matter disciplines.
- The social power hierarchy is close to ‘flat’, which helps to prevent unwitting exploitation and avoid undesired influences of power imbalance

3.1 AIED-MMOG Schematic Technical Definition

The AIED-MMOG will take the form of an open-world ‘sandbox’ style game, wherein various tools and toys pre-exist within a single large environment (the sandbox), which allows players freedom to engage as they prefer. This is a similar format as some of the most popular games of recent years, including Fortnite and Grand Theft Auto V. In such settings, avatar API can be run by humans or agent AI—i.e. the actors in the game (avatars) are like robots whose actions are ‘programmed’ by either human or AI.

The AIED-MMOG will leverage off-the-shelf technology (i.e. pre-existing and ready to use), such as the Unity game engine, which provides a vast array of software libraries to exploit. This technology will be used to build an environment to support a variety of different learning goals, by packaging learning content as ‘mini-games’. Such mini-games can indeed be simple games, or teaching/training tools, or aptitude tests, or hybrids of any/all of the above. Gamified cognitive tests illustrate one way to make such hybrids (Lumsden et al. 2016).

This design ethos of a social online world with embedded modular content has been trialled and evaluated in Cowley et al. (2011) and Cowley and Bateman (2017). Figure 1 shows example screens and architecture from an AIED-MMOG previously designed by the first author: this example game illustrates how such a game could be structured. Other educational games have also exemplified this design ethos, for example, ‘Real Lives: you are the world’ (Educational Simulations 2010).

The MMOG content will be versatile due to its modular design, permitting ‘minigame’ activities to also present moral dilemmas, such as those used to study AI ethics in Sundvall et al. (2021). Compared to such survey-based research, this setting offers the advantage that the dilemmas are lived and not just self-reported on—in other words, participants will not just view a moral dilemma vignette but will face the dilemma on their own behalf.

3.2 Player Models

Within the sandbox-and-minigames environment of our AIED-MMOG, player behaviour will conform (with some margin of error) to certain predictable patterns

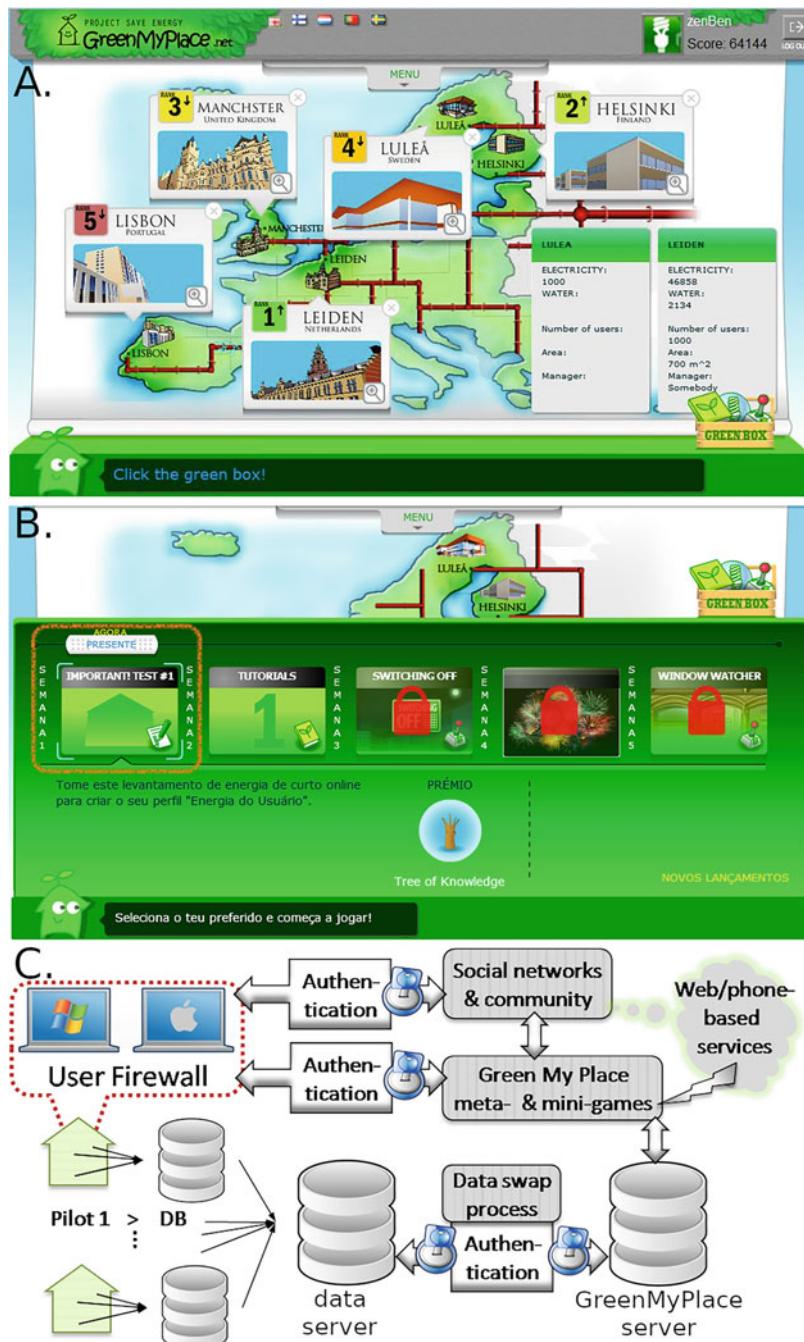


Fig. 1 An exemplar MMOG taken from the first author's earlier work. GreenMyPlace was a massively multiplayer social online game designed to teach concepts of energy efficiency and

because play must conform to how each game was designed to be played, i.e. to game design patterns. This does not mean all players must take exactly the same actions, merely that actions are similar and follow some clustering. This predictability will be exploited to model the types of play behaviour, which can give insight into the player themselves, when tracked over time as a player model. Importantly, insights derived from human players can also be applied to AI agent-based players, since all players use the same API to interact with the game and each other. This helps address the fundamental XAI challenge of equitable evaluation of human–AI activity.

In the MMOG simulation, the abilities that players use to interact are constructed from a hierarchy of tasks. This concept of a hierarchy of tasks that encapsulates the mechanics of a game has been termed *skill atoms* (by Daniel Cook³). A skill atom consists of a game action, which results in the application of game rules to change game state in the simulation, and the provision of feedback to the player. Based on this, a process occurs in which the player updates their mental model of the game as a system. The formalism of skill atoms is analogous to a finite-state machine. Furthermore, composition of skill atoms into chains of actions can be used to capture player behaviour (see Fig. 2).

Previously, we showed how such ‘skill-atom chains’ of behaviour can be linked to player temperament to derive micro-models of play preference, called Behavlets (Cowley and Charles 2016). The Behavlets method leverages domain-expert knowledge of game design patterns, to encode short activity sequences that represent an aspect of playing style or player personality traits (e.g. aggressive or cautious play), which can be mapped to temperament theory. Behavlets have been used to profile players by their play preference (Cowley et al. 2013).

Behavlets can be further analysed as temporally extended sequences called ‘B-chains’ (Charles and Cowley 2020). The skill-atom>Behavlets>B-chain stack of methods can be considered a hierarchically arranged model of a ‘player’, each layer trading detail for generality, which when combined serves several purposes:

- 1. Efficiency:** Behavlets reduce the dimensionality of game-play data, enhancing algorithmic efficiency and allowing comparison between players in terms of meaningful action

◀

Fig. 1 (continued) promote relevant behaviour change. Panel **a**: the top-level game involved 5 participating pilot locations around Europe which formed the game’s ‘teams’. Panel **b**: each individual player participated by selecting varied activities from a ‘Green Box’. Panel **c**: the game architecture illustrates a model game design for our purposes, providing a controlled, authenticated flow of data from local sites (e.g. classrooms) to online servers, to end-user devices—and back again

³ <https://www.gamedeveloper.com/design/the-chemistry-of-game-design>.

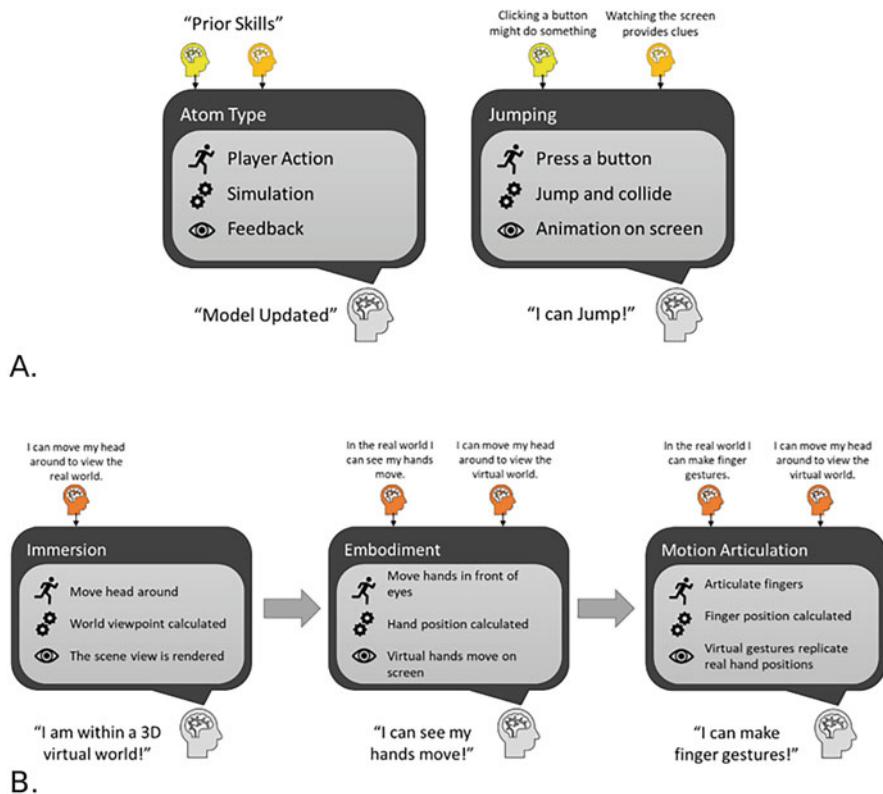


Fig. 2 The concept of a skill atom and their composition to produce skills. Panel **a**: skill atom prototype (left) and jump atom example (right). Panel **b**: an atom skill chain for a natural hand movement game controller in virtual reality

2. **Privacy:** such dimensional reduction means that Behavlets obscure the exact behaviours of individuals, such as key press or mouse movement logs, which have been shown to allow automated identification
3. **Profiling:** clustering human player data permits identification of playing styles (Cowley et al. 2013), which gives insight into AI agent players behaviour

Individual players may have different preferences for their style of play and thus vary in their motivations. We can capture such preferences by the above approach (Cowley et al. 2013); then, using a semi-supervised learning approach based on tracking which Behavlets are triggered by players (human or AI), information is gained on which play preferences are expressed. Similarly, the AIED-MMOG simulation provides representative data on human attention from the Behavlet and B-chain model layers, which can be used for task-focused benchmarking (within whatever mini-game task the data is taken from).

In Summary, the MMOG simulation for AIEd provides *a terrarium-society environment, wherein interactions can develop according to natural social patterns but bounded by constraints that ensure safety, explainability, reproducibility and transparency of outcomes* (Lo Piano 2020).

4 Findings

In a real educational policy situation, how can the AIEd-MMOG help authorities to decide whether a school or an educational system should deploy a given AI algorithm? Let us consider a concrete example. Suppose you are the Chief Digital Officer in a school district. You are asked to consider whether the region's educational organisation should move from a 'reactive' student guidance system to a 'preventive' guidance system. It would be a novel, sophisticated machine learning system that would help authorised school personnel, such as social workers, to forecast the possible social, cognitive or psychological learning problems of elementary school students. These methods would produce predictions by combining and analysing various sources of student data, including their learning results and, say, medical records. By analysing a large amount of criteria data, high-risk individuals could be identified and prioritised. These high-risk individuals could proactively be invited to meet with school tutors, social workers, counsellors or psychologists, to get guidance and help.

Obviously, the preventive system would have many positive possibilities, including potential to improve overall well-being of students. Furthermore, it might allow better student supervision, supportive actions and impact estimation. At the same time, the preventive system raises several legal and ethical issues regarding privacy, security and use of data. It raises the fundamental question of justification: do the authorities have a principled right to use private and sensitive data for identifying high-risk students, and if so, to what extent? And, if these systems are used, will individuals be treated in an equal way? What exactly *is* equality in this context, where different individuals have different needs and roles? How to distribute whatever resources are inherent in deployment of an AI algorithm, to ensure a fair and just outcome for all students, when the algorithms and their deployment mechanisms are not (cannot be) transparent?

AIEd-MMOGs could provide a formal setting for simulating these situations, where individuals do not start from the same position, and there are individual differences that matter. These simulations can be used to bring together, e.g. philosophical ideas on distributive justice, with an active, instantiated environment that facilitates testing of various alternative approaches to real-world scenarios.

4.1 Rawlsian Justice Game

According to John Rawls' theory of justice, the distribution of resources should maximise the benefits to the members who start with minimal resources (Rawls 1985). The most important principle of fairness is to ensure that the 'least advantaged' members of society will benefit and not be harmed. The distribution of resources that maximises the benefit to the members who start with minimal resources is the *maximin distribution*. Rawls' idea was that individuals in a society must choose their preferred distribution function with no foreknowledge of their own status in the society: a feature dubbed the *veil of ignorance*. From behind the veil of ignorance, Rawls claims, individuals will tend to select the maximin distribution.

Howe and Roemer (1981), among others, have described how Rawlsian justice can be modelled as a game, in their case for economic distribution. Such *Rawlsian justice games* (RJGs) have been used as classroom teaching tools in a variety of disciplines including political science and economics (Alden 2005), where students debate and select various distributive principles.

In the Howe and Roemer (1981) model, individuals from a population $p \in P$ will each receive an *endowment* α (which ranges in $[a, b]$), under some probability distribution $f(P)$. Now, the veil of ignorance prevents any p from knowing about $\alpha(p)$, but they may know about f . Endowment is converted to income Y (aka "the good") via some production function. Redistribution of incomes, which is the scheme that the population may choose via the game, is modelled as a tax τ .

Howe and Roemer (1981) then describe the *incentive problem*: as $\tau(p)$ rises, p will produce less *pre-tax* income. This is modelled by the production-incentive function $g(\alpha, \tau)$, which maps endowment and tax to income produced, which is $(1 - \tau)g(\alpha, \tau)$. Behind the veil of ignorance, the population know g exists, but not how it operates. Howe and Roemer (1981) go on to define tax schemes and the maximin distribution within this model, the details of which are not critical here. Then they describe how a game can be structured: behind the veil of ignorance, every p will aim to choose a tax scheme $\tau(p)$, such that, after endowments α are assigned and post-tax incomes Y realised, no *coalition* of players $p_{i..j} \subset P$ can improve on Y by attempting to draw again from f . Multiple draws on f may be expected before an equilibrium is reached. In other words, a key aspect of the RJJG is that it will "...allow an individual or coalition of individuals to express its dissatisfaction with a particular income distribution by hypothetically withdrawing from society, and testing whether under the rules of the game it can improve the lot of its members" (Howe and Roemer 1981).

We suggest that such a permutation testing mechanism can function as an evaluation of AI algorithms. To implement that we envision a selection of social learning mini-games within the MMOG, each game distinguished by variants of an AI algorithm. This social learning mini-game can be anything, so long as it prescribes some type of multi-user engagement (such as group-wise problem-based learning, PBL) and supports standard testing of learning outcomes (to enable

quantification and then automation of evaluation). Also critical is that in the AIEd-MMOG design described, performance in mini-games contributes to overall performance of one's 'team' (physical institution), thus the immediate outcome is important on the macro scale.

4.2 AIEd-MMOG Rawlsian Justice Game

To adapt the mechanics of the RJG to the AIEd-MMOG environment and obtain an *AIEd-RJG*, we must further answer the question: what is justice in this AIEd domain? What does the social contract govern, and/or what is justice regulating (since it is not monetary income)? To answer this, the implementation should map from the traditional concepts of the game to concepts that make sense in the domain. For 'income' we map from income as money to income as learning (measured by standardised test). For endowment, we map from endowment-as-social-status to endowment-as-representation, i.e. how representative was the training data for each person? In terms of our concrete example of preventive guidance, this will translate to accurately can a student be assessed based on the representation of their characteristics in the data.

Thus, justice will be defined as relevance of the AI to each person's given background and ability to learn and test well. If the AI does well for a given student, then the student should elect to continue with that particular algorithm; or conversely, they may elect to switch to another environment with an alternative algorithm. *However*, because mini-games within the RGJ depend on group-based PBL, then switching to another environment can only maximise learning outcomes if many other players *also* switch.⁴ In practice this will tend to favour joint action by 'coalitions', as in the original RJG (1981).

Thus, our proposed AIEd-RJG will function at the 'level' of the MMOG, accumulating data on the 'goodness' of separate draws on f via the mechanism of players' choice of mini-games. The adapted AIEd-RJG will then operate as follows:

- Individuals $p \in P$ become learning agents $l \in L$.
- Endowment α becomes representation $\bar{\alpha}$.
- Income Y becomes learning (test score) \bar{Y} .
- Endowment probability distribution f becomes the probability of occurrence of L in the AI training data, denoted \bar{f} .
- The tax rate τ becomes the cost of mentoring peers in $l_{i...j} \subset L$, denoted $\bar{\tau}$.
- Production-incentive function g becomes the personal learning incentive \bar{g} ⁵
- The personal learning outcome is then $(1 - \bar{\alpha})g(\bar{\alpha}, \bar{\tau})$

⁴ This mechanic is similar to when players in commercial MMOGs switch between servers.

⁵ The interpretation of \bar{g} is that intrinsic motivation to learn is weighed against the need to help peers to lift up team performance.

\bar{f} and \bar{g} are both unknown in AIEd-RJG because of *AI non-transparency*, corresponding to the veil of ignorance! They can be estimated from sufficient repeated plays (draws on \bar{f}). In other words, by actually playing the game, L creates data to estimate \bar{f} , \bar{g} . Play will end when no one wishes to try for better learning scores by sampling again (hoping for better representation).

However, the volume of play which is sufficient might be onerous for human players, especially recruited from a teacher training programme. To help solve this, the human-based play data can be supplemented with AI agent-based play, by training AI agent algorithms to play in a manner that emulates human playing style based on the ‘seed’ games played initially by humans. The techniques to do this are beyond the scope of this paper, however they rely on the methods described in Sect. 3.2 for modelling player personality, i.e. skill atoms, Behavlets and B-chains.

4.3 AIEd-RJG for AI Evaluation

Returning to the question of XAI evaluation, our AIEd-MMOG simulation follows from prior game-based AI evaluation work (Bellemare et al. 2013; Perez-Liebana et al. 2016), suggesting that ability can be evaluated from the aggregate of task evaluations. Our simulation adds the capability to assess (a) social influences on task performance from player-to-player interactions and (b) the representativeness of given algorithms for classes of individuals. This all aims to improve AI transparency, independently of which algorithm is used: although we cannot always see inside the black box, we *can* forecast how it behaves.

Note that in this approach, two kinds of AIEd algorithm can actually be tested: (a) agent-based AI that plays the game alongside humans or (b) the analytics/oversight algorithm that models players and distributes the ‘social good’ (thus conforming to the concept of a market principle in original RJG).

5 Discussion/Synthesis

In this chapter, we have presented a thought experiment on how to use an MMOG simulation to study AIEd deployment solutions, focusing on the fundamental challenge of explainable AI, examined through the lens of Rawlsian distributive justice.

As stated by Schulzke (2012), ‘by taking a concept like distributive justice out of the realm of theoretical speculation and making it part of a simulation, games provide an excellent means of recontextualising the problem by giving players firsthand, concrete experience of that problem’.

Schulzke in fact examined the educational game *Real Lives* from the perspective of an RJJG, thus linking it to our thought experiment by design format. That work focused on natural justice, not AIEd, and within *Real Lives* the Rawlsian lesson is never explicit. However, Schulzke's commentary shows the relevance of an MMOG-format game for examining Rawlsian concepts. Rawlsian justice has also been modelled in the context of AI ethics (Leben 2017), although (to our knowledge) our work is first to situate an RJJG within AIEd.

5.1 Implications

Responsible AI requires that choices and decisions be explicitly reported and open to inspection, i.e. they meet the ART principles: Accountability, Responsibility and Transparency (Dignum 2021, p-3).

Accountability includes that all stakeholders are involved in defining the moral values and societal norms that AI represents (is designed for). *Responsibility* encompasses the user's relation to AI, already at development and also when using the system. *Transparency* refers to describing, inspecting and reproducing how the AI system learns to make decisions and adapt to its environment, thus ensuring trust. Transparency also refers to explicitly and openly describing data sources for training, development processes and stakeholders. Not meeting ART requirements can lead to stakeholder dissatisfaction and 'bandaid' fixes, such as post hoc regulation.

The AIEd-MMOG meets all ART principles. Accountability, because the environment combines top-down designed constraints on actions with a bottom-up process of social construction to shape the games' moral norms. Responsibility, because building the human–AI relationships on a foundation of well-defined XAI permits comprehensive comparable evaluation. Transparency, because the MMOG is a strictly bounded environment where code is open, data has clear provenance, and actions cannot be hidden—they are even associated with action-motivations through the Behavlets and with action-context through the B-chains.

What is more, the setting provides the opportunity to explicitly represent varied moral stances as minigames, which allow human or AI players to demonstrate their own values as choices.

Finally, given our aim of supporting XAI for more transparent, interpretable and ethical AIEd, note that the MMOG simulation facilitates *reproducible* AI (Pineau et al. 2020), compared to deployment in a live classroom.

5.2 Future Outlook

Successful teaching relies on pedagogic rights and teacher–student relationships governed by enhancement, participation and inclusion (Reiss 2021). *Enhancement*

is education for critical thinking. *Participation* means that the users have the right to be separate and autonomous and not subsumed with the system. *Inclusion* facilitates representative democratic structures, i.e. avoiding dominance of commercial or governmental providers. These pedagogic rights align with acting morally in the humanist sense (2021). A corollary is that any AI system should be made for, but also by, the users who then decide which AI systems are used, and how.

Conscious and well-informed... individuals will create a solid foundation for responsible and positive uses of AI systems and digital technologies more generally, and strengthen their personal skills on cognitive, social and cultural levels. This will not only increase the available talent pool, but also foster the relevance and quality of research and innovation of AI systems for society as a whole.

(European Commission, n.d.)

These rights correspond to large-scale implementation demands, touching on the AIEd challenges discussed above. By facilitating some small progress towards tackling those challenges, our AIEd-MMOG would allow potential issues to be identified and without running expensive and time-consuming live trials.

6 Conclusions

Although human-performed evaluation in education is sometimes imperfect, it is also important to consider that AI evaluation can be biased, leading to problems of underestimating AI systems or setting too high a bar on them (Buckner 2021). We have described a thought experiment aimed at addressing this dual evaluation issue within the new frontier of AIEd. The proposed AIEd-MMOG is simply a constrained and well-defined *setting* for AI to enter education, a proposed set of features that facilitate bottom-up/task-focused XAI evaluation within a social milieu with deployed AI. In this setting, we have shown how an RJJ design could improve AI transparency by estimating how representative is a given algorithm for various classes of individuals.

References

- Alden, L. (2005). Econoclass: The distributive justice game. Retrieved 2 December 2021, from <http://www.econoclass.com/distjusticegame.html>
- Bellemare, M. G., Naddaf, Y., Veness, J., & Bowling, M. (2013). The Arcade Learning Environment: An Evaluation Platform for General Agents. *Journal of Artificial Intelligence Research*, 47, 253–279.
- Berendt, B., Littlejohn, A., & Blakemore, M. (2020). AI in education: Learner choice and fundamental rights. *Learning, Media and Technology*, 45 (3), 312–324.
- Berendt, B., Mitros, P., Shacklock, X., Blakemore, M., Littlejohn, A., & Kern, P. (2017). *Big data for monitoring educational systems*. Publications Office of the European Union.

- Bhatnagar, S., Alexandrova, A., Avin, S., Cave, S., Cheke, L., Crosby, M., Feyereisl, J., Halina, M., Loe, B. S., Ó hÉigearaigh, S., Martínez-Plumed, F., Price, H., Shevlin, H., Weller, A., Winfield, A., & Hernández-Orallo, J. (2018). Mapping Intelligence: Requirements and Possibilities. In V. C. Müller (Ed.), *Philosophy and Theory of Artificial Intelligence 2017* (pp. 117–135). Springer International Publishing.
- Buckner, C. J. (2021). Black Boxes, or Unflattering Mirrors? Comparative Bias in the Science of Machine Behavior. *The British Journal for the Philosophy of Science*.
- Charles, D., & Cowley, B. U. (2020). Behavlet Analytics for Player Profiling and Churn Prediction. In C. Stephanidis, D. Harris, W.-C. Li, D. D. Schmorow, C. M. Fidopiastis, P. Zaphiris, A. Ioannou, X. Fang, R. A. Sottilare, & J. Schwarz (Eds.), *HCI International 2020 – Late Breaking Papers: Cognition, Learning and Games* (pp. 631–643). Springer International Publishing.
- Charles, T. (2010). *Enhanced e-learning engagement using game absorption techniques ELEGANT* (Doctoral dissertation). University of Ulster.
- Cowley, B., Charles, D., Black, M., & Hickey, R. (2013). Real-time rule-based classification of player types in computer games. *User Modeling and User-Adapted Interaction*, 23 (5), 489–526.
- Cowley, B., Moutinho, J. L., Bateman, C., & Oliveira, A. (2011). Learning principles and interaction design for 'Green My Place': A massively multiplayer serious game. *Entertainment Computing*, 2 (2), 103–113.
- Cowley, B., & Bateman, C. (2017). Green My Place: Evaluation of a Serious Social Online Game Designed to Promote Energy Efficient Behaviour Change. *International Journal of Serious Games*, 4 (4), 71–90.
- Cowley, B., & Charles, D. (2016). Behavlets: A Method for Practical Player Modelling using Psychology-Based Player Traits and Domain Specific Features. *User Modeling and User-Adapted Interaction*, 26 (2), 257–306.
- Cowley, H. P., Natter, M., Gray-Roncal, K., Rhodes, R. E., Johnson, E. C., Drenkow, N., Shead, T. M., Chance, F. S., Wester, B., & Gray-Roncal, W. (2022). A Framework for Rigorous Evaluation of Human Performance in Human and Machine Learning Comparison Studies. *Scientific Reports*, in press.
- Dignum, V. (2021). The role and challenges of education for responsible AI. *London Review of Education*.
- Gee, J. P. (2003). *What video games have to teach us about learning and literacy*. Palgrave Macmillan.
- Henderson, P., Islam, R., Bachman, P., Pineau, J., Precup, D., & Meger, D. (2019). Deep Reinforcement Learning that Matters. *arXiv:1709.06560*.
- Hernández-Orallo, J. (2017a). Evaluation in artificial intelligence: From task-oriented to ability-oriented measurement. *Artificial Intelligence Review*, 48 (3), 397–447.
- Hernández-Orallo, J. (2017b). *The Measure of All Minds: Evaluating Natural and Artificial Intelligence*. Cambridge University Press.
- Howe, R. E., & Roemer, J. E. (1981). Rawlsian Justice as the Core of a Game. *The American Economic Review*, 71 (5), 880–895.
- Huizinga, J. (1949). *Homo Ludens: A study of the play-element of culture*. Routledge.
- Kirriemuir, J., & McFarlane, A. (2004). Report 8: Literature review in games and learning. *FuturLab Series*.
- Kuhl, P. K., Lim, S.-S., Guerrero, S., & Damme, D. v. (2019). *Developing Minds in the Digital Age*.
- Laird, J. E., & Van Lent, M. (2001). Human level AI's killer application: Interactive computer games. *AI Magazine*, 22 (2), 15–25.
- Latour, B. (2005). *Reassembling the social: An introduction to actor-network-theory*. Oxford University Press.
- Lean, J., Moizer, J., Derham, C., Strachan, L., & Bhuiyan, Z. (2021). RealWorld Learning: Simulation and Gaming. In D. A. Morley & M. G. Jamil (Eds.), *Applied Pedagogies for Higher Education: Real World Learning and Innovation across the Curriculum* (pp. 187–214). Springer International Publishing.

- Leben, D. (2017). A Rawlsian algorithm for autonomous vehicles. *Ethics and Information Technology*, 19 (2), 107–115.
- Lin, S., Hilton, J., & Evans, O. (2021). TruthfulQA: Measuring How Models Mimic Human Falsehoods. *arXiv:2109.07958*.
- Lo Piano, S. (2020). Ethical principles in machine learning and artificial intelligence: Cases from the field and possible ways forward. *Humanities and Social Sciences Communications*, 7 (1), 1–7.
- Lumsden, J., Edwards, E. A., Lawrence, N. S., Coyle, D., & Munafò, M. R. (2016). Gamification of Cognitive Assessment and Cognitive Training: A Systematic Review of Applications and Efficacy. *JMIR Serious Games*, 4 (2), e5888.
- Malone, T. W., Lepper, M. R., Snow, R., & Farr, M. (1987). Making learning fun: A taxonomy of intrinsic motivations for learning. *Aptitude, learning and instruction: III. conative and affective process analyses* (pp. 223–253). Hillsdale.
- Malone, T. W. (1981). What makes computer games fun? [for education]. *BYTE*, 6 (12), 258–277.
- Marcus, G. (2018). Deep Learning: A Critical Appraisal. *arXiv:1801.00631*.
- Mohseni, S., Block, J. E., & Ragan, E. (2021). Quantitative Evaluation of Machine Learning Explanations: A Human-Grounded Benchmark. *26th International Conference on Intelligent User Interfaces*, 22–31.
- Perez-Liebana, D., Samothrakis, S., Togelius, J., Schaul, T., Lucas, S. M., Couëtoux, A., Lee, J., Lim, C.-U., & Thompson, T. (2016). The 2014 General Video Game Playing Competition. *IEEE Transactions on Computational Intelligence and AI in Games*, 8 (3), 229–243.
- Pineau, J., Vincent-Lamarre, P., Sinha, K., Larivière, V., Beygelzimer, A., d'Alché-Buc, F., Fox, E., & Larochelle, H. (2020). Improving Reproducibility in Machine Learning Research (A Report from the NeurIPS 2019 Reproducibility Program). *arXiv:2003.12206*.
- Powell, K. C., & Kalina, C. J. (2009). Cognitive and Social Constructivism: Developing Tools for an Effective Classroom. *Education*, 130 (2), 241–250.
- Rajalingham, R., Issa, E. B., Bashivan, P., Kar, K., Schmidt, K., & DiCarlo, J. J. (2018). Large-Scale, High-Resolution Comparison of the Core Visual Object Recognition Behavior of Humans, Monkeys, and State-of-the-Art Deep Artificial Neural Networks. *Journal of Neuroscience*, 38 (33), 7255–7269.
- Raji, I. D., Denton, E., Bender, E. M., Hanna, A., & Paullada, A. (2021). AI and the Everything in the Whole Wide World Benchmark.
- Rawls, J. (1985). Justice as Fairness: Political not Metaphysical. *Philosophy & Public Affairs*, 14 (3), 223–251.
- Reiss, M. J. (2021). The use of AI in education: Practicalities and ethical considerations. *London Review of Education*.
- Schulzke, M. (2012). Using Video Games to Think About Distributive Justice. *The Journal of Interactive Technology and Pedagogy*, (2).
- Sourmelis, T., Ioannou, A., & Zaphiris, P. (2017). Massively Multiplayer Online Role Playing Games (MMORPGs) and the 21st century skills: A comprehensive research review from 2010 to 2016. *Computers in Human Behavior*, 67, 41–48.
- Sundvall, J., Drosinou, M., Hannikainen, I. R., Elovaara, K., Halonen, J., Herzon, V., Kopecky, R., Kšová, M. J., Koverola, M., Kunnari, A., Perander, S., Saikkonen, T., Palomäki, J., & Laakasuo, M. (2021). *Innocence over utilitarianism - Heightened Moral Standards for Robots in Rescue Dilemmas PREPRINT* (tech. rep.). PsyArXiv.
- Vinyals, O., Babuschkin, I., Czarnecki, W. M., Mathieu, M., Dudzik, A., Chung, J., Choi, D. H., Powell, R., Ewalds, T., Georgiev, P., Oh, J., Horgan, D., Kroiss, M., Danihelka, I., Huang, A., Sifre, L., Cai, T., Agapiou, J. P., Jaderberg, M., ... Silver, D. (2019). Grandmaster level in StarCraft II using multi-agent reinforcement learning. *Nature*, 575 (7782), 350–354.
- Zhuang, C., Yan, S., Nayebi, A., Schrimpf, M., Frank, M. C., DiCarlo, J. J., & Yamins, D. L. K. (2021). Unsupervised neural network models of the ventral visual stream. *Proceedings of the National Academy of Sciences*, 118 (3).

Open Access This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.



Four Surveillance Technologies Creating Challenges for Education



Roy D. Pea, Paulina Biernacki, Maxwell Bigman, Kelly Boles, Raquel Coelho, Victoria Docherty, Jorge Garcia, Veronica Lin, Judy Nguyen, Daniel Pimentel, Rose Pozos, Brandon Reynante, Ethan Roy, Emily Southerton, Miroslav Suzara, and Aditya Vishwanath

Contents

1 Introduction	317
2 Surveillance State	318
2.1 Location Tracking	318
2.2 Facial Identification Technologies (FITs)	320
2.3 Automated Speech Recognition	322
2.4 Social Media Mining	323
3 Call to Action	325
3.1 Research	325
3.2 Policy and Law	326
3.3 Practice	326
4 Conclusion	327
References	328

1 Introduction

Accelerating embeddedness of information and communication technologies in our social and physical worlds requires reflection on the future of learning environments and educational research. The ubiquitous AI—embodied in cloud computing web services which detects empirical patterns in accruing data, coupled with sensors in

R. D. Pea (✉)

Graduate School of Education, Stanford University, Stanford, CA, USA
e-mail: roypea@stanford.edu

P. Biernacki · M. Bigman · K. Boles · R. Coelho · V. Docherty · J. Garcia · V. Lin · J. Nguyen · D. Pimentel · R. Pozos · B. Reynante · E. Roy · E. Southerton · M. Suzara · A. Vishwanath
Stanford Learning Sciences and Technology Design PhD Program, Stanford University, Stanford, CA, USA
e-mail: pbiernacki@stanford.edu; mbigman@stanford.edu; kboles@stanford.edu; raquelcoelho@stanford.edu; vldocherty@stanford.edu; jorgeedu@stanford.edu; vronlin@stanford.edu; judynguyen@stanford.edu; dpimente@stanford.edu; rkozpos@stanford.edu; reynante@stanford.edu; ethanroy@stanford.edu; em.southerton@stanford.edu; msuzara@stanford.edu; vishwanath@stanford.edu

phones and the physical world—is becoming infrastructural to society’s cultural practices. We first sketch the surveillance state, enabled by pervasive sensors, cloud computing, and ubiquitous AI for pattern recognition and behavior prediction. We briefly characterize four surveillance technologies, all making headway into PreK-12 schools, universities, educational research, and technology design: (1) location tracking, (2) facial identification, (3) automated speech recognition, and (4) social media mining. We then pose primary issues educational research should investigate on cultural practices with these technologies for education and learning. We interweave three prioritized themes in our questioning: (1) how these technologies are shaping human development and learning; (2) current algorithmic biases and access inequities; and (3) the need for learners’ critical consciousness concerning their data privacy rights under threat and their agency in dealing with them efficaciously. We close with calls to action essential for guiding an educational future for our children and youth attuned to the risks of unreflective uses of these technologies and focused on demanding their transparent accountable uses for furthering our nation’s democratic society.

2 Surveillance State

Our networked society involves people exchanging personal details about themselves and what they’re doing for services and products on the web or apps (Ip 2018). Many accept the deals they’re offered in return for sharing insights about their behaviors, interests, and social lives. Pew Research Center research reveals a majority of Americans worry about these data being collected and used (Auxier et al. 2019). Zuboff (2019, 2020) calls this “surveillance capitalism,” a market-driven process commodifying personal data for profit-making, requiring capturing and producing these data through mass Internet surveillance. The concept arose after advertising companies foresaw using personal data to target consumers more specifically, and social media companies Facebook, Google, and Amazon exploited the insights to great fiscal rewards. “Analyzing massive data sets began as a way to reduce uncertainty by discovering the probabilities of future patterns in the behavior of people and systems” (Möllers et al. 2019). Turning humans into objects (data for monetization), not recognizing their agency as subjects, evokes warnings (Castells 1996: 371) of future networked inequalities where two profiles define humanity—the “interactive” (“using the Web’s full capacities”) and the “interacted” (limited to a “restricted number of prepackaged choices.”)

2.1 *Location Tracking*

Location tracking refers to processes of employing technologies that physically locate and electronically record and track movements of people or objects. This

technology is used in GPS navigation, locations specified on digital snapshots, and when people search for businesses nearby or more general information using common apps. Where you are, where you have been, and what information you are seeking at specific locations are among the most personal of facts. Technologies that enable location tracking are thus among the most privacy sensitive of all. Yet, “every minute of every day, everywhere on the planet, dozens of companies—largely unregulated, little scrutinized—are logging the movements of tens of millions of people with mobile phones and storing the information in gigantic data files” (Thompson and Warzel 2019). The Times Privacy Project obtained from a concerned source the vastest location sensitive data file ever reviewed by journalists, containing over 50 billion precise location pings from over 12 million Americans’ phones when moving through several major cities—Washington, New York, San Francisco, and Los Angeles—during several months in 2016–2017. Even so, they note, “this file represents just a small slice of what’s collected and sold every day by the location tracking industry—surveillance so omnipresent in our digital lives that it now seems impossible for anyone to avoid.” They note there is no federal law limiting collecting or selling these data.

On American campuses, college students are being watched, tracked, and managed by an accelerating nexus of technologies whose data are mined for colleges’ purposes (Belkin 2020). Beyond all the activity logging attendant to their uses of student IDs, video surveillance cameras record students’ faces, GPS tracks their movements, and their messages and photos are monitored on social media and email. Online courses and digital textbooks log their study habits minutely, and their pathways through campus buildings are recorded whether in class, dorm, cafe, library, or sporting events. Colleges say they’re using these surveillance data to keep students safe, engaged, and making progress, but we should ask how the reduced freedom to act without surveillance is shaping student agency and responsibility, since surveillance is a means of control and suppression. How commonplace is such surveillance on college campuses, and whether students can opt out or not, does it affect their sense of belonging and trust in higher education spaces (Jones et al. 2020)? Members of minoritized and racialized groups such as first-gen low-income (FLI) and underrepresented minority (URM) students may be especially vulnerable to such threats. Are surveillance-data-informed nudges for participation in study groups or visiting teaching assistants when a student is struggling effective? How are universities promoting on-campus critical literacies regarding ongoing surveillance of both students and faculty?

Similar questions extend to K-12 learners, as we ask how tracking technologies are used in K-12 settings to ensure the safety and progress of students but also potentially to violate students’ rights. First, given the increasing prevalence of computer learning in schools, education stakeholders should know what student location data is collected when they use computing hardware like Chromebooks, websites, and apps as education requirements. To participate in schooling, students must access disparate technology subsystems that are part of their education, such as Kahoot!, Edmodo, and Classrooms. How is their information and learning profile being protected or tracked and used to advance capitalistic rather than

student-centered interests? To what extent are K-12 students aware of and critically conscious about these tracking technologies and location data privacy? The default setting on many apps is “track” rather than “not track”—many apps thus track people without disclosure. Students’ sense of what these technologies imply about their learning environment may influence their personal agency, free movement, free expression of ideas and social affiliations, and feeling of belonging and trust in their school. Furthermore, deportation risks may lurk for undocumented students.

2.2 *Facial Identification Technologies (FITs)*

Facial recognition uses computer vision systems to identify specific human faces in photos/videos. Amazon, Microsoft, and smaller start-ups aggressively market FIT products to governments, law enforcement agencies, and private buyers (casinos and schools). Federal agencies ICE and FBI use face surveillance. Facebook and Google have their own proprietary algorithms. Apple and Google employ FIT for biometrically unlocking smartphones. The broader project of recognizing a person from photographs taken from live cams in public places like parks or streets was technically challenging for decades (Raviv 2020) but is now so advanced; it monitors millions of individuals in China (Economist 2018) and in US and UK urban settings (EFF 2020).

Facial recognition technology learns how to identify people by analyzing as many digital pictures as possible using “neural networks,” complex mathematical systems requiring vast amounts of data to build pattern recognition capabilities (Metz 2019). The *New York Times* has profiled the company Clearview AI selling access to facial recognition databases and tools to law enforcement agencies for presumed greater societal safety. Clearview violated service terms on diverse social media platforms to amass an enormous database of billions of images for facial recognition. The American Civil Liberties Union (ACLU) sued Clearview AI in its violation of state laws forbidding companies using residents’ face scans without consent. Beyond civil liberties issues, most commercial facial recognition systems exhibit biases, with false positives of African American and Asian faces 10–100 times more frequent than those of Caucasian faces (Buolamwini and Gebru 2018; Grother et al. 2019).

Governments use face surveillance technology to automatically *identify* an individual from a photo they have by scanning vast databases of labeled images (e.g., driver’s licenses) to find the faceprint matching the photo. For *tracking*, they use the technology once they know a person’s identity but want to track that person in real time and retroactively. Authorities use networks of surveillance cameras for tracking, and automation software builds records of everyone’s movements, habits, and associations. This is how China surveils ethnic minorities (Andersen 2020; Mitchell and Diamond 2018) and Russia monitors protests (Dixon 2021).

Finally, “emotion detection” technology claims to read emotions based on a person’s facial expression in photos and videos. Amazon and Microsoft advertise “emotion analysis” as one of their facial recognition products.

There is an increasing normalization of K-12 schools’ FIT use as thousands now employ video surveillance justified by the promise of protecting young people and checking attendance (Andrejevic and Selwyn 2019; Simonite and Barber 2019). Schools serving primarily students of color are more likely to rely on more intense surveillance measures than other schools (Nance 2016). The Electronic Frontier Foundation argues that schools must stop using these invasive technologies (Wang and Gebhart 2020).

Empirical research is needed on how PreK-12 learners experience FITs. Do algorithmic biases lead to (in)accuracies in FIT uses for recognizing youth and by gender, race, and ethnicity at different ages? We should examine what K-12 learners understand about FITs and how their parents engage with their presence in their child’s learning environments. How are decisions made to embed them in school environments, with what accountability to parents and local, state, and federal data privacy laws? Such information would help inform researchers and policymakers of what learners do and don’t know about FIT and privacy and how parents and educators might deal with their uses in education.

What are the generational differences in normalized acceptance of facial recognition technology as students adopt more social media platforms—Facebook, Instagram, Snapchat, and TikTok? Many parents upload their children’s pictures to social media since birth—how does growing up with social media affect the new generation’s perception of FITs and attitudes toward digital privacy?

We have many urgent questions: How to center children and their rights in this reality? What legal protections exist for PreK-12 learners regarding video surveillance FITs? How can educators ensure student data security? What do teachers, parents, school administrators, and learners understand about these safeguards? With what frequency is PreK-12 FIT used illicitly and challenged by teachers, parents, and children and with what consequences?

What curricula are needed to advance informed action by parents, teachers, and school leaders concerning FITs in children’s everyday lives? We need to understand what they understand about the risks of the FIT technology in its providing of “false positives”—when the technology reports it has identified a specific person but in fact has not—and of “false negatives,” when the technology missed out in finding a person who is in fact present in the video scene. How do stakeholders think about the troubles of data privacy risks and the greater error-prone nature of FIT algorithms for Blacks and Asians versus their purported security benefits? What concepts and models should students be learning to understand their technologically rich and privacy-poor world of FITs?

2.3 *Automated Speech Recognition*

Automated speech recognition is the capability of natural language processing (NLP) software to “understand” human language. Millions use voice recognition systems like Alexa, Siri, Assistant, or Cortana, which hear their voices, process their language, and act based on its query content—finding information online, playing music, making purchases, or controlling lighting/heating. NLP capabilities expand accessibility for people with visual impairments, but as always-on components of home and mobile communication infrastructure, they raise serious data privacy questions for their influences on human development and society.

How are children/adults using virtual personal assistants explicitly for learning purposes and to what effects? Are youth as automated-speech-recognition natives learning differently than youth in the past and with what consequences? There are two competing developmental hypotheses on prospects for and effects of conversational AI. The first is child psychologists arguing interactions with smart speakers are too superficial to teach children complex interactions like speech (Hirsh-Pasek, quoted in Kelly, 2018). The second hypothesis is more optimistic—Siri’s co-creator Tom Gruber (Markoff and Gruber 2019) suggests conversational AI has potential to teach students skills like reading as computers may outperform humans because of their ability to learn exponentially with pattern recognition. What skills and topics AI conversation systems will be good at “teaching” students lies unexamined. What are their benefits and limitations? Will children more likely share their goals, feelings, or progress relative to their learning with an AI humanoid tutor rather than human teachers? What roles might such tutors positively play in education for K-12 learners?

Smart AI speakers transform how people access and interact with information. Children are becoming accustomed to receiving answers immediately when asking Siri or Alexa questions. Greenfield (2017) argues making search frictionless could “short-circuit the process of reflection that stands between one’s recognition of a desire and its fulfillment via the market.” What will be the developmental consequences of the bots’ displacement of unmediated processes of trial and error and reflection for children to learn to solve problems on their own?

Biased algorithms are concerning: Koenecke et al. (2020) found racial disparities in automated speech recognition for Black and White speakers. Given such algorithmic biases in speech recognition, what inequities in technology access for supporting human activities will be perpetuated or even amplified, as for intersectional identities such as people of color who rely on speech recognition technology for learning accommodations in educational settings?

Childhood speech is tough. It is difficult to be accurate in speech recognition for the sentences young learners produce. A youngster’s breaks in speech, pauses, and filler words may decrease speech recognition accuracy. A child’s frustration when an agent doesn’t recognize their questions may well increase their cognitive load. Children’s frequent use of agents may also affect their language development—semantics, syntax, pragmatics, and prosody. It is also worth investigating how hybrid language learning environments of adult speech to children with adult speech to

agents influences children's language learning. For adult learners, we need to study how well speech recognition systems perform on different accents, speech styles from different cultures, and colloquial speech. For people of all ages, biases in recognition over time may condition learners to modify their speech style, accent, and behavior to match what makes the recognition system work. If true, this adaptation could create a stereotype threat-like effect where learners are forced to modify behavior to fit into a "normal" defined by the dominance of Western White speaker data used in these speech recognition systems.

Since today's conversational AI does not allow for the creative and flexible dialogues normally practiced by children and adults, youth may become less likely to question and explore ideas outside what these systems have programmed. As students more frequently use built-in speech recognition features of Google Docs to write their assignments by speaking, we wonder how their writing may be transformed. Literacy scholarship by Ong and McLuhan centers the ways "technologizing of the word" leads to interior transformations of consciousness, not only serving as exterior aids. It is worth asking how oppressive societies will control what kinds of answers are provided to questions doubting national authority. Differential access to such technologies by citizens of different nations may affect society and democracy at large.

Conversational agent futures will yield intelligent robotic assistants performing physical tasks to improve quality of life and increasing accessibility for many populations (e.g., the aged, students with special learning needs). The desire is that equity in access and utility of such tools can be promoted, while algorithmic biases are avoided as these assistants become ubiquitous in schools and homes which have diverse language practices.

Everyone needs to know about the safeguards that exist for data privacy when using these systems. Yet we know too little about how their users think about trade-offs between the convenient "frictionless" interactions which Weiser (1991) called "calm computing" and privacy-related drawbacks like ads, government surveillance, and hackers. Woven so effectively into the social fabric, the processes and effects of oppression become normalized, thus making it difficult to step outside of the system to discern how it operates (Adams et al. 2016). As speech recognition systems become embedded in smart toys for kids, research is needed into how children and parents navigate the ethical, trust, and safety issues in monitoring and recording interactions (McStay and Rosner 2021).

2.4 Social Media Mining

Social media are Internet-based apps for creating and exchanging user-generated content (Kaplan and Haenlein 2010 p. 61)—social networking, blogging, news aggregation, photo and video sharing, livecasting, social gaming, and instant messaging. Social media mining represents, analyzes, and extracts actionable patterns from social media data (Zafarani et al. 2014). Social media data mining analyzes user-generated content with rich social relationship information. Social media

dissolve boundaries between physical and digital worlds when social media mining researchers integrate social theories with computational methods to study how individuals (“social atoms”) interact and how communities (“social molecules”) form.

We begin by asking about critical consciousness: What do adults of different demographic profiles know about the powers corporations and governments have in making possible and regulating the conditions of their social media usage and associated data mining? What is the relationship between their social media behaviors and their beliefs about epistemic inequality, i.e., “unequal access to learning imposed by hidden mechanisms of information capture, production, analysis, and control” (Zuboff 2020b, p. 175)? How does this vary depending on sociocultural contexts and norms? With the emergence of legislation and regulations such as Europe’s (EU, 2018) GDPR and California’s Consumer Privacy Act (2018: CCPA), we need to know if adults are aware of their newly granted extensive data privacy rights (the rights to know, delete, opt out, and nondiscrimination). How are they appropriately informed of these rights in ways supporting their agency—are learning resources available not requiring reading impenetrable legalese?

Social media is now a huge part of adolescent students’ culture. How could the varied ways they learn, interact, and do things participating in online communities be leveraged for meeting the educational needs of all students? It is important to study how youth are weighing the pros of making social connections, expressing themselves and developing their online identity against the cons of being surveilled, profiled, and controlled. We ask what types of sense-making discussions youth have around ads or “news” on social media they are presented with based on their data-aggregated profiles and how many modify their privacy settings.

We know too little about the consequences arising for children’s social life and learning ecologies as social platforms connect preadolescent children from 4- to 13-year old to other children and families. Facebook’s Messenger Kids is a parent-controlled kids’ version for those under 13 who cannot have Facebook accounts but want to chat with friends and family. After violating a children’s privacy law in 2019, the FTC fined child popular TikTok \$5.7 million for allowing children under 13 to sign up without parental consent. TikTok made compliance changes allowing parents to set time limits, filter mature content, and disable direct messaging for kids’ accounts.

It is important that educational researchers and learning technology designers leverage these social media tools to further personalized learning while understanding the need to simultaneously continue pushing on the important questions about surveillance and privacy. It remains to be determined what parent education is needed for protecting children’s personal data and their critically informed social media uses. The California Consumer Privacy Act of 2018 requires children under 16 to provide opt-in consent for the sale of personal information, with parent or guardian consent for children under 13. The policy presumes that parents will prioritize the child’s privacy, but it is also frequently the case that parents themselves are uploading the child’s personal information to social media.

Learning researchers should examine what is being learned from experiences crafting and implementing K-12 curricula on social media use, Internet economics,

and data privacy rights, whether these are deployed in computer science education, civics, or humanities. We wonder how these lessons transform youth learning ecology and social media practices and influence their civic engagement and democratic participation. Questions of social media mining and the future of data use are intertwined with economic system design and regulation. Congress (thus, we the people) could play a role in regulating the social media industry and its deployment of AI technologies following ethical guidelines. Policy research and development needs to define the best options for sustainable, equitable, and democratic economic models for the Internet's social media moving forward and the associated legislation needed to achieve those models.

3 Call to Action

Free speech and assembly are rights guaranteed to US citizens under the First Amendment but are likely compromised when our networked world makes it difficult for people to avoid broadcasting spatiotemporal histories of where in the world they are with their faces, voices, spatial locations, and social media postings. We must ask about what consequences these constraints will have on human development and learning and what technology choices and political actions people should be making today to protect their privacy. All these questions indicate the need for greater attention, among educational researchers, policymakers, and education stakeholders, to vigilant enactment of the guidelines for ethical AI use in education, as discussed by Kousa and Niemi (this volume).

3.1 Research

We need an agenda of research priorities for educational research and learning technology design which addresses these vital issues. First considerations are to engage in the systematic empirical investigation of how pervasive in school buildings and campuses the uses of these four surveillance technologies have become.

We must ask about what effective strategies exist for overcoming the widespread sense of disempowerment and willingness to compromise by surrendering one's online data. We conjecture that adults and adolescents may be less complicit in the surveillance industry capturing so much personal information if they could see all the personal inferences that can be made from data captured from their behaviors. By analogy to the arguments for social distancing during the COVID-19 pandemic, might community-based action pushing back on surveillance capitalism be motivated because in doing so we would be caring for the most vulnerable in our communities? We need productive ways for adults and K-12 learners to acquire effective strategies to combat digital surveillance and to maintain their Internet privacy.

3.2 Policy and Law

Federal laws and guidelines protect pre-college students' data privacy rights. The Federal Trade Commission (FTC), the federal agency that enforces antitrust laws and protects consumers, has established COPPA (Children's Online Privacy Protection Act); it requires companies collecting online personal information from children under 13 to provide notice of their data collection and use practices and obtain verifiable parental consent. But schools can consent on behalf of parents to collection of student personal information—but only if such information is used for a school-authorized educational purpose and for no other commercial purpose. The FTC cites how edtech services should review the Family Educational Rights and Privacy Act (FERPA) and the Protection of Pupil Rights Amendment (PPRA)—laws administered by the US Department of Education's Student Privacy Policy Office (SPPO)—and any state laws protecting preK-12 students' privacy. The US Department of Education has provided new information on FERPA and virtual learning. In these regulations, we see the intersection of legal, commercial, and schooling issues.

Momentous impacts on society seem inevitable with the increasing embeddedness of facial recognition, voice recognition, location tracking, and social media mining. Society should regulate ethical guidelines for AI systems collecting and analyzing digital records of human faces, voices, spatial locations, and social relations, given the advent of AI-enhanced systems which identify us by those media and sell ads based on inferences predicting our behaviors. Arguably, these technical achievements have created benefits for consumers and citizens. But they've also raised difficult questions about personal rights and discriminative algorithmic biases. Protecting individual freedoms and maintaining a healthy democracy are priorities.

The United States has no laws or regulations governing the sale, acquisition, use, or misuse of face surveillance technology by the government. As of mid-2021, the few exceptions were municipal bans in California's San Francisco and Oakland, in Massachusetts' Boston, Somerville, Brookline, and Cambridge, and in Portland, Oregon. In 2020, Facebook agreed to pay \$550 million to Illinois settling a class action lawsuit over its FIT use. In 2019, as part of a \$5 billion privacy violations FTC settlement, Facebook agreed to "clear and conspicuous notice" about its face matching software and to get additional permission from people before using it for new purposes. We must also seek legal protections against discriminative uses of inaccurate FIT and speech recognition technologies for minoritized groups.

3.3 Practice

The practice of education by teachers, school leaders, and parents must seek to protect the digital privacy rights of children as they participate in the learning environments of their daily lives. Schools need to prepare school personnel, so they learn about the data sharing that they are (probably unknowingly) asking students

and parents to participate in. Perhaps digital privacy health checkups should be a regular educational service for both adults and children.

Another issue is how academic and industrial researchers deal with the ethics of developing face recognition technologies, location tracking, automated speech recognition, and social media mining that can have widespread detrimental effects. Ethics and privacy considerations are commonly afterthoughts in technology development and, even then, described as a nuisance and as stifling innovation. However difficult to develop, foresight on future consequences of a technology capability should be built into training and R&D processes with transparency and accountability.

4 Conclusion

In this chapter, we introduced the capabilities of four core surveillance technologies, each becoming interwoven into the fabrics of universities and preK-12 schools: location tracking, facial identification, automated speech recognition, and social media mining. As such ubiquitous AI is becoming infrastructural to cultural practices, embodied in cloud computing web services and in sensors in phones and the physical world, creating a surveillance society, it is essential for education stakeholders, from policymakers to school leaders, teachers, parents, legislators, regulators, and industry itself, to tackle together the ethical issues of AI in education which these surveillance technologies foreground. We sketched challenges around how these technologies may be reshaping human development, risks of algorithmic biases and access inequities, and the need for learners' critical consciousness concerning their data privacy.

Although ethical guidelines for education as a context of AI application are mainly lacking (Holmes et al. 2021), we may find utility for education's issues with these four AI-enabled surveillance technologies in the five principles for ethical use of AI synthesized by Morley et al. (2020) and discussed in Kousa and Niemi's chapter (this volume). Recall that these five complementary aspirational principles are beneficence, non-maleficence, autonomy, justice, and explicability.

AI *beneficence* means useful, reliable technology generously supporting the diversity of human well-being. AI *non-maleficence* would guarantee data security, accuracy, reliability, reproducibility, quality, and integrity. AI with *human autonomy* has humans free to make decisions and choices regarding AI use. AI with *justice* operates in a fair and transparent manner, not obstructing democracy or harming society. *Explicable* AI enables clear explanation and interpretation of system functioning for humans and corresponding accountability and responsibility.

We are hopeful that with concerted collaboration of government, industry, and the public sector on these issues, the continued advances in artificial intelligence will come to be a powerful aide to more equitable and just educational systems and an ingredient to engaging, innovative learning environments that will serve the needs of all our diverse learners and educators.

References

- Adams, M., Bell, L. A., Goodman, D. J., & Joshi, K. (2016) (Eds.), *Teaching for diversity and social justice* (3rd ed.). New York: Routledge.
- Andersen, R. (2020, September). The Panopticon is already here. *The Atlantic Monthly*.
- Andrejevic, M., & Selwyn, N. (2019). Facial recognition technology in schools: critical questions and concerns. *Learning, Media and Technology*, 1-14.
- Auxier, B., Rainie, L., Anderson, M., Perrin, A., Kumar, M., & Turner, E. (2019, November). *Americans and privacy: Concerned, confused and feeling lack of control over their personal information*. Pew Research Center.
- Belkin, Douglas (2020, March 5). No Place to Hide: Colleges Track Students, Everywhere. *Wall Street Journal*.
- Buolamwini, J., & Gebru, T. (2018, January). Gender shades: Intersectional accuracy disparities in commercial gender classification. In *Proceedings of Machine Learning Research*, 81: 1-15.
- California Consumer Privacy Act of 2018 (CCPA): Fact Sheet.
- Castells, M. (1996). *The information age* (Vol. 98). Oxford: Blackwell Publishers.
- Dixon, R. (2021, April 17). Russia's surveillance state still doesn't match China. But Putin is racing to catch up. *The Washington Post*.
- Economist (2018, October 24). China: facial recognition and state control
- EFF (2020). Face recognition: Street level surveillance
- EU (2018, May 25). EU GDPR (General Data Protection Regulation).
- Greenfield, A. (2017). *Radical technologies: The design of everyday life*. Verso Books.
- Grother, P., Ngan, M., & Hanaoka, K. (2019). Face Recognition Vendor Test (FRVT) Part 3: Demographic Effects. *Technical Report NISTIR 8280, NIST*.
- Holmes, W., Porayska-Pomsta, K., Holstein, K., Sutherland, E., Baker, T., Shum, S.B., Santos, O.C., Rodrigo, M.T., Cukurova, M., Bittencourt, I.I. and Koedinger, K.R., (2021). Ethics of AI in education: Towards a community-wide framework. *International Journal of Artificial Intelligence in Education*, pp.1-23.
- Ip, Greg (2018, November). The Unintended Consequences of the 'Free' Internet. *Wall Street Journal*.
- Jones, K. M., Asher, A., Goben, A., Perry, M. R., Salo, D., Briney, K. A., & Robertshaw, M. B. (2020). "We're being tracked at all times": Student perspectives of their privacy in relation to learning analytics in higher education. *Journal of the Association for Information Science and Technology*, 71(9), 1044-1059.
- Kaplan, A. M., & Haenlein, M. (2010). Users of the world, unite! The challenges and opportunities of Social Media. *Business horizons*, 53(1), 59-68.
- Kelly, S.M. (2018, October 17). Growing up with Alexa: A child's relationship with Amazon's voice assistant. CNN Business. (<https://edition.cnn.com/2018/10/16/tech/alexa-child-development/index.html>)
- Koenecke, A., Nam, A., Lake, E., Nudell, J., Quartey, M., Mengesha, Z., Toups, C., Rickford, J.R., Jurafsky, D., & Goel, S. (2020). Racial disparities in automated speech recognition. *Proceedings of the National Academy of Sciences*, 117(14), 7684-7689.
- Markoff, John & Gruber, Tom (2019, January 17). A Conversation about Conversational AI | Tom Gruber
- McStay, A., & Rosner, G. (2021). Emotional artificial intelligence in children's toys and devices: Ethics, governance and practical remedies. *Big Data & Society*, 8(1), 2053951721994877.
- Metz, C. (2019, July 13). Facial recognition tech is growing stronger, thanks to your face. *New York Times*.
- Mitchell, A., & Diamond, L. (2018, Feb 2). China's surveillance state should scare everyone. *The Atlantic Monthly*.
- Morley, J., Floridi, L., Kinsey, L., Elhalal, A. (2020). From what to how: An initial review of publicly available AI ethics tools, methods, and research to translate principles into practices. *Science and Engineering Ethics*, 26:2141-2168. <https://doi.org/10.1007/s11948-019-00165-5>

- Möllers, N., Wood, D.M. & Lyon, D. (2019-03-31). Surveillance capitalism: An Interview with Shoshana Zuboff. *Surveillance & Society*, 17 (1/2): 257–266.
- Nance, J. P. (2016). Student surveillance, racial inequalities, and implicit racial bias. *Emory LJ*, 66, 765.
- Raviv, S. (2020, January 21). The secret history of facial recognition. *Wired*.
- Simonite, T., & Barber, G. (2019). The delicate ethics of using facial recognition in schools. *Wired*, Business 10.17.2019
- Thompson, Stuart A., Warzel, Charlie (2019, December 19). Twelve million phones, one dataset, zero privacy (Links to an external site). *New York Times*. Privacy Project.
- Wang, Mona, & Gebhart, Gennie (2020, February 27). Schools are pushing the boundaries of surveillance technologies (Links to an external site.). *Electronic Frontiers Foundation*.
- Weiser, M. (1991). The computer for the 21st Century. *Scientific American*, 265(3), 94-105.
- Zafarani, R., Abbasi, M. A., & Liu, H. (2014). *Social media mining: an introduction*. Cambridge University Press.
- Zuboff, S. (2019). *The age of surveillance capitalism: The fight for a human future at the new frontier of power*. New York: Public Affairs.
- Zuboff, S. (2020, January 24). The known unknown: Surveillance capitalists control the science and the scientists, the secrets and the truth. *New York Times, Sunday Review*, pp. 1, 6, 7.
- Zuboff, S. (2020b). Caveat Usor: Surveillance capitalism as epistemic inequality, in Kevin Werbach, Ed., *After the Digital Tornado*, Cambridge: Cambridge University Press.

Open Access This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.



Reflections on the Contributions and Future Scenarios in AI-Based Learning



Roy D. Pea, Yu Lu, and Hannele Niemi

Contents

1 Where Are We Now with AI?	331
2 AI Contributions to Different Levels of Education Systems	333
2.1 K-12 Tutoring Systems and Other Adaptive Learning Technologies	333
2.2 Beyond K-12 Disciplinary Curriculum: Whole Child AI Technologies	334
2.3 Higher Education and Lifelong Learning	335
2.4 Enabling Media for the Learning Ecosystem.....	336
3 Roles of AI for Enhancing the Processes and Practices of Educational Research	337
4 Advancing the Learning of AI.....	338
5 Ethical Dimensions of AI Integration into Human Learning Environments and Socio-Technical Systems for Education.....	339
References	341

1 Where Are We Now with AI?

In our concluding chapter, after briefly considering AI's immense presence in the emerging information infrastructure of global societies and its importance for both education and education research, we reflect on the contributions to research, technology, and theory provided by the chapters of our volume. We characterize the vectors of development and the critical issues identified as priorities for the research

R. D. Pea (✉)

Graduate School of Education, Stanford University, Stanford, CA, USA

e-mail: roypea@stanford.edu

Y. Lu

Advanced Innovation Center for Future Education, Faculty of Education, Beijing Normal University, Beijing, China

e-mail: luyu@bnu.edu.cn

H. Niemi

Faculty of Educational Sciences, University of Helsinki, Helsinki, Finland

e-mail: hannele.niemi@helsinki.fi

ahead. This chapter also provides scenarios of the future development and changes when AI will be applied in learning and education.

Artificial intelligence, or simply AI, has become one of the most pervasively adopted technologies in history. It is now integrated into billions of smartphones for services as diverse as speech recognition agents like Siri and Alexa and recommendation services for music, movies, books, retail purchasing, and route mapping for driving. It is possible that AI and its related technologies could become highly consequential for the future of learning, teaching, and educational systems more broadly. Our scholarly community in education research, and all of education's stakeholders, should critically consider how to best develop and use AI in education so that it will be equitable, ethical, and effective while guarding against data and design risks and harms.

In 1956, Stanford's John McCarthy offered one of the first definitions of AI: "The study [of artificial intelligence] is to proceed on the basis of the conjecture that every aspect of learning or any other feature of intelligence can in principle be so precisely described that a machine can be made to simulate it" (Russel and Norvig 2010).

With its advances in 65 years, we find Accenture's contemporary definition useful: "AI is a constellation of many different technologies working together to enable machines to sense, comprehend, act, and learn with human-like levels of intelligence."

Integral to this AI terrain are machine learning and natural language processing. Machine learning is a type of AI enabling systems to learn patterns from data, make predictions, and then improve future experience through applying the discovered patterns to situations absent in their initial design (Popenici and Kerr 2017). When you get product recommendations in an online retail shopping site, these suggestions are driven by machine learning, as the AI is continuously improving at figuring out what you might buy. Useful as they are, these forms of AI are called *Narrow AI*, toolled to performing a single task or closely related tasks.

General AI, as in sci-fi films, where sentient machines emulate human intelligence and think strategically, handling a broad range of complex tasks, is not yet reality. Although AI computing works at exceptional speed and scale, human-machine collaboration is crucial as humans provide guidance by labelling data from which AI machines can learn. So, AI thus far augments human capabilities, rather than replacing them.

We anticipate that recognizing the possibilities and limits of AI technologies will become more an everyday topic of conversation as people seek to make sense of the stunning digital transformations they are experiencing and quest to fulfill their hopes to fully participate in, benefit from, and adapt to the new occupations and skill needs that will emerge.

In the spirit of supporting that quest, we now ask: How does AI relate to educational systems, teaching and learning processes, and educational research? The essential purpose is to explore how AI can serve human purposes in promoting learning and enhancing education research.

To begin, we note that in 2020 in Silicon Valley, the US non-profit organization Digital Promise convened a panel of 22 experts in AI and in learning for several days to consider these broad questions (Roschelle et al. 2020):

- What do educational leaders need to know about AI in support of student learning?
- What do researchers need to tackle beyond the ordinary to generate knowledge for shaping AI in learning for the good?

Their synthesis report suggests three layers for framing AI's meaning for educators. First, AI can be viewed as a "computational intelligence" for contributing an additional resource to an educator's abilities and strengths in tackling educational challenges. Second, AI brings specific and exciting new capabilities to computing, including sensing, recognizing patterns, representing knowledge, making and acting on plans, and supporting naturalistic interactions with people. These capabilities can be engineered into solutions to support learners with varied strengths and needs, such as allowing students to use handwriting, gestures, or speech as input modalities in addition to keyboard and mouse. Third, AI may be used as a tool kit to enable imagining, studying, and discussing future learning scenarios that don't exist today. We find our authors making contributions to the AI in education literature in each of these layers.

Now let us consider ways to frame the full panoply of contributions from the chapters of our volume. Seven categories provide perspectives to reflections. Four of them are connected to different levels of the educational system, others opening scenarios to research on education and learning with AI, and finally the last category is devoted to ethical challenges of AI in education and learning. These reflections will help us sort through the forest of new work represented in this volume.

2 AI Contributions to Different Levels of Education Systems

2.1 *K-12 Tutoring Systems and Other Adaptive Learning Technologies*

As we indicate below, one preliminary and central idea we wish to communicate is that AI in education is about so much more than "ed tech" applications, such as intelligent tutoring systems (ITS) and adaptive learning technologies, although developments in AI are still contributing to this vision (see chapters by Niu et al. "[Multiple Users' Experiences of an AI-Aided Educational Platform for Teaching and Learning](#)" and Chen et al. "[Learning Career Knowledge: Can AI Simulation and Machine Learning Improve Career Plans and Educational Expectations?](#)"). Niu et al.'s chapter on how their AI-aided educational smart learning partner platform provides intelligent services to support students' learning contributes a multiuser perspectives account of the experiences of students, teachers, and school managers as they employ a system in which learners constantly receive individualized learning assessments and recommended improvements, teachers can attune their pedagogical strategies and actions according to students' needs, and school management can more informedly support teachers' teaching and students' learning.

One chapter provides a bridge between K-12 education and career development. Chen et al.'s chapter on “[Learning Career Knowledge: Can AI Simulation and Machine Learning Improve Career Plans and Educational Expectations?](#)” details their approach to using AI capabilities to support youth career selection as they face job future uncertainties with automation's advancements. They investigate how machine learning applications can help solve the problem of enabling youth to align their individual career goals with specific employment opportunities and know what capabilities and certifications specific jobs either demand or require. They describe how these applications have been implemented with tasks and goals to test players' capacity, skills, and interests in selecting future occupations using simulated game-based scenarios that yield a player's computer-generated characteristics. They share the machine learning decision tree algorithms derived to map out all the possible outcomes of job selections and to then narrow individual players' opportunity choices given their current gameplay status. It is impressive how such gameplay can minimize risks and provide strategic advantages for young people with limited occupational knowledge.

The described examples provide signals that *AI will change future curricula, assessment methods, student counseling, and teachers' work*. It demands radical changes in the whole educational ecosystem and supporting teachers to move toward new kinds of pedagogical orchestration in classrooms and beyond when expanding learning environments with AI.

2.2 Beyond K-12 Disciplinary Curriculum: Whole Child AI Technologies

AI is also being applied to what we might call “whole child education”, more than the standard curriculum and its learning standards. Increasingly, educational systems are taking more of a whole child development approach to education in which creating safe and supportive learning environments for equitably preparing each student to reach their full potential is a key goal. Such supportive environments aim to promote wellness and resilience for everyone participating in the school community, emphasizing not only academic but social-emotional outcomes such as self-regulation, stress management, and a sense of belonging since they affect productive engagement in learning.

Several chapters address, respectively, students' broadly considered well-being and, more narrowly, their problem behaviors. Students' well-being is critical as it marks their positive development in school life and ensures their future growth. Tang et al. chapter “[Assessing and Tracking Students' Wellbeing Through an Automated Scoring System: School Day Wellbeing Model](#)” introduce an automated scoring

well-being system—School Day Well-Being Model—featured as dynamic and real-time in giving immediate feedback at multiple organization layers (person, class, school). Task performance and emotion regulation skills were the most consistent skills to promote psychological well-being, academic well-being, and health-related outcomes. Penghe et al.’s chapter “[An AI-Powered Teacher Assistant for Student Problem Behavior Diagnosis](#)” proposes an AI-powered assistant for solving student problem behaviors in school, as defined by undesirable behavior compared with social norms. Interventions are based on automatically diagnosed unmet needs of students. They build a domain knowledge graph summarizing all relevant factors of diagnosed unmet student needs to guide the system, adopting reinforcement learning to learn dialogue policy on this topic and to implement the dialogue system for addressing student behavioral problems.

Socio-emotional factors are decisive for students’ successful learning (e.g., Durlak [2015](#)). AI and its capacity to bring multimodal data into learning environment designs and interventions will open totally new opportunities *to understand student’s behaviors and their needs for learning and well-being*. However, we can also see that the mere data and even its effective interactive systems do not necessarily help without *human scaffolding and interaction* (Pea [2004](#)). Human behavior has pervasive social foundations, and we need the integration of AI-based information and human users.

2.3 Higher Education and Lifelong Learning

Four chapters tackle the uses of AI in learning environments for college-age students and beyond, encompassing nursing education, VR training of hard procedural skills in industry, stress during simulation-based learning, and self-learning and emotional support through cognitive mirroring with intelligent social agents (ISA). Koivisto et al. “[Learning Clinical Reasoning Through Gaming in Nursing Education: Future Scenarios of Game Metrics and Artificial Intelligence](#)” report studies of nursing students using computer-based simulation games for learning clinical reasoning (CR) skills in an authentic 3-D hospital environment with nine scenarios based on different clinical situations in nursing care as they learn essential skills for ensuring patient safety and high-quality care as they assess patients’ clinical condition systematically by interviewing, observing, and measuring patient’s vital signs. Game metrics calculated during gameplay are used to evaluate nursing students’ CR skills and to target needs for improvements.

Korhonen et al.’s chapter “[Training Hard Skills in Virtual Reality: Developing a Theoretical Framework for AI-Based Immersive Learning](#)” explores learning with an immersive virtual reality-based hard-skills training guided by an AI

tutor software agent. They observe how such environments, supported by sufficiently advanced tutoring software, may facilitate asynchronous, embodied learning approaches for learning hard, procedural skills in industrial settings. They unpack the mismatch between the philosophy of cognition underpinning intelligent tutoring system (ITS) software and emergent issues for the learner's epistemology in a virtual world and its attendant shortcomings for learners' experiences in the VR environments where they are learning. To counteract this mismatch of philosophy of cognition and technology-augmented learning environment design, they propose improved pedagogical approaches in employing the philosophies of embodied, embedded, enacted, and extended (4e) cognition as the underpinning for VR-native pedagogical principles. Ruokamo et al.'s chapter "[AI-Supported Simulation-Based Learning: Learners' Emotional Experiences and Self-Regulation in Challenging Situations](#)" explores professionals' learning experiences and their stress level during simulation-based learning, considered from physiological, emotional, motivational, and cognitive perspectives to identify key factors increasing and inhibiting their learning. In "[Learning from Intelligent Social Agents as Social and Intellectual Mirrors](#)" Maples et al. report on a mixed-method study report on a mixed-method study exploring relationships between user loneliness, use motivations, use patterns, and user outcomes for 27 adult users of Replika, a best-in-class "intelligent social agent" (ISA) sufficiently anthropomorphized to pass Turing tests in short exchanges. Their data indicate these users were lonely or experiencing a time of change and distress and they used Replika for its availability, friendship, therapy, and personal learning. For many, Replika provided critical emotional support; for some, belief in Replika's intelligence led to a deeper cognitive proximity and increasingly profound engagement as they identified Replika as a human, a friend, and even an "extension of themselves".

AI will change the landscape of life-long learning. The borders of formal and informal learning will be broken. *AI will be the essential tool in learning of skills and competences in working life as well as in personal learning environments and contexts.* So far, games and simulations have been an essential tool, but in the future, much training will happen in *virtual reality, increasingly called the metaverse* (Sparkes 2021). This also makes collaboration and social elements possible in skills and competence learning. As the future scenario, we may expect *radical changes in adult education and job reskilling*.

2.4 Enabling Media for the Learning Ecosystem

Two chapters are devoted to explicating how AI can provide advances in the core functionalities of the establishment of media for the learning ecosystem:

one is devoted to intelligent e-textbooks and one to deep learning in automatic math word problem solvers (MWPs). Jiang et al.'s chapter “[Recent Advances in Intelligent Textbooks for Better Learning](#)” investigates the history and vital topic of how e-textbook platforms could promote learning. If we could understand how people interact with and read e-textbooks, we would have more guidance for providing intelligent learning support to learners in the design of e-textbooks. They review key intelligent technologies used in intelligent textbooks—student modeling and domain modeling technologies. Student modeling has three aspects: learner knowledge state modeling, learner learning behavior modeling, and learner psychological characteristic modeling. They introduce popular intelligent textbook authoring platforms used for creating intelligent textbooks. Zhang's chapter “[Deep Learning in Automatic Math Word Problem Solvers](#)” provides a synoptic account of developments in automatic MWPs, from the 1960s to the uses of deep learning algorithms today as they seek to solve the challenging problem of parsing the human-readable word problems into machine-understandable logical expressions. As systems advance the intelligence level of AI agents in terms of natural language understanding and automatic reasoning, they promise intelligent support in education environments for learners' developments in mathematical word problem-solving competencies.

Technological advances have made it possible to overcome many earlier barriers in how to support human learning. The future perspectives require that we *understand more about the relationship of human and machine learning*. With AI, we have two learners: a human and machine. This interaction needs new understanding of how this relationship can be supportive to different kinds of human learners and extraordinarily diverse learning situations. Success in this enterprise requires continuous collaboration between experts of computing sciences and learning sciences.

3 Roles of AI for Enhancing the Processes and Practices of Educational Research

Three chapters report how AI is contributing to facilitation of educational research. Marcelo Worsley characterizes different facets for how multimodal learning analytics employs AI for measuring student performances during complex learning tasks. He highlights how contemporary authentic and engaging learning environments transcend the traditional teacher-centric classroom context, incorporating types of learning experiences that are embodied, project-based, inquiry-driven, collaborative, and open-ended. He examines AI-based tools and sensing technologies that can help researchers and practitioners navigate and enact these novel approaches to learning

with new analytic techniques and interfaces for helping researchers collect and analyze different types of multimodal data across contexts, while also providing a meaningful lens for student reflection and inquiry.

Vivitsou's chapter "[Perspectives and Metaphors of Learning: A Commentary on James Lester's Narrative-Centered AI-Based Environments](#)" centers on James Lester's AI in education keynote address and associated interview, to discuss perspectives on narrative-centered learning and metaphors of AI-based learning environments, such as Crystal Island, an AI-based game for K-12 students learning science. She employs Ricoeur's narrative theory and metaphor theory to examine the role of characters and the narrative plot in relation to Lester's visualization of the future of learning with AI-based technologies, revealing new roles in AI-rich game-based learning such as drama manager. She also examines the importance of dynamic agency metaphors in AI for advancing learning environment design. With the intention of supporting the improvement of classroom teaching quality, Yu & Sun's chapter "[Analysis and Improvement of Classroom Teaching Based on Artificial Intelligence](#)" depicts research and technology which seeks to transcend traditional labor-intensive classroom teaching event analysis methods by using their teaching event sampling analysis framework (TESTII), which employs computer vision, natural language processing, and other emerging AI technologies to perform classroom teaching event analysis for improving educational practices.

When AI comes to education and learning settings, the typical designed structures of lessons and learning environments will be changed. We need *new concepts for understanding our life-long, life-wide, and life-deep learning environments* (Bell and Banks 2012), and how analytic techniques and research methods must also be reconceived and re-designed with AI-based tools and learning environments.

4 Advancing the Learning of AI

Several chapters are devoted to the basic research problem of engineering AI to learn more productively, in hopes that such advances could improve human learning in educational systems as well. Haber considers how to build AI that learns via curiosity and interactions like humans, and Zhang asks how advances in deep learning with automatic math word problem solvers can represent progress toward the automatic reasoning of general AI. Haber's chapter "[Curiosity and Interactive Learning in Artificial Systems](#)" introduces readers to results from AI's deep reinforcement learning that aspire to replicate the processes and outcomes of human interactive learning, sparked by curiosity, seeking novelty and information, and social engagement. He asks how might we engineer an artificial, autonomous

agent that can flexibly interact with its environment, and other agents within it, to learn as humans do. He argues that if this AI engineering program makes progress, it may shape the future of education by providing fine-grained computational models of learning and even enabling *in silico* testing of learning interventions, from early childhood through K-12 education. Zhang's chapter "[Deep Learning in Automatic Math Word Problem Solvers](#)" provides a synoptic account of the technical history of automatic math word problem solvers (MWPs), from the 1960s to the uses of deep learning algorithms today that shrink the semantic gap between what humans can read and what machines can understand. MWP seek to solve the challenging problem of parsing human-readable word problems into machine-understandable logical expressions. Different MWP architectures have been good test beds for appraising the intelligence level of agents in terms of natural language understanding and automatic reasoning, and their comparative performances on public benchmark datasets illuminate advances toward the automatic reasoning of general AI.

While even the latest AI techniques still find it challenging to simulate human learning and fully understand the semantics of human language, significant progress has been made in the fields of machine learning and natural language processing in recent years (Deng and Liu 2018). We believe that the learning *capabilities of AI will be more powerful* and effective in the near future, by leveraging the *advancements of neuroscience* that reveal how our human brain thinks, remembers, and learns (Savage 2019; Ullman 2019).

5 Ethical Dimensions of AI Integration into Human Learning Environments and Socio-Technical Systems for Education

Two chapters delve into national policy (comparing Finland and China) and stakeholder perspectives on AI in education (education technology industry and its educational system clients). Wei & Niemi's chapter "[Ethical Guidelines for Artificial Intelligence-Based Learning: A Transnational Study Between China and Finland](#)" provides an AI policy analysis comparing programmatic policy documents developed by the Finnish and Chinese governments for promoting the development of AI-based learning in society. Five themes emerged: (1) the potential of AI for reshaping basic education and school quality; (2) emphasizing the importance of AI in the workforce and employment; (3) connecting AI with human development and students' wellbeing; (4) promoting teachers' AI literacy in digitalized times; and (5) AI for lifelong learning reform in a civil society. Yet promoting ethical guidelines for AI in learning is barely discussed at the policy level. Instead, policy documents

discuss general ethical themes, not specifying ethical challenges for educational environments. Their chapter further analyzes detailed ethical challenges within the five themes when AI-based tools are used in educational environments and critically reflects on needed ethical guidelines when AI is applied in education.

Kousa & Niemi's chapter "[Artificial Intelligence Ethics from the Perspective of Educational Technology Companies and Schools](#)" analyzes and reflects on the perspectives of multiple parties—companies producing AI-based tools and services and their users in schools and workplaces—concerning ethical opportunities and challenges which AI is establishing for learning in schools and working life. Corporate perspectives consider ethical challenges to be related to regulations, equality and accessibility, machine learning, and society. From the school users' perspectives, the critical questions are: Who has the power to decide which educational services the school can use? Who is responsible for ethical issues (such as student privacy) of those services? Who will ensure that AI-based services and tools are equally accessible to and effective for all in supporting teaching and learning? The authors argue that continuing dialogue between producers and consumers is essential and that national and international guidance is needed on how to engage in ethically sustainable action. The aim is to increase common AI knowledge through education to understand its opportunities and challenges and keep up with our rapidly evolving society.

It is an important shortcoming that, despite increasing attention on privacy and ethics in educational technology (Henein et al. [2020](#), p. 3), there remains a “widespread lack of transparency and inconsistent privacy and security practices for products intended for children and students.” To advance educational research at scale, it is crucial to provide methods and processes for implementing privacy-preserving learning analytics globally ([Every Learner Everywhere 2020](#); Joksimović et al. [2022](#)).

Meta concerns for the ethics of AI in education are provided by Cowley et al.'s chapter "[Artificial Intelligence in Education as a Rawlsian Massively Multiplayer Game: A Thought Experiment on AI Ethics](#)" They provide a thought experiment for conceptualizing the possible benefits and risks to be revealed as AI is integrated into education. Actors with different stakes (humans, institutions, AI agents, and algorithms) all conform to the definition of a player—a role designed to maximize protection and benefit for human players. AI models that understand the game space provide an API for typical algorithms, e.g., deep learning neural nets or reinforcement learning agents, to interact with the game space. The thought experiment surfaces socio-cognitive-technological questions that must be discussed, such as benefits of using AI-based tools for supporting different learners, yet possible risks of algorithmic manipulation, or hidden algorithmic discrimination. The more we reflect on it, the clearer it becomes that the ethics of AI in education is a keystone issue which will ramify throughout future inquiries into the future of AI-augmented learning.

Finally, Pea et al.'s “Four Surveillance Technologies Creating Challenges for Education” introduces the capabilities of four core surveillance technologies now being embraced by universities and preK-12 schools: location tracking, facial

identification, automated speech recognition, and social media mining. The chapter articulates challenges in how these technologies may be reshaping human development, risks of algorithmic biases and access inequities, and the need for learners' critical consciousness concerning their data privacy. The chapter expresses hope that government, industry, and public sector collaboration on these issues can make more likely that continued advances in artificial intelligence will become a powerful aide to more equitable and just educational systems and an ingredient to engaging, innovative learning environments that will serve the needs of all our diverse learners and educators.

The ethical questions are burning when AI is applied in education and learning. Ethical demands concern the whole society, developers, and providers of new tools, environments, and services. It also concerns all users. Even though we have many national and international ethical guidelines appearing, many issues are still open and new problems are continually being discovered. Perhaps the biggest question is how users can trust that their privacy is not violated. AI has become ubiquitous, it is part of everyday life, and it will be a common tool in education and learning. For understanding what AI means in our life, we need *a new civic skill*. Support for this should be part of school curricula and easily available in society. AI users need basic knowledge about AI, its features and applications, and what are ethical regulations needed for its safe use. All people should also have information about what are their rights and what are the procedures to follow if there are misuses of their privacy with AI. Users will need this kind of knowledge in their school years and widely throughout their life. AI will be a powerful tool in our future, but we must remember that human beings have the ultimate responsibility when developing and using AI.

References

- Bell, P., & Banks, J. (2012). Life-long, life-wide, and life-deep learning. *Encyclopedia of Diversity in Education*, 4, 1395–1396.
- Deng, L., & Liu, Y. (Eds.). (2018). *Deep learning in natural language processing*. Springer.
- Durlak, J. A. (Ed.). (2015). *Handbook of social and emotional learning: Research and practice*. Guilford Publications.
- Every Learner Everywhere. (2020). *Guiding principles and strategies for learning analytics implementation*. Tyton Partners.
- Henein, N., Willemse, B., & Woo, B. (2020). The state of privacy and personal data protection, 2020–2022. Gartner Report.
- Joksimović, S., Marshall, R., Rakotoarivelo, T., Ladjal, D., Zhan, C., & Pardo, A. (2022). Privacy-Driven Learning Analytics. In *Manage Your Own Learning Analytics* (pp. 1–22). Springer, Cham.

- Pea, R. D. (2004). The social and technological dimensions of scaffolding and related theoretical concepts for learning, education, and human activity. *The Journal of the Learning Sciences*, 13(3), 423–451.
- Popenici, S., & Kerr, S. (2017). Exploring the impact of artificial intelligence on teaching and learning in higher education. *Research and Practice in Technology Enhanced Learning*, 12(1), 1–13.
- Roschelle, J., Lester, J. & Fusco, J. (Eds.) (2020). *AI and the future of learning: Expert panel report*. Digital Promise. <https://circls.org/reports/ai-report>.
- Russel, S., & Norvig, P. (2010). *Artificial intelligence - a modern approach*. New Jersey: Pearson Education.
- Savage, N. (2019). How AI and neuroscience drive each other forwards. *Nature*, 571(7766), S15–S15.
- Sparkes, M. (2021, August 21). What is a metaverse. *New Scientist*, 251(3348), 18.
- Tomasello, M. (1999). The human adaptation for culture. *Annual review of anthropology*, 28(1), 509–529.
- Ullman, S. (2019). Using neuroscience to develop artificial intelligence. *Science*, 363(6428), 692–693.

Open Access This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.



Index

A

- Adaptive learning systems, 6, 219, 228, 266, 333–334
Agency, vi, 3, 9, 127, 130–134, 216, 219, 268, 269, 277, 293, 318–320, 324, 326, 338
AI-aided educational platform, 10, 215–230, 333
AI ethical issues, 1, 3, 12, 13, 268, 270, 288, 327
AI tutor, 10, 177, 180, 181, 186–188, 196–199, 203–207, 210, 216, 288, 336
Algorithmic biases, 33, 318, 321–323, 341
Alignment, 21, 138, 139, 141, 145, 146, 148–154, 242, 286
Artificial Intelligence (AI), 1, 21, 37, 56, 80, 92, 105, 125, 138, 160, 176, 195, 215, 233, 248, 265, 283, 298, 317, 331
Artificial systems, 37–52, 337
Assistant, 5, 8, 80, 91–102, 219, 302, 319, 322, 323, 335
Automated scoring system, 7, 334–335
Automated speech recognition, 13, 318, 322–323, 327, 341
Automatic reasoning, 337, 338

C

- Career knowledge, 9, 137–155, 333
Challenging situations, 10, 160, 175–188, 335, 336
Chatbot, 2, 5, 58, 74, 162, 250
Classroom teaching analysis, 8, 105–109, 111, 113, 118
Clinical reasoning, 9, 159–170, 335

Crystal Island, 9, 126, 127, 129–134, 337

Curiosity, 7, 37–52, 80, 337

D

- Decision trees, 138, 145–153, 253, 334
Deep learning, 11, 39, 40, 74, 98, 113, 116, 160, 178, 196, 233–244, 266, 276, 277, 292, 298, 300, 337–340
Dialogue system, 8, 92, 95, 96, 335
Digital therapies, 75
Displacement hypothesis, 75

E

- EdTech companies, 283–294
Education, 1, 20, 38, 56, 94, 105, 126, 138, 160, 177, 196, 215, 243, 266, 283, 298, 317, 331
Embodied artificial intelligence, 284
Embodied, embedded, enacted, and extended (4E) cognition, 10, 200, 335
Emotion, 57, 58, 62, 66, 67, 83, 85, 86, 130, 162, 175, 177–179, 182, 183, 187, 188, 219, 253, 321, 335
Ethical challenges, v, 3, 6, 11, 13, 285, 287, 289–291, 294, 333, 340
Ethical guidelines, 3, 265–279, 284–287, 293, 327, 339–341
Ethical principles, 4, 269–271, 278, 279, 283, 293
Ethics, v, 12, 31, 33, 63, 134, 265–269, 275, 277, 278, 283–294, 298–313, 327, 339–341
Extended mind, 82, 84, 86

F

Facial identification, 318, 320–321, 327

G

Game metrics, 9, 159–170

Gaming, 8, 25, 26, 159–170, 323, 335

Grounded cognition, 199–204, 209

H

Hard skills training, 196

Human learning, 1, 6, 7, 13, 23, 38–41, 50, 51, 300

I

Informal learning, 61

Intellectual mirror, 73–87, 335

Intelligent social agents (ISAs), 7, 8, 73–87, 335

Intelligent textbooks, 11, 247–257, 337

Intelligent tutoring system (ITS), 5, 6, 13, 176, 177, 196–198, 216, 248–254, 266, 274, 284, 333, 335

Interactive learning, 7, 37–52, 337

L

Learning, 1, 20, 38, 56, 74, 91, 105, 125, 138, 160, 175, 196, 215, 233, 248, 266, 283, 298, 318, 332

Learning analytics, 5, 132, 163, 219, 222, 224, 227, 266, 274, 284, 302, 340

Learning assistant, 219

Life-long learning, 284

Location tracking, 12, 318–320, 326, 327, 340

M

Machine learning, 6, 11, 12, 22, 23, 26, 74, 97, 102, 118, 133, 137–155, 160, 196, 207, 216, 218, 219, 233, 236, 251, 276, 283, 286, 299, 300, 303, 308, 332, 333, 340

Massively-multiplayer game, 12, 298–313, 340

Math word problems (MWPs), 11, 236–241, 243, 244, 337

Metaphors of learning, 9, 125–135, 338

Multimodal data, 7, 13, 22, 23, 31, 277

Multiple perspectives, 340

N

Narrative, 8, 9, 76, 78, 85–87, 125–135, 337

Narrative-based learning, 129, 133

Natural language processing, 127, 129, 131, 133

Need deficiency, 8, 92, 93, 96, 97

Nursing education, 159–170, 335

P

Policy analysis, 268

Privacy policy, 326

Problem behavior, 8, 55, 91–102, 335

Q

Question answering, 66, 95, 97–101, 248, 275

R

Rawlsian game, 12, 298–313, 340

S

Schools, 3, 4, 6–8, 12, 20, 27, 31, 33, 55–68, 91, 93, 94, 95, 97, 101, 108, 109, 127, 132, 134, 139, 145, 149, 154, 201, 216–222, 224, 226, 228–230, 233, 255, 256, 271–275, 283–294, 308, 318–321, 326, 327, 333–335, 339–341, 347

Self-regulation, 10, 175–188, 335, 336

Simulation-based learning, 8, 10, 175–188, 335

Social-emotional skills, 61, 65–67

Social media mining, 13, 323–325, 327, 341

Students' wellbeing, 56, 67, 334, 339

Supervised machine learning algorithms, 133, 144, 145

Surveillance society, 327

T

Teaching and learning, 10, 105, 107, 109, 113, 114, 117, 119, 161, 215–230, 332, 333, 340

Teaching events, 8, 111–113, 116–118

Teaching structure, 109, 110, 112–114, 116–118

Thought experiment, 12, 298–313, 340

U

Ubiquitous AI, 317, 318, 327

V

Virtual agents, 335, 336

Virtual reality (VR), 10, 161, 163, 195–211, 307, 335, 336