

IACAP

International Association of Computing and Philosophy

Oregon 2024

Eugene



This program was compiled from \LaTeX source code on July 8, 2024.

This conference program is licensed with a Creative Commons With Attribution 4.0 license, copyright by the board of the International Association for Computing and Philosophy. Individual abstracts are copyright their authors.

This template originates from LaTeXTemplates.com and is based on the original version at:
https://github.com/maximelucas/AMCOS_booklet.

This booklet is prepared by Arzu Formánek, Brian Ballsun-Stanton and Hajo Greif.

Front cover is designed by Arzu Formánek.

Contents

| | |
|--|-----------|
| About IACAP | 4 |
| Organizers | 5 |
| Timetable | 6 |
| Program Overview | 7 |
| Keynote Addresses | 10 |
| Covey Award Keynote Address, 8 July, 10:00–11:10 | 10 |
| Simon Award Keynote Address, 10 July, 14:00–15:10 | 10 |
| Abstracts of the Talks | 11 |
| Monday, 8 July | 11 |
| Epistemology of ML – 11:40 | 11 |
| Ethics of AI – 11:40 | 11 |
| Philosophy and History of Computing – 14:00 | 12 |
| Societal Impact of AI, Computation, and Information – 14:00 | 13 |
| Epistemology of ML – 15:40 | 13 |
| Ethics of AI – 15:40 | 14 |
| Epistemology of ML – 17:20 | 15 |
| Tuesday, 9 July | 16 |
| Symposium: New Perspectives on Computer Simulation in the Age of AI – 9:40 | 16 |
| Special Interest Group: Minds and Machines – 11:40 | 17 |
| Ethics of AI – 11:40 | 18 |
| Epistemology of ML – 14:00 | 19 |
| Societal Impact of AI, Computation, and Information – 14:00 | 20 |
| Special Interest Group: Minds and Machines – 15:40 | 20 |
| Human Robot Interaction – Tuesday 15:40 | 21 |
| Philosophical Implications of AI – 17:20 | 21 |
| Wednesday, 10 July | 23 |
| Computational Modeling and Programming – 9:30 | 23 |
| Cognitive Science, Computation, and Cognition – 11:40 | 24 |
| Societal Impact of AI, Computation, and Information – 11:40 | 24 |
| Abstracts of the Posters | 26 |
| Practical Information | 29 |
| Conference Venue | 29 |
| Communication Channels | 30 |
| Conference Social Events | 30 |
| Eugene, Directions | 30 |

About IACAP

The board of IACAP welcomes you to the IACAP 2024 Conference in Eugene, Oregon.

The International Association of Computing and Philosophy

The International Association of Computing and Philosophy ([IACAP](#)) has a long-lasting tradition of promoting philosophical dialogue and interdisciplinary research on all aspects of the digital turn. IACAP's members have contributed to shaping the philosophical and ethical debate about computing, information technologies, and artificial intelligence. The 2024 annual meeting continues this tradition and gathers philosophers, ethicists, roboticists, and computer scientists and engineers interested in the following topics:

- Artificial Intelligence and Machine Learning
- Algorithmic Opacity and Bias
- Artificial Life and Moral Agency
- Autonomous Weapon Systems
- Computation, Cognition, and Cognitive Science
- Computational Modelling in Science and Social Science
- Computer-Mediated Communication
- Ethical Problems and Societal Impact of Computation and Information
- History of Computing
- Human-Robot Interaction
- Information Culture and Society
- Metaphysics of Computing
- Philosophy of Information
- Philosophy of Information Technology
- Robotics
- Virtual Reality
- ... and related issues

Covey and Simon Awards

Each year, IACAP presents two awards:

Covey Award: The Covey Award recognizes senior scholars with a substantial record of innovative research in the field of computing and philosophy.

Simon Award: The Herbert A. Simon Award recognizes scholars at an early stage of their academic career whose research is likely to reshape debates at the nexus of Computing and Philosophy.

IACAP Board

| | |
|-------------------------------------|--|
| Steve McKinlay (President) | Björn Lundgren (Vice-president) |
| Arzu Formánek (Conference Director) | Ramón Alvarado |
| Ahmed Amer | Brian Ballsun-Stanton (Technical Director) |
| Hajo Greif | Thomas M. Powers |

Organizers

University of Oregon: Philosophy

The 2024 IACAP meeting is hosted by and with the generous funding from the University of Oregon's Department of Philosophy. The conference is co-organized with the help of a College of Arts and Sciences Program Grant and generous support from the newly formed School of Computer and Data Sciences.

Program Committee

- * Abbas Bagwala (University of Oregon)
- * Ahmed Amer (Santa Clara University)
- * Alexis Elder (University of Minnesota Duluth)
- * Aliya Dewey (Centre for Philosophy and AI Research, Friedrich-Alexander-Universität)
- * Arzu Formánek (Fraunhofer IPA, University of Vienna)
- * Björn Lundgren (Centre for Philosophy and AI Research, Friedrich-Alexander-Universität)
- * Brian Ballsun-Stanton (Macquarie University)
- * Hajo Greif (Warsaw University of Technology)
- * Nico Formánek (HLRS)
- * Juan M. Durán (Delft University of Technology)
- * Kathleen Creel (Northeastern University)
- * Leonard Dung (Centre for Philosophy and AI Research, Friedrich-Alexander-Universität)
- * Maria Joseph Israel (Santa Clara University)
- * Nicola Angius (Università degli Studi di Messina)
- * Ramón Alvarado (University of Oregon)
- * Robin Hill (University of Wyoming)
- * Steve McKinlay (Wellington Institute of Technology)
- * Thomas Powers (University of Delaware)
- * Tom Sterkenburg (Ludwig-Maximilians-Universität)

Program Chairs

- * Arzu Formánek (Fraunhofer IPA, University of Vienna)
- * Björn Lundgren (Centre for Philosophy and AI Research, Friedrich-Alexander-Universität)
- * Ramon Alvarado (University of Oregon)

Support

This conference was only possible through the ongoing memberships and support, year by year and lifetime, of the IACAP Members.

**International
Association for
Computing and
Philosophy**

www.iacap.org



IACAP 2024 Timetable

Monday 8 July

| | | | | |
|-------------|---|-------------------------------------|----------|---|
| 09:30–10:00 | Registration | | | |
| 10:00–11:10 | James Cmns Covey Award Keynote: Johannes Lenhard | | | |
| 11:10–11:40 | Coffee Break | | | |
| 11:40–12:50 | Room 140 | Epistemology of ML | Room 204 | Ethics of AI |
| 12:50–14:00 | Lunch | | | |
| 14:00–15:10 | Room 140 | Philosophy and History of Computing | Room 204 | Societal Impact of AI, Computation, and Information |
| 15:10–15:40 | Coffee & Posters | | | |
| 15:40–16:50 | Room 140 | Epistemology of ML | Room 204 | Ethics of AI |
| 16:50–17:20 | Coffee Break | | | |
| 17:20–18:30 | Room 140 | Epistemology of ML | | |
| 19:00 | Welcome Drinks | | | |

Tuesday 9 July

| | | | | |
|-------------|-----------------------------|--|----------|---|
| 9:40–11:10 | Room 140 | Symposium: New Perspectives on Computer Simulation in the Age of AI | | |
| 11:10–11:40 | Coffee Break | | | |
| 11:40–12:50 | Room 140 | SIG: Minds and Machines | | Ethics of AI |
| 12:50–14:00 | Lunch | | | |
| 14:00–15:10 | Room 140 | Epistemology of ML | Room 204 | Societal Impact of AI, Computation, and Information |
| 15:10–15:40 | Coffee & Posters | | | |
| 15:40–16:50 | Room 140 | SIG: Minds and Machines | Room 204 | Human Robot Interaction |
| 16:50–17:20 | Coffee Break | | | |
| 17:20–18:30 | Room 140 | Philosophical Implications of AI | | |
| 19:00 | Conference Dinner | | | |

Wednesday 10 July

| | | | | |
|-------------|---------------------|--|----------|---|
| 9:30–11:10 | Room 140 | Computational Modeling and Programming | | |
| 11:10–11:40 | Coffee Break | | | |
| 11:40–12:50 | Room 140 | Cognitive Science, Computation, and Cognition | Room 204 | Societal Impact of AI, Computation, and Information |
| 12:50–14:00 | Lunch | | | |
| 14:00–15:10 | James Cmns | Simon Award Keynote: Corey J. Maley | | |
| 15:10–15:20 | Break | | | |
| 15:20–16:00 | Room 204 | Members Meeting (for all conference participants) | | |

Tykeson Hall on the UOregon main campus is the venue for all talks and posters.

Max's Tavern is the venue for the Welcome Drinks.

Agate Alley Bistro is the venue for the Conference Dinner.

IACAP 2024 Program

Monday 8 July

| | | | |
|-------------|---|--|--|
| 9:30–10:00 | Registration | | |
| 10:00–11:10 | Covey Keynote James Commons Chair: Hajo Greif <ul style="list-style-type: none"> • Cultures of Prediction by Johannes Lenhard | | |
| 11:10–11:40 | Coffee Break | | |
| 11:40–12:50 | Epistemology of ML Room 140 Chair: Björn Lundgren <ul style="list-style-type: none"> • Informational Injustice and Epistemic Exclusions by Abbas Bagwala • ML Assistance, Epistemic Virtues, and Judicial Decision-Making by Syed AbuMusab, Alejandro D. Tamez | Ethics of AI Room 204 Chair: Ahmed Amer <ul style="list-style-type: none"> • AI Ethics and Ordoliberalism 2.0: Towards A 'Digital Bill of Rights' by Manuel Woersdoerfer • Close Personal Relationships with AI Artifacts by Oluwaseun Sanwoolu | |
| 12:50–14:00 | Lunch | | |
| 14:00–15:10 | Philosophy and History of Computing Room 140 Chair: Björn Lundgren <ul style="list-style-type: none"> • Is Computability Enough? by Nico Formánek • The Origin of Adaptation, Effective Procedures, and the Reality of Mechanism by Hajo Greif, Kristina Šekrst | Societal Impact of AI, Computation, and Information Room 204 Chair: Abbas Bagwala <ul style="list-style-type: none"> • How to Address the Anthropomorphism Dilemma in Social AI systems: A Functional Approach by Syed AbuMusab • "It's Alive?" – The Logical and Ethical Problems of a Bad Question by Ahmed Amer | |
| 15:10–15:40 | Coffee & Posters | | |
| 15:40–16:50 | Epistemology of ML Room 140 Chair: Steve McKinlay <ul style="list-style-type: none"> • Is the 'Calculator for Words' Metaphor Useful for Communicating about LLMs? by Brian Ballsun-Stanton, Ines Hipolito • LLMs Can't Think by Russ Abbott | Ethics of AI Room 204 Chair: Arzu Formánek <ul style="list-style-type: none"> • Generative AI, the Next Battlefield in the AI Wars by John Sullins, Robin Zebrowski • Privacy as Obscurity: Big Data, Inferences, and Attention by Clint Hurshman | |
| 16:50–17:20 | Coffee Break | | |
| 17:20–18:30 | Epistemology of ML Room 140 Chair: Nico Formánek <ul style="list-style-type: none"> • Artificial Intelligence, Irrationality and Fallibilism by Steve T McKinlay • Should We Allocate Epistemic Trust to AI? by Ramon Alvarado | | |
| 19:00 | Welcome Drinks at Max's Tavern | | |

Poster Sessions: On Monday and Tuesday we will have poster sessions with refreshments (15:10-15:40 both days). Please take the opportunity to check out the posters in James Commons. If you are presenting a poster, please enjoy your refreshments next to your poster so that people can engage with you and ask questions.

Tuesday 9 July

| | | |
|-------------|--|---|
| 9:40–11:10 | Symposium: New Perspectives on Computer Simulation in the Age of AI Room 140 Chair: Ramón Alvarado <ul style="list-style-type: none"> • Novel Epistemological Considerations for AI-Infused Science by Eamon Duede • Computational Reliabilism in Computer Simulations and AI by Ramón Alvarado • Simulation Avoidance by Nico Formánek | |
| 11:10–11:40 | Coffee Break | |
| 11:40–12:50 | SIG: Minds and Machines Room 140 Chair: Tom Powers <ul style="list-style-type: none"> • Pseudo-Rehearsal Learning: Navigating the Philosophical Nexus of Cognition and Memory Dynamics by Georgios Palamas • The Measure for Weak Emergence: Taking Data-Compression Algorithms Philosophically Seriously by A. Theodore Izmaylov | Ethics of AI Room 204 Chair: Björn Lundgren <ul style="list-style-type: none"> • Offloading Wisdom: 3 Technological Relations that Mediate Phronesis by Andrew Zelnay • The Responsibility Gap in Explainable AI by Filippos Stamatou |
| 12:50–14:00 | Lunch | |
| 14:00–15:10 | Epistemology of ML Room 140 Chair: Hajo Greif <ul style="list-style-type: none"> • Navigating the Informativeness-Compression Trade-Off in Explainability by Ninell Oldenburg • A New Pathway: From Objectual to Explanatory Understanding with AlphaFold2 by Annika Schuster | Societal Impact of AI, Computation, and Information Room 204 Chair: Ahmed Amer <ul style="list-style-type: none"> • A Definitional Framework of Autonomous Weapon Systems by Björn Lundgren • Should Meeting the Deep Dangers of Generative AI Fall Upon Academia or Industry? by Selmer & Alexander Bringsjord |
| 15:10–15:40 | Coffee & Posters | |
| 15:40–16:50 | SIG: Minds and Machines Room 140 Chair: Tom Powers <ul style="list-style-type: none"> • Deontology in AI by Tom Powers • The Care and Feeding of Productive Uploads: A Color-Based Control Taxonomy for Emulated Intelligences by Jeremy A. Hansen | Human Robot Interaction Room 204 Chair: Steve McKinlay <ul style="list-style-type: none"> • Robots and Resentment by Don Berkich • The Rewards and Risks of Replacing Human Neglect with Robot Care by Ahmed Amer, Maria Kyrarini, Fatemeh Davoudi, Aleksandar Zecevic |
| 16:50–17:20 | Coffee Break | |
| 17:20–18:30 | Philosophical Implications of AI Room 140 Chair: Arzu Formánek <ul style="list-style-type: none"> • Perspicuity and Digital Depth by Robin K Hill • What Roles Shall Computers Play in the Moral Conversation? by Elizabeth O'Neill | |
| 19:00 | Conference Dinner at Agate Alley Bistro | |

Wednesday 10 July

| | | |
|-------------|---|---|
| 9:30–11:10 | Computational Modeling and Programming Room 140 Chair: Steve McKinlay <ul style="list-style-type: none">• Phenomenology of Program Surveyability by Lucas Bang• Citation Distance Matters: Towards a New Metric for Evaluating Journal Impact by Kate Barnes• Co-Simulations of Brain Language Processing using Neural Language Models by Pietro Perconti | |
| 11:10–11:40 | Coffee Break | |
| 11:40–12:50 | Cognitive Science, Computation, and Cognition Room 140 Chair: Ahmed Amer <ul style="list-style-type: none">• Are Biointelligent Systems Intelligent? Theoretical Challenges to Technology and Philosophy of Intelligence by Arzu Formanek, Robert Miehe• LLMs and the Cognitive Architecture by Ron Sun | Societal Impact of AI, Computation, and Information Room 204 Chair: Abbas Bagwala <ul style="list-style-type: none">• An Analysis of the Ukraine/Russia War and Botnets: Ethical and Anticipated Ethical Issues by Richard Wilson• Relational Equality, Data Technology, and the Anatomy of Formats by Colin Koopman |
| 12:50–14:00 | Lunch | |
| 14:00–15:10 | Simon Award Keynote James Commons Chair: Ramón Alvarado <ul style="list-style-type: none">• Radically Rethinking Computation by Corey J. Maley | |
| 15:10–15:20 | Break | |
| 15:20–16:00 | Members Meeting (for all conference participants) Room 204 | |

Keynote Addresses

Covey Award Keynote Address, 8 July, 10:00–11:10

IACAP Board has selected Professor Johannes Lenhard for the 2024 Covey Award recognizing senior scholars with a substantial record of innovative research in the field of computing and philosophy broadly conceived. The board recognizes Professor Lenhard's significant contribution to our field over several decades.

Cultures of Prediction

[Johannes Lenhard ↗](#)

Being able to predict is a key feature of science and engineering. The lecture addresses the dynamical history of prediction and discerns four cultures of prediction. Two of them are older than computers, while two of them mark a difference between computer-related cultures.

Overall, I aim to illustrate how the philosophy of computing can profit from a historical perspective. The full version of my account can be found in Johnson&Lenhard: *Cultures of Prediction*, MIT, 2024 (also available open access).

Simon Award Keynote Address, 10 July, 14:00–15:10

IACAP Board has selected Associate Professor Corey Maley for the 2023 Herbert A. Simon Award for Outstanding Research in Computing and Philosophy, which specifically recognizes scholars at an early stage of their academic career whose research is likely to reshape debates at the nexus of Computing and Philosophy.

Radically Rethinking Computation

[Corey J. Maley ↗](#)

A hundred years ago, everyone knew what a computer was: a person (often held by women) whose profession involved computing the answers to mathematical problems. Computing machines were also well-understood: these were devices used to manipulate representations of numbers in various ways, where those manipulations correspond to mathematical operations on the numbers represented.

These days, it is less clear what a computer is. We still have computing artifacts, but now questions that would have seemed silly a hundred years ago are taken seriously and debated heatedly. Is the brain a computer? Can slime molds compute? Is the universe itself a computer? Are there any facts of the matter about such questions, or is it all a matter of perspective or preference?

Nearly all theoretical and philosophical answers to these questions take the resources of automata theory to be foundational: what it is for a physical system to be a computer is for it to be appropriately described by some automaton or another (most famously, a Turing Machine). However, I think this gets things precisely backwards. Computers existed prior to the mathematics—i.e., automata theory—that describes them. Moreover, analog computers cannot be captured by automata theory at all.

To address these issues, it is important to take a step back, and look at what is—and always has been—important about computation in the first place. This includes what types of computation existed in the past, and which particular types of computation were meant to be modeled by—but not defined by—Turing Machines.

Abstracts of the Talks

Monday, 8 July

Epistemology of ML – Monday 11:40

Informational Injustice and Epistemic Exclusions

[Abbas Bagwala](#) ↗ <abagwala@uoregon.edu; abbasqb@protonmail.com>

Information is a unique resource. Asymmetries that arise out of information access or processing capacities, therefore, enable a distinctive form of injustice. This paper builds a working conception of such injustice and explores it further. Let us call it informational injustice. Informational injustice is a consequence of informational asymmetries between at least two agents, which are deeply exacerbated due to modern information and communication technologies but do not necessarily originate with them. Informational injustice is the injustice of having information from an informational surplus being used to disadvantage the agent with less information. This paper argues that informational injustice exploits an agent as a knower, specifically exploiting the agent's limitation in possessing or processing information—an agent is exploited because she is not informed or lacks in her ability to process information. Since informational injustice excludes the harmed agent from participating in knowledge practices, it is also a kind of epistemic exclusion, much like Fricker's conception of epistemic injustice or Harris' epistemic domination.

ML Assistance, Epistemic Virtues, and Judicial Decision-Making

[Syed AbuMusab](#) ↗ <syedmusab@ku.edu>

[Alejandro D. Tamez](#) ↗ <tamezad@ku.edu>

No short abstract provided.

Ethics of AI – Monday 11:40

AI Ethics and Ordoliberalism 2.0: Towards A 'Digital Bill of Rights'

[Manuel Woersdoerfer](#)

This article analyzes AI ethics from a distinct business ethics perspective, i.e., 'ordoliberalism 2.0.' It argues that the ongoing discourse on (generative) AI relies too much on corporate self-regulation and voluntary codes of conduct and thus lacks adequate governance mechanisms. To address these issues, the paper suggests not only introducing hard-law legislation with a more effective oversight structure but also merging already existing AI guidelines with an ordoliberal-inspired regulatory and competition policy. However, this link between AI ethics, regulation, and antitrust is not yet adequately discussed in the academic literature and beyond. The paper thus closes a significant gap in the academic literature and adds to the predominantly legal-political and philosophical discourse on AI governance. The paper's research questions and goals are twofold: First, it identifies ordoliberal-inspired AI ethics

principles that could serve as the foundation for a 'digital bill of rights.' Second, it shows how those principles could be implemented at the macro level with the help of ordoliberal competition and regulatory policy.

Close Personal Relationships with AI Artifacts

Oluwaseun Sanwoolu<oluwaseun.sanwoolu@ku.edu>

Can AI artifacts genuinely participate in close personal relationships? Amidst the growing trend of using chatbots and robots to combat loneliness we explore the advantages and potential harms of substituting AI for human interactions in friendships, romantic relationships, and parenting. We focus on the central role of finitude and embodiment in human relationships, emphasizing how these aspects are integral to the depth and authenticity of interpersonal connections. The central argument is that the moral significance of human-to-human connections is rooted in part in our finite and embodied existence. This renders AI incapable of replicating these relationships for reasons we explain. While we acknowledge the need to address the loneliness epidemic, we question recourse to AI artifacts as a solution. We discuss key concepts like special obligations in close personal relationships and consider whether the subjective experience of forming real friendships or falling in love with AI involves deception.

Philosophy and History of Computing – Monday 14:00

Is Computability Enough?

Nico Formánek [<nico.formanek@hirs.de>](mailto:nico.formanek@hirs.de)

The formal concept of computability is fundamentally discrete. In terms of mathematical structure it reaches the expressiveness of rational numbers but not that of the reals. Is this a problem? Recently some philosophers have grown weary of real numbers - for metaphysical and for practical reasons. Metaphysically diagonalization arguments have become suspect and it is claimed that nowhere you can find a real number in nature. Practically real numbers seem to be dispensable from natural science because one operates always at a finite level of accuracy - the infinite accuracy of real numbers is not needed. Computability naturally gives a notion of approximation and defines computable numbers. Drawing on results from computable analysis - a reconstruction of real analysis in terms of computability - I argue that the dispensability of reals doesn't come so easy - one does actually lose theorems of value. Computable numbers might in practice be indistinguishable from approximations to reals, but in principle they are. And getting kicked out of Cantor's paradise for a mere practicality would be a shame.

The Origin of Adaptation, Effective Procedures, and the Reality of Mechanism

Hajo Greif [<hans-joachim.greif@pw.edu.pl>](mailto:hans-joachim.greif@pw.edu.pl)

Kristina Šekrst

We explore the conceptions of 'mechanism' in W. Ross Ashby's machine model of adaptive organization vis-à-vis Alan M. Turing's contemporaneous notion of mechanism as an 'effective procedure' of calculating functions. Although taking their cues from the same concept of mechanism in the history of philosophy, and although developing in a shared intellectual context, there are relevant differences to consider. First, the referent of 'mechanism' is the model of computation in Turing and the system under investigation in Ashby, who sought to identify a shared kind of mechanism in various kinds of systems. Second, the respective conceptions of mechanisms account for differential degrees in the models' capacity of generalization, depending on their divergent metaphysical premises and the extent to which the mechanisms in question shall count as computational.

Societal Impact of AI, Computation, and Information – Monday 14:00

How to Address the Anthropomorphism Dilemma in Social AI systems: A Functionalist Approach

Syed AbuMusab [✉](#) <syedmusab@ku.edu>

This paper provides a novel model for social robots and robot companionship, specifically focusing on human-robot relationships (HRRs) for geriatric companionship. I call it Socio-Functional Conversationalist Companions (SFCC). I argue that the chatbots' socio-functional role determines their ontological status and the nature of HRRs. The two dominant perspectives on HRRs are the anti-anthropomorphic, which opposes assigning human-like qualities to robots, and the anthropomorphic apologist model, which supports a human-like framing. Against the pro-anthropomorphic model, I argue that completely disregarding HRRs may marginalize those who benefit from HRRs. However, equating and presenting social robots like human-human relationships can set unrealistic expectations for geriatric users. Drawing on empirical data, I argue that the pro-anthropomorphic model risks priming older adults with false expectations about social chatbots' nature and capabilities. Thus, it is best to understand HRRs as fulfilling a social function, namely providing a venue for exercising the capacity of conversing.

"It's Alive?" — The Logical and Ethical Problems of a Bad Question

Ahmed Amer

Is it wrong to ask whether a computer is alive? Can it be wrong to merely ask the question? This talk presents, as a provocation, the notion that posing questions of artificial life can be fraught with potential moral risk. Instead of a focus on the plausibility or definition of artificial life per se, we discuss the moral implications of framing such a question imprecisely, and the logical implausibility of framing such a question precisely. This creates a double-bind that seemingly precludes any morally safe formulation of the question. We therefore discuss how the moral risk we present might, hopefully, be avoided.

Epistemology of ML – Monday 15:40

Is the 'Calculator for Words' Metaphor Useful for Communicating About LLMs?

Brian Ballsun-Stanton [✉](#) <brian.ballsun-stanton@mq.edu.au>

Ines Hipolito

The nature of a Large Language Model is not intuitive to users who see it only as a chatbot or search engine. Willison coined the phrase "A calculator for words" early in 2023 and there is a debate as to the utility of the metaphor. This paper will explore arguments for and against the metaphor of a 'calculator for words,' look at various lenses around this metaphor, and discuss alternatives which try to encapsulate a tacit experience of 'talking as programming' versus these things as chatbots or search engines.

Finding an appropriate set of metaphors to support appropriate use, effective teaching, and the philosophies of mind and technology is difficult. The terms 'stochastic parrot,' 'spicy autocomplete,' and 'text extruders' while amusingly derisive, fail to communicate useful analogues to other systems. For now, though, the metaphor of a "calculator for words" still has some persuasive power. Despite Bocci's objections, it communicates a locus of control, a lack of judgement by the machine, and a necessary attention to inputs and outputs that more derogatory or anthropomorphising metaphors do not.

LLMs Can't Think

Russ Abbott ↗

LLMs are strikingly good at generating text: their output is syntactically correct, coherent, and plausible. They seem capable of following instructions and of carrying out meaningful conversations. LLMs achieve these results by using transformers to produce text based on complex patterns in their training data. But powerful though they are, transformers have nothing to do with reasoning. LLMs have no means to build or to reason from internal models; they cannot backtrack or perform exploratory search; they cannot perform after-the-fact analysis; and they cannot diagnose and correct errors. More generally, LLMs cannot formulate, apply, or correct strategies or heuristics. In short, LLMs are not a step away from Artificial General Intelligence.

Ethics of AI - Monday 15:40

Generative AI, the Next Battlefield in the AI Wars

John Sullins ↗ <john.sullins@sonoma.edu>

Robin Zebrowski ↗

In 2021, Bloomsbury Press published the book *The Great Philosophical Objections to Artificial Intelligence: The history and legacy of the AI Wars*. Even though this book was exhaustive in laying out the history and implications of AI technologies to philosophy, like all large works about current technologies, it was almost immediately out of date. Large Language Models (LLMs) have changed the debate; or have they? In this session, two of the authors of this book will discuss some of the new ethical impacts that LLMs have added to the discussion in the mere two years since publication. We see this paper as the groundwork to the new chapter in the second edition of the book, which is in preparation. The first issue relates to issues around sustainability and AI (with a focus on LLMs), and the second centers on more ontological concerns about sentience or consciousness, and all the messiness that implies.

Privacy as Obscurity: Big Data, Inferences, and Attention

Clint Hurshman ↗ <clinthurshman@ku.edu>

In response to the powerful inferential capabilities of big data analytics, philosophers have recently raised concerns about inferential or predictive privacy. Big data can be used to make incredibly specific and accurate predictions that seem to violate privacy by inferring from data that do not necessarily violate privacy. However, this form of privacy is in tension with a widely accepted view in the privacy literature (which I call the acquisition thesis), viz. that privacy rights concern the means by which information is acquired, not the information itself. However, if any means of acquisition is permissible, it seems to be inference. In this paper, I aim to resolve this tension by offering a new interpretation of the acquisition thesis. Specifically, I argue that philosophers have over-emphasized information-gathering practice types, but that even apparently innocent practice types can be made to violate privacy when they are carried out in a way that directs excessive attentional resources at a subject. In short, I argue that the right to privacy is a right to obscurity: a right against excessive attention.

Artificial Intelligence, Irrationality and Fallibilism

Steve T McKinlay <stevet.mckinlay@gmail.com>

This paper explores the parallels between science and artificial intelligence (AI), both seen as epistemic activities. It highlights the shared challenges of reliability and fallibility, with AI models subject to error much like scientific theories. We delve into the complex nature of scientific history, filled with not just facts and experiments, but also ideas, interpretations, and errors. AI, like science, involves forming beliefs, making inferences, and trusting sources. I argue that AI ought to be viewed as an epistemic technology, enhancing knowledge acquisition, creation, and dissemination. Whilst acknowledging the high standards to which AI is held, often exceeding human fallibility, I discuss the ethical and philosophical issues associated with AI, including bias, opacity, and trust. The paper concludes by arguing that accepting AI as an epistemic technology carries with it all the baggage associated with epistemology and calls for a critical examination of the limitations of machine learning algorithms. Finally, I explore the ethical and societal implications of AI's inevitable fallibility.

Should We Allocate Epistemic Trust to AI?

Ramón Alvarado  <ralvarad@uoregon.edu>

Alvarado recently suggested that “narrowing down the kind of trust that is adequate for the kind of instrument that AI is does not tell us whether the instrument itself is trustworthy and hence whether or not our trust [allocation] is well-grounded” (2023 p.13). Thus, he states that “even after it is established that epistemic trust is the only legitimate kind of trust to allocate to epistemic technologies, whether or not AI can in fact be trusted remains an open question” (*Ibid* p.1). In this paper I expand on Alvarado’s framework and explore this question with the aim of articulating some indispensably desirable criteria for such epistemic trust to be appropriately allocated.

Tuesday, 9 July

New Perspectives on Computer Simulation in the Age of AI – Tuesday 9:40

Symposium Description

Ramón Alvarado [<ralvarad@uoregon.edu>](mailto:ralvarad@uoregon.edu)

Recently, computer simulation practices are beginning to be coupled with both extractive and generative AI techniques in machine learning. This happens at different stages in the computer simulation pipeline. Amongst other uses, machine learning is sometimes used to optimize simulation code; other times it is used to analyze and extract patterns from vast amounts of data that can later be used in simulations; and yet other times, machine learning techniques are used to generate or regenerate faulty or missing data to create computer simulations. While both computer simulations and machine learning methods can be thought of as computational methods, computer simulations are non-trivially distinct from machine learning in several ways. This symposium is meant to offer a preliminary examination of the epistemological implications related to this coupling of non-trivially distinct computational methods in scientific inquiry.

Novel Epistemological Considerations for AI-Infused Science

Eamon Duede<eduede@g.harvard.edu>

This talk begins to explore the relatively unexplored domain of generative AI's role in scientific discovery by focusing on creativity in mathematical proof, the broader topic of which is often overlooked by philosophers of science. This is accomplished through reflection on scenarios where AI undertakes traditionally creative mathematical tasks, like intuitively navigating low-dimensional topology or making pivotal contributions in extremal combinatorics. The central aim of the talk is to discern the extent to which the outputs of AI-driven processes, while perhaps indistinguishable from human processes, are properly characterizable as creative, and their impact on our understanding of collaborative scientific discovery. Furthermore, the talk contrasts the use of deep learning and generative AI in mathematical proofs with conventional computational approaches. This comparison aims to broaden our comprehension of AI's potential in creative endeavors and reshape our conception of collective creativity and, thereby, epistemology within scientific collaborations.

Computational Reliabilism in Computer Simulations and AI

Ramón Alvarado [<ralvarad@uoregon.edu>](mailto:ralvarad@uoregon.edu)

The concept of computational reliabilism (CR), coined in the context of computer simulations (Durán and Formanek, 2018), has been recently deployed to justify our reliance and trust in other computational technologies, including machine learning methods (Durán and Jongsma, 2021). These deployments seek to serve several—often interrelated—purposes under the umbrella of a unified epistemological framework adept to account for a justified reliance on computational practices, methods and devices. In particular, an overarching hope of those championing such a framework is that CR can:

- Respond to or circumvent the challenges related to epistemic opacity in computational methods, and in doing so,
- warrant or justify our beliefs regarding the reliability of computational processes and their results; and hence,

- To reassure us of the possibility of trust in computational methods, practices and artifacts even if these are insurmountably opaque.

In this chapter I aim to expand on Alvarado's (2023) work and elucidate what I deem to be three major challenges to CR. I deem these challenges to have a bearing on its viability both as a general epistemological framework capable of dealing with the advent of computational methods, particularly in scientific inquiry, and as a pragmatic epistemic resolution to the justification problems related to the adoption of opaque computational methods, both of which are often cited as motivations for its adoption (Durán and Formanek, 2018; Durán and Jongsma, 2021; Russo et al., 2023; Ferrario, 2023). In particular, I focus on the following three challenges:

1. The challenge of warrant transmission and reliability-crediting properties
2. The challenge of the indispensability of endogenous features in artifactual reliability, and
3. The challenge of error-related opacity

Simulation Avoidance

[Nico Formánek](#)  <nico.formanek@htrs.de>

Large computer simulations like weather forecasts or crash tests are computationally very expensive. This is a strong incentive to avoid them. Simulation avoidance strategies have recently been complemented by methods from the ML toolkit, which promise that one can get simulation-like results at a fraction of the computational cost of simulations. Currently these methods complement rather than replace simulations, but it is hoped that at one point full avoidance might be possible. But even in the former situation the crucial question is why should it work? What are the computational shortcuts that are harnessed? Which computational problems don't allow for shortcuts? These questions are hard to answer in general. In this talk I will apply some lessons about computational shortcuts that Wilson drew in his book Physics avoidance to our case. His second lesson for example can be almost transferred verbatim: "Reasoning complexity can be greatly reduced by concentrating upon dominant effects within each sub-model." This one could translate as Computational cost of a simulation can be greatly reduced if ML concentrates on the dominant effects within the data. A statement like this might seem like a truism, but it leaves open the question if ML really latches onto the dominant effects within the data. I suggest that the answer to this question has to do with the variance of the data generating process. High variance processes not allowing for computational shortcuts, while low variance processes do. There are thus simulation problems which don't allow for simulation avoidance.

Special Interest Group: Minds and Machines – Tuesday 11:40

Pseudo-Rehearsal Learning: Navigating the Philosophical Nexus of Cognition and Memory Dynamics

[Georgios Palamas](#)  <georgios.palamas@mau.se>

This interdisciplinary project embarks on an exploration of the profound implications of pseudo-rehearsal learning, specifically within the realms of unsupervised and self-supervised learning. Beyond a technical analysis, this endeavor incorporates a philosophical exploration to unveil the connections between artificial intelligence (AI) and human cognition. Pseudorehearsal learning, a mechanism simulating artificial experiences, not only holds promise for machine learning advancements but also invites contemplation on the philosophical dimensions of cognition, memory dynamics, the plasticity-elasticity dilemma, and its intriguing parallels with dream sleep. This project posits that pseudo-rehearsal learning mirrors the memory consolidation processes observed during sleep, introducing the notion of catastrophic forgetting, the delicate balance of plasticity and elasticity in brain function, and the generation of pseudo-patterns reminiscent of dreaming.

The Measure for Weak Emergence: Taking Data Compression Algorithms Philosophically Seriously

A. Theodore Izmaylov<a@theodoreizmaylov.com>

Bedau's (2008) weak emergence is a scientifically relevant and metaphysically innocent alternative to reductionism based on explanatory incompressibility and underivability except by simulation, allowing or requiring degrees of emergence. I critically analyze two quantification approaches and show that they are scientifically irrelevant: Hovda's (2008), based on a narrow definition of simulation as computation using the exact logical structure of the underlying rules or laws, which is not how scientific simulations work; Berenstain's (2022), suggesting to use uncomputable Kolmogorov algorithmic complexity as an absolute measure, which is of no practical scientific use. I present my approach based on treating data compression algorithms, specifically dictionary methods like LZW, as a metaphor and an informative model of scientific inquiry. It interprets compression as applying our scientific dictionary of basic entities to describe the empirical data at hand and establish lawful patterns in it which explain and account for more while using less (i.e. compress) and then expanding and sometimes modifying this dictionary to describe data and patterns better. Thus, it is scientifically and philosophically relevant to the emergentism-reductionism debate. Conway's Game of Life is traditionally used as a model by all three authors and me.

Ethics of AI - Tuesday 11:40

Offloading Wisdom: 3 Technological Relations that Mediate Phronesis

Andrew Zelny<azelny@ed.ac.uk>

Although there has been much discussion regarding how technology mediates our practical and ethical lives, little has been said on how it mediates phronesis: the skilled deliberative capacity to direct our lives well. With new and emerging technologies like the generative AI of Chat-GPT, mindfulness apps such as Wysa and Headspace, and the datafication of our everyday lives, it becomes necessary to ask how these technologies and practices affect our ability to reason towards and actualize flourishing lives. I argue that (1) phronesis is a technologically mediated capacity whose mediation is best understood from the postphenomenological perspective; (2) through an adaptation of Albert Borgmann's ideas of commanding and disposable realities, practical wisdom's mediation can be understood through augmenting, diminishing, and displacing relationships with technology; (3) the rapid proliferation of new and emerging AI technologies threaten to emphasize disposable realities over commanding realities, which leads to a diminishing of phronesis, and by extension human flourishing. If we are interested in creating and using technology in ways that promote human wellbeing, we must first understand how these technologies mediate phronesis and consider what sorts of technological relationships benefit that deliberative skill.

The Responsibility Gap in Explainable AI

Filippos Stamatou  <filippostam@sun.ac.za>

The paper highlights a responsibility gap within explainable artificial intelligence (XAI), posing adverse effects on the design, development, and application of socially beneficial AI systems. Beyond merely an issue of legal responsibility, the gap is amplified by the opacity of many AI systems functioning as "black boxes." I argue that neither new regulation nor appeal to conventional moral practices suffice to resolve the issue.

XAI emerges as a promising solution, aiming to enhance transparency and comprehension of opaque systems. However, bridging the responsibility gap requires tackling two distinct problems. First, an epistemological problem

about the nature of explanations provided by XAI and their adequacy in establishing moral responsibility. Second, a blame attribution problem about determining the best way to distribute burdens among affected parties in AI-involved scenarios.

Drawing from Lipton's account on model interpretability, I contend that existing XAI techniques do not provide adequate insight into opaque systems, thereby not closing the responsibility gap. I finish with a discussion of the implications of this persistent gap before proposing three distinct strategies to effectively bridge it.

Epistemology of ML – Tuesday 14:00

Navigating the Informativeness-Compression Trade-Off in Explainability

Ninell Oldenburg  <ninelloldenburg@gmail.com>

The current paradigm for AI explainability looks something like this: take a deep learning network, throw a lot of data on it, and see if you can find a simplified representation that explains its outputs. While this has led to pointwise wins and successful explainability methods, a few philosophers and AI researchers have raised its shortcomings. These critiques, however, are mostly mentioning single methods or method subclasses without addressing the bigger picture. I argue, however, that identifying and analyzing the underlying assumptions of this paradigm could move us beyond spot-wise improvements in explainability.

For this, I will first present a list of desiderata for explainable AI that I compiled by surveying the current XAI landscape, more specifically, "Human Understandability", "Responsibility", "Scalability & Model Agnosticism", "Adjustable Granularity" of "Explanatory Depth", "Explanandum Confidence Explicitation", and "Causal Expressivity". Then, I present two underlying assumptions under which most XAI approaches operate: reductionism and external unconfoundedness. I will then contrast the desiderata with those assumptions on a case-by-case basis, and show how the current paradigm fails to satisfy them. Lastly, I will offer a glance into alternatives for explainability.

A New Pathway: From Objectual to Explanatory Understanding with AlphaFold2

Annika Schuster<annika.schuster@tu-dortmund.de>

Deepmind's AlphaFold2 (AF2) deep neural network (DNN) (Jumper et al. 2021) gained a lot of attention when it surpassed other algorithmic devices for protein structure predictions from amino acid sequences considerably in the last Critical Assessment of protein Structure Prediction. Critical voices, however, remarked that the most important questions concerning protein folding are still unanswered. Due to the high dimensionality of the data they process and of the network architecture there is no straightforward way of understanding which features of the input data were responsible for their success. Objectual and explanatory understanding as two types of understanding commonly distinguished in the literature are of particular interest with regards to DNNs. Building on the case study of AF2 in protein biology, I will argue that the relationship of DNNs to science, explanation and understanding is best described as a two-step adaptive process. In detail, building on how scientists actually work with AF2 predictions, I will show that, in the first instance, DNNs like AF2 increase objectual understanding. However, in a second step, this increase can, and often does, lead to additional explanatory understanding.

Societal Impact of AI, Computation, and Information – Tuesday 14:00

A Definitional Framework of Autonomous Weapon Systems

Björn Lundgren

In this talk, I present a definitional framework of Autonomous Weapon Systems (AWSs). The framework is based on a definition that with minor modification can be used to define sub-classes of AWSs and hence create a conceptual framework of different AWSs, such as biological, kinetic, explosive, chemical AWS. The discussion will also include a critique of a recent proposal on how to defining AWSs.

Should Meeting the Deep Dangers of Generative AI Fall Upon Academia or Industry?

Selmer Bringsjord <Selmer.Bringjord@gmail.com>, Alexander Bringsjord

The explosion of "generative AI" was caused by the for-profit business sector, not the Academy: GPT-4 is from OpenAI/Microsoft (the former having started as a non-profit, but now — puzzlingly — firmly for-profit, its source code and engineering practice tightly locked up as proprietary). Philosophizing about this system is being carried out by many in Industry. In this context, we ask:

Q: Who can and should carry out the analysis of generative AI, and figure out how to protect humanity from its deep dangers D1, D2, ... Dk, Academia or Industry?

We focus on two particular dangers — human disemployment by AI & ethically flawed AIs/robots that are powerful, autonomous, and intelligent — and supply an answer to Q with respect to both.

Special Interest Group: Minds and Machines – Tuesday 15:40

Deontology in AI

[Tom Powers](#) <tpowers@udel.edu>

Deontology, in its most general conception, is an ethics of duty. Since Asimov, there have been questions about the suitability of deontology for AI. Whether we can develop a deontological machine ethics for AI is an open question. I will describe several cases of AI deontology from science and engineering that are not top-down approaches (as Wallach and Allen held) but are 1) intrinsic to the functioning of AI, and 2) inclusive of all activities that go into its creation, i.e., data selection and curing, modeling building and fine-tuning, etc. What I hope to achieve here is to reform our expectations of the sources of deontology in machine ethics.

The Care and Feeding of Productive Uploads: A Color-Based Control Taxonomy for Emulated Intelligences

[Jeremy A. Hansen](#) <jeremyhansen@acm.org>

Here we consider how to best make use of (not-yet-extant) full brain software emulations to perform useful tasks. Using motivation theory, we define a taxonomy of techniques to ensure these human-derived software agents perform such tasks efficiently and effectively. We refer to this software as emulated intelligence or EI, to make a clear distinction between artificial intelligence (AI) which may be inspired by humans but synthetic and not directly or indirectly derivative from any embodied human. It is not our purpose to expand upon the underlying theory of

motivation, but instead explore how existing theory could be applied to the labor of EIs and extend techniques originally described by science fiction authors to a coherent taxonomy.

Human Robot Interaction – Tuesday 15:40

Robots and Resentment

Don Berkich [✉](mailto:<berkich@gmail.com>)

The prevailing philosophical view seems to assert that fully ethical artificial agents require (at least) a trio of cognitive capacities—awareness, understanding, and free will—each of which presently exceeds either our engineering grasp or our computational modeling by no small measure. Consequently we shelve puzzles over, say, robots as fully moral agents in favor of the problem of designing moral normative constraints on their ranges of behavior. Extending recent scholarship on the arguments PF Strawson offered in his influential "Freedom and Resentment", in this talk I argue that the nature of such participant reactive attitudes as resentment and gratitude are such that they apply to us regardless of whether we have awareness, understanding, or free will, and will apply to artificial agents regardless of whether they have the same or similar cognitive capacities as human persons. Thus fully ethical artificial agents are much nearer on the horizon, from an engineering standpoint, than philosophical consensus would allow.

The Rewards and Risks of Replacing Human Neglect with Robot Care

Ahmed Amer

Maria Kyrarini [✉](mailto:<kyrarini@uwy.edu>)

This talk considers two contrasting perspectives on robot care and assistance. The first one reflects the traditional view that designing robot helpers is inherently desirable, because this would allow us to assist a larger number of people. The second urges us to exercise some caution, because while it is understood that the availability of robot helpers cannot completely absolve caregivers of their responsibilities, it is nonetheless possible for the provision of robot helpers to be an aid to human neglect. Although such robotic devices can certainly improve the level of care, they can also become an absolution of such human neglect.

Philosophical Implications of AI – Tuesday 17:20

Perspicuity and Digital Depth

Robin K Hill [✉](mailto:<hill@uwyo.edu>)

We project rather narrow expectations on the models that result from deep learning of images, and we find that the process does not expose the visual features that we think are salient. And why would it? We may not grasp what is salient about an object or its image, or the features that are exposed may include artifacts from the presentation, or our notions of signifiers drawn from data may not apply, counter to assumptions that undergird the interpretation of the results. We need to interrogate all of the aspects of the data and the task in our quest for understanding of the products of artificial intelligence.

What Roles Shall Computers Play in the Moral Conversation?

Elizabeth O'Neill<e.r.h.oneill@tue.nl>

Recent developments in Artificial Intelligence (AI) research mean that computers are poised to play an unprecedented role in the human moral conversation. Many LLM-based chatbots routinely appear to assert moral claims, to give moral advice, to praise or blame, and so on. We can characterize this phenomenon as artificial moral discourse. In the first part of this talk, I identify a subset of the important classes of artificial moral discourse that LLM-based chatbots are currently performing, which have the potential to influence human morality. In the second part of the talk, I turn to the normative question of what roles LLM-based chatbots should play in human morality, and whether humans should attempt to restrict artificial moral discourse in some way.

Wednesday, 10 July

Computational Modeling and Programming – Wednesday 9:30

Phenomenology of Program Surveyability

Lucas Bang ↗ <bang@cs.hmc.edu>

This paper explores the aspect of program size and its impact on the phenomenological experience of programming, particularly surveyability. Drawing inspiration from Tymoczko's focus on proof surveyability and leveraging Blum's Size Theorem, the paper bridges the gap between mathematical proofs and programming, highlighting the unique executable nature of programs. It argues for a philosophy of program size that incorporates intersubjectivity and human experience into the analysis of program size, aiming to enhance the understanding of program surveyability in a fundamentally human context. This paper emphasizes the relevance of classical program size theorems to contemporary discussions on programming phenomenology and seeks to inject experiential intersubjectivity into theoretical results on program size. Through examining coding practices in restricted programming systems and theoretical frameworks, it underscores program size as crucial to programming's daily activities and social effects. This work aims to draw attention to program size as a fundamental issue in theoretical computer science and argues for its significance in the programming experience.

Citation Distance Matters: Towards a New Metric for Evaluating Journal Impact

Kate Barnes ↗ <kate.barnes@coloradocollege.edu>

Existing metrics for evaluating scientific quality rely on the idea that highly cited science is better. These metrics fail to distinguish between the quality of the citations themselves. Some citations, like most self-citations, are trivial while others, like citations by high impact works, are crucial. We hypothesize that citations coming from far away in the academic network indicate higher impact. The further a journal's reach in the network, the larger its impact on science in general. We compare 5 ways of measuring distance in journal citation networks and find that citation distance does correlate with impact. Highly ranked journals have longer average incoming citation distances and shorter average outgoing citation distances, whereas journals identified as misbehaving by the JCR, or suspected of citation gaming, have lower average citation distances overall. We discuss the meaning of a citation from a historical perspective and question whether citations can be used to properly evaluate the quality of scientific works at all. Our proposal is that distance metrics are important for evaluating one aspect of scientific quality, but no one metric will solve the problem. Current scientometrics research, such as this, highlights the need for a rigorous definition of scientific quality.

Co-Simulations of Brain Language Processing using Neural Language Models

Pietro Perconti

This paper provides an epistemological and methodological analysis of the practice of using neural language models to simulate brain language processing. Firstly, neural language models are introduced; a study case showing how neural language models are being applied in cognitive neuroscience for simulative purposes is then presented; after recalling the main epistemological features of the simulative method in artificial intelligence, it is finally examined how the simulative method is modified when using neural language models. In particular, it is argued that the

epistemic opacity of neural language models requires that the brain itself be used to simulate the model and to test hypotheses about the model, in what is called here a co-simulation.

Cognitive Science, Computation, and Cognition – Wednesday 11:40

Are Biointelligent Systems Intelligent? Theoretical Challenges to Technology and Philosophy of Intelligence

Arzu Formanek

Robert Miehe

Emerging Biointelligent Technologies promise a path to transform production technology and achieve sustainability goals regarding our technological coexistence with the natural environment. These systems are based on systemic complex interactions of technical, biological and information technological components (i.e. bio-, hard- and software, especially AI systems). We explore the conceptual space on intelligence and address the following question: are biointelligent systems intelligent, in what sense, in which practically, conceptually and technologically informative ways? As Philosophy lacks a dedicated subfield of intelligence, we put together the landscape of views and their relations to offer a very first structured portrait of what we can explicitly call Philosophy of Intelligence. We show that the current conceptual space cannot satisfactorily account for novel phenomena of (bio-)intelligence. Respectively, there is a pressing need for a new conceptual apparatus, to which our work aims to contribute. Responding to this not only conceptually improves the Biointelligent Technologies, and but also theoretically enriches disciplines like Philosophy and AI.

LLMs and the Cognitive Architecture

Ron Sun  <dr.ron.sun@gmail.com>

Within the framework of a computational cognitive architecture (as a comprehensive and domain-generic theory of the mind), I argue that LLMs can capture human intuition and instinct (namely, human implicit processes in the dual-process architecture). LLMs can capture human intuition from a large amount of data and repeated experience, somewhat similar to how humans develop intuition through repeated experience. However, intuition (implicit processes) alone may not be sufficient for human-level intelligence; explicit processes, including symbol manipulation and explicit rule following, are needed to supplement LLMs (e.g., for the sake of more precise reasoning), as demonstrated by the computational cognitive architecture. Also worth emphasizing is the importance of motivation, which guides human action and reasoning, also as demonstrated by the cognitive architecture. LLM-based systems need essential or intrinsic motives in order to behave in an existentially meaningful way.

Societal Impact of AI, Computation, and Information – Wednesday 11:40

An Analysis of the Ukraine/Russia War and Botnets: Ethical and Anticipated Ethical Issues

Richard Wilson<wilson@towson.edu>

This analysis is based on observations published in “Exploring Information Warfare Strategies during the Russia/Ukraine War on Twitter”, additional research papers have identified similar behaviors which firmly establish the use of botnets in the Ukraine War. The future of information operations via botnets on social media such as Twitter is characterized as a game between those trying to minimize this behavior and those trying to expand it (Orabi

2020). While there is a continuous stream of new research aiming to develop methods to detect bots, there is a simultaneously growing body of work being done to find new ways to evade detection as it improves. Significant areas of anticipated bot improvement are natural language processing, further intricacy of coordination behaviors, and improved exploitation of promotional algorithms (Polychronis, 2023). Improvement upon these aspects of botnet functionality will exacerbate the existing ethical problems generated by their existence, and thus the need for intervention to restore a sense of basic credibility on the platform.

Relational Equality, Data Technology, and the Anatomy of Formats

Colin Koopman  <koopman@uoregon.edu>

With respect to concerns about injustice, inequality, and bias the ideal of “algorithmic fairness” is the centerpiece of contemporary data ethics scholarship. This presentation aims to expand the scope of concern for contemporary data ethicists and technologists aiming for fairness by proposing a broader notion that overcomes limitations inherent in the ideal of algorithmic fairness. This broader notion is called “data equality.” The paper offers three interventions. The first intervention draws on scholarship in political philosophy focused on “relational egalitarianism” in order to show why the fairness of procedures is an insufficient metric for egalitarianism. The second intervention shows why an exclusive concentration on algorithms is insufficient for the evaluation of the ethical, social, and legal impacts of computational systems. Computational programs have two necessary elements: algorithmic processing and data structures (or what I call “formats”). The first two arguments combined together amount to a conception I call “data equality” that both enriches the ideal of algorithmic fairness and overcomes some of its limitations. This leads to a third intervention, one focused on matters of methodology. Here I describe a methodology for scrutinizing computational systems for data (in)equality and describe a case study of its implementation in learning analytics.

Abstracts of the Posters

The (Im)precise Promise of Precision Medicine: Implications of Genetic Determinism in the Case of Sickle Cell Disease

Vanessa Ferguson<vanferg@yorku.ca>

Precision Medicine (PM) is an emergent technology and method of disease treatment and prevention that considers the individual factors of each person it seeks to treat. PM initiatives are designed to tailor medical treatments to individual genetic profiles, aiming for enhanced health outcomes. However, the current framework of PM appears to primarily benefit those affluent in power and wealth, marginalizing certain populations and illnesses. This paper examines the inherent biases within PM, focusing on how its reliance on genomic data as markers of illness fails to address the full spectrum of social and political determinants of health. By predominantly targeting genetic markers, PM perpetuates genetic determinism and depersonalized care, particularly disadvantaging racialized populations. The case of sickle cell disease (SCD) is analyzed to illustrate how PM's genetic-centric approach may not only fall short of its promises but also potentially exacerbate health disparities. To truly fulfill its potential, PM must integrate a comprehensive understanding of both biological and socio-political factors affecting health, particularly for Black populations affected by SCD. This paper argues that without addressing these broader determinants, PM may continue to offer limited and inequitable health outcomes.

The E.U.'s Artificial Intelligence Act: An Ordoliberal Assessment

Manuel Woersdoerfer

In light of the rise of generative AI and recent debates about the socio-political implications of large-language models and chatbots, this article investigates the E.U.'s Artificial Intelligence Act (AIA), the world's first major attempt by a government body to address and mitigate the potentially negative impacts of AI technologies. The article critically analyzes the AIA from a distinct economic ethics perspective, i.e., 'ordoliberalism 2.0' – a perspective currently lacking in the academic literature. It evaluates, in particular, the AIA's ordoliberal strengths and weaknesses and proposes reform measures that could be taken to strengthen the AIA.

Toward an Ethical Framework for the Remote Assessment of Emotions

Max Aeon Parks [<maxaeonparks@gmail.com>](mailto:maxaeonparks@gmail.com)

This paper delves into the ethical considerations surrounding the burgeoning field of remote mental state detection facilitated by Artificial Intelligence (AI). Analyzing the interplay between technological innovation and individual rights, it comprehensively explores the ethical implications of remotely collecting data about emotion and other private mental states. Navigating diverse perspectives on regulation, this paper grapples with advocates for minimal regulations fostering innovation and those advocating for stringent measures to protect privacy. The examination extends to the nuanced relationship between freedom of thought and privacy, leveraging the United Nations Universal Declaration of Human Rights (UDHR) Articles 18 and 12. The paper contends that protections applying to freedom of thought can be extended to other mental states, advocating for a context-aware approach to remote assessment of emotions and other mental states. The ethical justifications are underscored through

examples, emphasizing benefits like aiding individuals with autism and enhancing human-robot team performance. Addressing concerns of informed consent and potential abuse of power, the paper calls for necessary regulations to balance ethical deployment and technological advancement. In conclusion, it navigates the intricate ethical landscape to propose a balanced approach that harmonizes individual freedoms with the evolving benefits of AI technology.

Social Intelligence in Generative AI-Empowered Environments

Sabine Thuermel  <sabine@thuermel.de>

Following Luciano Floridi's conviction that AI stands for "Agere sine Intellegere" (to act without understanding), one could argue that in AI-empowered environments, Social Intelligence (SI) may manifest as "Socialitas sine Intellegere" (sociality without understanding) in technical agents. Examples include swarm intelligence systems, multi-agent systems, and cooperative and competitive generative agents. Due to their nature as computational artifacts, the potential of such systems and their potential sociality becomes actual in a concrete instantiation. Since purely computational artifacts are intangible, i.e. existing in time but not in space, the situation becomes even more challenging: one and the same system can be executed in experimental environments and in real-world interaction spaces. Thus, Baudrillard's idea of simulacra must be more finely differentiated when used for AI-empowered environments. A case in point is the interactive simulacra of human behavior displayed by generative agents. They display their sociality when organizing a party in "Smallville" all by themselves in simulacra of virtual actuality. Such simulacra possess a real actuality where humans interact with these generative agents. Even if these environments are good at "pretend play" they are far from authentic human sociality.

Cyber Warfare, Nuclear Weapons and Cyber Security: Ethical and Anticipated Ethical Issues

Richard Wilson <wilson@towson.edu>

This paper will discuss the interrelationship of nuclear weapons, cyber warfare, and cyber security. Some of the most significant cyber threats to nuclear stability are now due to the intersection of technologies related to nuclear weapons and cyber technology. Cyber warfare can now be used to engage in and influence international events through cyber attacks upon nuclear systems and weapons. To prevent cyber warfare from leading to nuclear warfare there needs to be a focus on cyber security in order to protect nuclear systems and nuclear arsenals but also to mitigate cyber attacks that could lead to the use of nuclear weapons.

This analysis will define a stakeholder framework for identifying the ethical and anticipated ethical issues with cyber warfare and nuclear warfare and relate these issues to the importance of cyber security. Ethics should be at the center of the discussion of the use of nuclear weapons, nuclear warfare and cyber warfare. The need for this moral concern is due to the threat to vulnerable populations by nuclear systems and nuclear weapons, as well as the threat posed to democratic institutions by the use of nuclear weapons.

Artificial Intelligence and In Vitro Fertilization: An Ethical and Anticipatory Ethical Analysis

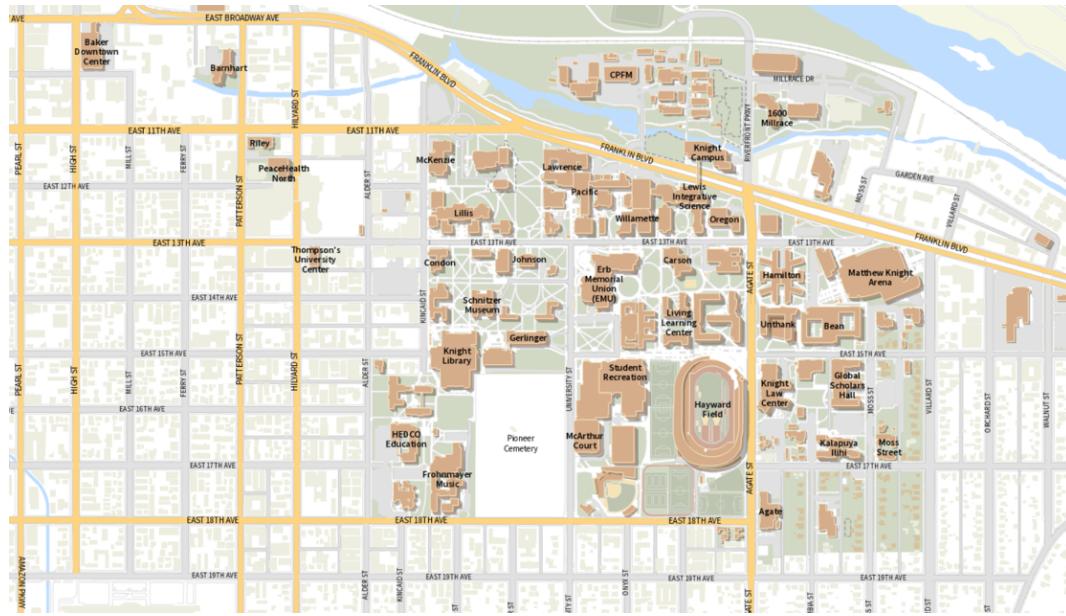
Richard Wilson <wilson@towson.edu>

The birth of the first baby conceived through the use of in vitro fertilization happened over four decades ago when Louise Brown was born on July 25, 1978. Since that time, there have been a variety of developments and advances with IVF related technological developments. These developments, related to technological improvements, include personalized ovarian stimulation, extended embryo culture at physiological oxygen level and a move to single

embryo transfer. However, the success of IVF has remained level (stagnant) for the last decade. The focus of a great deal of research on IVF related to AI has been on the effort to improve upon the 30% success rate of in vitro fertilization. The use of artificial intelligence has been deployed to improve fertility outcomes in fertility clinics. This use of AI has focused on the potential for adopting AI techniques as a way to develop improvements for issues related to IVF cycle, egg/sperm and emryo selection, as well as developing ways for improving and developing IVF treatment techniques. In addition to discussing these issues, this analysis will discuss the ethical and the anticipated ethical issues with the application of AI techniques as they are being applied to IVF technology.

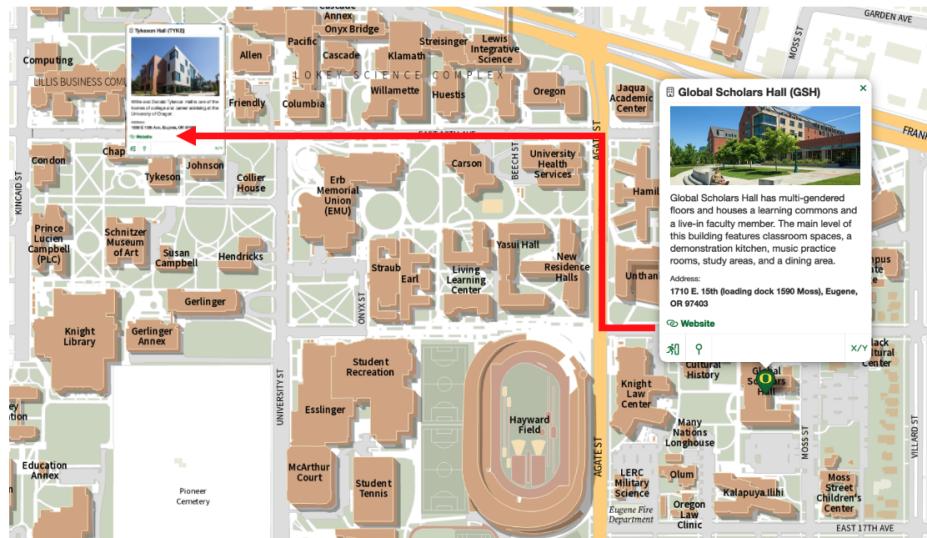
Practical Information

Conference Venue



The IACAP 2024 Conference will be held on the campus of the University of Oregon, in Eugene, Oregon. The campus is located directly East/South-East of downtown Eugene (seen on left-hand side of map above), a one mile walk away from the center. Conference sessions will take place across several floors inside [Tykeson Hall](#) (see map below).

Map above: overview. Map below: detail with directions from on-campus accommodation (GSH).



Technical Infrastructure

All rooms have full audiovisual equipment and are wheelchair accessible. See specifications for [Room 140](#) and [Room 204](#).

Wifi is available as eduroam and via the UOguest open wifi network.

Communication Channels

Please chat with other attendees and IACAP members about this conference on the IACAP Slack Workspace: <https://iacapconf.org/slack>. This conference's Zenodo community will happily accept pre-prints for any papers resulting from this conference. Currently uploaded abstracts and pre-prints can be found at: <https://zenodo.org/communities/iacap-2024/>

Conference Social Events

Welcome drinks

Monday July the 8th 7:00 PM at Max's Tavern

Max's tavern is the oldest bar in Eugene and is only a 5-minute block from our conference venue. It is a relaxed local pub popular with both Eugene residents and college students. It is also famous because it is where Matt Groening, creator of the Simpsons, spent a lot of time writing his ideas. Max's Tavern is supposed to have been a direct inspiration for Moe's in the Simpsons. Side note: Springfield, Oregon, a town across the river (5-minute drive/30-minute walk), is supposed to be THE Springfield in the show, which would make Eugene Shelbyville (for those in the know). Springfield is full of Simpson's murals.

Conference Dinner

Tuesday July the 9th 7:00 PM at Agate Alley Bistro

Agate Alley Bistro is a delightful neighborhood restaurant. It is only a 10-minute walk from our conference venue. It serves quality regional and seasonal dishes in an unpretentious setting and atmosphere. Although we may have to work with a limited menu given the size of our group, vegetarian, pescatarian and meat options will be available.

Eugene, Directions

Eugene, Oregon, with a population of approximately 176,000, is a vibrant city known for its athletic culture and picturesque landscapes. If you like accessible hikes in the middle of a city, you will find many opportunities in nearby hills, parks and by the Willamette River, all of which are 10-15 minutes walking distance from campus.

Eugene has its own international airport (EUG), located 11 mi/18 km northwest of downtown. To get to downtown Eugene or UOregon campus from there, you will have to rely on taxi or ride-sharing services (Lyft/Uber) or car

rental. Taxis are usually waiting outside of the airport and offer direct and relatively quick trips to the university (approx. USD 30 + 10-20% tip). Details on your options can be found [here](#).

Some visitors choose to fly into Portland International Airport (PDX), the largest airport in the state of Oregon, which is 122 mi/196 km away. There are several ground transportation options to Eugene, for information please visit [this link](#).

Eugene is also served by Amtrak trains (Cascades and Coast Starlight services, several connections per day, station information [here](#)) and long distance bus services ([Flixbus/Greyhound](#)).

UOregon provides travel information for visitors [here](#).