

Generative Artificial Intelligence in Learning Analytics: Contextualising Opportunities and Challenges through the Learning Analytics Cycle

LIXIANG YAN, Monash University, Australia

ROBERTO MARTINEZ-MALDONADO, Monash University, Australia

DRAGAN GAŠEVIĆ, Monash University, Australia

Generative artificial intelligence (GenAI), exemplified by ChatGPT, Midjourney, and other state-of-the-art large language models and diffusion models, holds significant potential for transforming education and enhancing human productivity. While the prevalence of GenAI in education has motivated numerous research initiatives, integrating these technologies within the learning analytics (LA) cycle and their implications for practical interventions remain underexplored. This paper delves into the prospective opportunities and challenges GenAI poses for advancing LA. We present a concise overview of the current GenAI landscape and contextualise its potential roles within Clow's generic framework of the LA cycle. We posit that GenAI can play pivotal roles in analysing unstructured data, generating synthetic learner data, enriching multimodal learner interactions, advancing interactive and explanatory analytics, and facilitating personalisation and adaptive interventions. As the lines blur between learners and GenAI tools, a renewed understanding of learners is needed. Future research can delve deep into frameworks and methodologies that advocate for human-AI collaboration. The LA community can play a pivotal role in capturing data about human and AI contributions and exploring how they can collaborate most effectively. As LA advances, it is essential to consider the pedagogical implications and broader socioeconomic impact of GenAI for ensuring an inclusive future.

CCS Concepts: • **Applied computing** → *Collaborative learning; Computer-assisted instruction; Learning management systems.*

Additional Key Words and Phrases: learning analytics, human-AI collaboration, generative artificial intelligence, ChatGPT, Midjourney, educational technology

ACM Reference Format:

Lixiang Yan, Roberto Martinez-Maldonado, and Dragan Gašević. 2024. Generative Artificial Intelligence in Learning Analytics: Contextualising Opportunities and Challenges through the Learning Analytics Cycle. 1, 1 (November 2024), 17 pages. <https://doi.org/XXXXXX.XXXXXXX>

1 INTRODUCTION

Generative artificial intelligence (GenAI), catalysed by the public release of ChatGPT in November 2022, has sparked a new era of innovation and productivity, particularly in the educational sector. Tools like ChatGPT, powered by state-of-the-art large language models (LLMs), have illustrated their potential in reshaping the educational landscape, fostering enhanced engagement and interactive learning experiences [47, 101]. A comprehensive scoping review has identified 53 different applications of LLMs in automating educational tasks [114]. Subsequent studies delving into ChatGPT and other LLM tools illustrated their capabilities in delivering comprehensive feedback, surpassing the efficacy

Authors' addresses: Lixiang Yan, Monash University, Australia; Roberto Martinez-Maldonado, Monash University, Australia; Dragan Gašević, Monash University, Australia.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

© 2024 Association for Computing Machinery.

Manuscript submitted to ACM

of human educators in articulating student performance [15]. These models also outperformed average students in reflective writing tasks [56] and innovated conversational assessments that could potentially transform conventional digital formative assessments [116]. Yet, the spotlight on ChatGPT represents just the tip of the iceberg in the vast GenAI ecosystem. The maturity of text-to-image GenAI models like DALL-E 2 and Midjourney has facilitated the creation of dynamic teaching resources, supporting multimedia learning in specialised fields such as medical education [62] and craft training [102]. Meanwhile, speech-to-text models, such as OpenAI's Whisper, could potentially transform educational content transcription [80] and automate the capture of collaborative discourse [7]. The GenAI landscape further broadens with text-to-code models like OpenAI's Codex and text-to-audio models like Meta's Voicebox [52], offering learning analytics (LA) researchers and practitioners a diverse toolkit for data analytics and communication. As these tools continue to evolve, their integration into the LA cycle [12] holds promise in addressing some of the pressing challenges in the field, especially the lack of attempts to intervene in the learning environment [67]. Nevertheless, how these novel technologies can be embedded in the LA cycle [12] and benefit the development of practical LA solutions remains largely unknown. This position paper explores the opportunities and challenges GenAI poses for advancing learning analytics. The contribution of this work to the LA community is that we contextualised the opportunities and challenges of GenAI within Clow's [12] LA cycle and summarised six endeavours for future research. These contributions could pave the way for future LA research and practice to integrate GenAI tools effectively, ensuring that they align with educational goals and uphold the principles of inclusive education while navigating the ethical complexities introduced by these advanced technologies.

2 CURRENT LANDSCAPE OF GENERATIVE AI

We first provide a concise overview of the current landscape of GenAI, focusing on the state-of-the-art models, promising applications, and supporting infrastructures instead of the underlying algorithms or technological mechanisms, as the aim of this paper is to explore the practical implications of GenAI in LA. While LLMs like GPT-4 and Llama-2 have attracted most of the attention in academic research, especially in education [10, 47, 101], GenAI's capabilities extend beyond merely text-to-text generation but also to other modalities, such as image, video, and speech generation [9, 62, 80]. Within text-based GenAI, there are many application scenarios beyond chatbots like ChatGPT. For example, the idea and exploration into AI agents (e.g., AgentGPT [83]) were based on GPT-4. These AI agents are autonomous and adaptive AI that work towards a goal and make decisions independently without continuous user input. They could potentially self-generate prompts based on the outputs of the previous prompt, making them a potential candidate for working proactively with humans [35] and simulating human behaviours [76]. Similar to text generation, code generation also utilises LLMs but with additional training on code-related content (e.g., OpenAI's Codex and Meta's Code Llama). These text-to-code generation models excel when paired with a code interpreter like in ChatGPT-4, where they can automatically conduct data analysis on provided datasets and with natural language prompts [106].

On the other hand, image generation models like Midjourney, Stable Diffusion, and Dall-E are capable of generating high-quality images from natural language text prompts. While these text-to-image generation models have been previously criticised for their inconsistency and lack of controllability over multiple generations, which might limit the practical utility in creating narrated or specialised contents (e.g., illustration of cardiac arrest training) [9, 62], recent progress, such as Midjourney's new 'inpainting' function and the latest Dall-E 3 have made them more controllable and consistent for image generation tasks. Likewise, video generation models like Runway's Gen-2 and Meta's Make-A-Video [91] are capable of creating video from text, image, or a combination of both. However, the practical utility of this line of GenAI models still needs to be discovered as the technologies become more mature and accessible. Speech-wise,

text-to-speech models like Meta's Voicebox [52] are capable of generating and synthesising audio that mirrors the tone and voice of a human in multiple different languages. Whereas, speech-to-text models like OpenAI's Whisper [30] have demonstrated their ability to automatically generate transcripts from audio inputs with performance comparable and even surpass human transcription when the audio is severely influenced by background noise [30]. Preliminary studies that utilised Whisper to automatically transcribe educational videos have already shown promising results in enriching the modality of existing educational resources [80].

Apart from these foundational generative models, there are also many supporting infrastructures that have been deployed to address some of their challenges. In particular, for LLMs, the emergence of vector embedding databases like Milvus (milvus.io) aims to address the limited context length of most existing models (e.g., GPT-3.5/4 and Llama 2). This limitation is often imposed on these LLMs to optimise computational resources as the self-attention mechanism, which is core to Transformer models, struggles with very long sequences due to its quadratic complexity [103]. Vector embedding is the representation of objects, such as words or entities, as vectors in a continuous vector space to capture semantic or relational meaning [65]. Such embedding databases can also act as an external knowledge database for LLMs, where they can draw knowledge from and use these databases to limit the context of AI responses. For example, a common workflow for developing in-document questions and response systems involves using an application framework (e.g., Langchain and Llama Index) to segment the document content into different chunks, generate the vector embedding of each chunk, and store them in an embedding database, which can be retrieved and compared with the embedding of prompts to identify the most relevant chunks. Together, the prompt and these identified chunks can be sent to LLMs like GPT-4 to generate contextually accurate and reliable responses [14].

3 GENERATIVE AI IN LEARNING ANALYTICS

Clow's [12] generic LA cycle provides a well-grounded framework for exploring the potential influence of GenAI in LA research and practices as it highlights the four essential steps in LA, including identifying the learner, collecting the relevant data, generating and presenting analytics, and delivering and evaluating interventions. As shown in figure 1, GenAI have multiple implications for each of these steps. The following sections elaborated on each of these steps and discussed the relevant theoretical and empirical basis.

3.1 Learner

The foundation of LA is centred on understanding learning and learners [25]. A pivotal initial step in the LA cycle is identifying the learner and the contexts in which learning occurs. This step is essential for ensuring that LA aligns with its primary objective: to enhance learning for learners in learning environments [51]. Before the prevalence of GenAI, pinpointing the learner was relatively straightforward. Typically, a learner could be an individual pursuing a degree or education at a university or in school, or participating in a Massive Open Online Course (MOOC), other online education platforms, or even physical space. Under such circumstances, LA researchers and practitioners could confidently ascertain the learner's identity and their learning environment. Given this clarity, the data collected, either from the learning products (e.g., assessments) or the learning processes (e.g., interactions in online forums), can be deemed reliable to a certain degree for modelling learners' metacognitive (e.g., self-regulated learning [86, 109] and socially shared regulation of learning [39, 41]), cognitive (e.g., memory and reflection [27, 49]), and affective processes (e.g., emotion and motivation [16, 95]).

As GenAI becomes more capable, the underlying assumption of knowing learners' identity might face challenges. GenAI tools like ChatGPT and Claude, which students can utilise for various learning tasks, have blurred the boundary

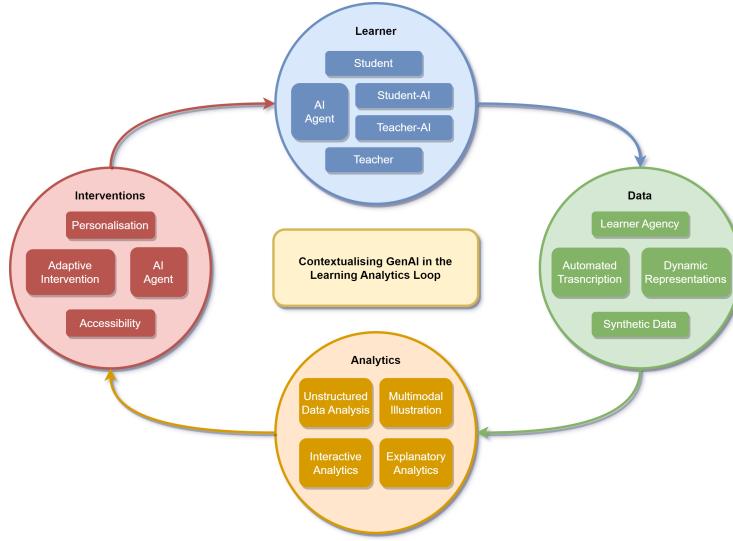


Fig. 1. Contextualising the implications of generative artificial intelligence within each step of Clow's learning analytics cycle.

between learners and tools. These tools have been demonstrated to generate learning products indicative of high metacognitive, cognitive, and social skills. Specifically, recent research indicates that ChatGPT can produce reflective writings that surpass the quality produced by average undergraduate and graduate pharmacy students. Notably, these AI-generated writings were indistinguishable from student submissions by both educators and state-of-the-art AI detection systems (e.g., OpenAI's AI text classifier) [56]. Such capabilities also extend to structured and high-stake assessments. Preliminary findings suggest that GenAI tools, like GPT-3 and ChatGPT, successfully passed the United States Medical Licensing Exam [50], passed four real exams at the University of Minnesota Law School [10], and answered basic operations management and process-analysis questions from Wharton School of Business examinations [96]. While these test-taking performances might be expected considering potential data contamination, where the vast training data of these LLMs already contained the answers, there is also evidence showing signs of emotional intelligence in GenAI. A study highlighted that AI-generated responses to patient queries in an online forum were deemed more empathetic and of higher quality than responses from physicians by experts [2].

Given GenAI's capabilities, the products of learning, such as written assignments and test scores, can easily be influenced. On the other hand, the processes of learning, which include the steps, strategies, and cognitive activities learners engage in, may remain more resistant to GenAI interference. While GenAI can provide learners with information or even automate certain tasks, the actual cognitive processes, emotions, and motivations that constitute their learning process are still human-centric [40, 47]. However, as GenAI become more sophisticated, there is also the potential for it to not only complete learning products but also mimic certain learning processes. For instance, a student might employ an AI tool to read and summarise content, bypassing the cognitive process of comprehension. While this might lead to higher task efficiency, it raises questions about the depth and authenticity of their learning process and challenges the very essence of what it means to be a 'learner' [87].

Therefore, instead of solely focusing on validating learners' identities, the LA community should pivot towards understanding the evolving nature of the learner in the age of GenAI. Learners will increasingly be expected to interact

with and utilise GenAI tools in their professional capacities [47]. This calls for a shift in our conceptualisation of learners. Drawing from novel insights and frameworks like hybrid human-AI regulation of learning [66], artificial and human cognition [90], human and AI collaboration in socially shared regulation of learning [40], hybrid intelligence [42], and human-AI teaming [119], we advocate for a renewed understanding of learners. Rather than seeing GenAI as a threat to learning integrity, we should view it as an opportunity to transform learning in the context of human-AI collaboration [45, 68]. This evolving learner definition has significant implications for other stages of the LA cycle, offering exciting avenues for both theoretical and empirical research. The LA community can play a pivotal role in capturing data about human and AI contributions and exploring how they can collaborate most effectively.

3.2 Data

LA is a data-centred field that leverages learners' data to understand and optimise their learning process [51]. Yet, not all data is relevant. The second step of the LA cycle is to capture relevant data for understanding and improving learners' learning process [12]. These data typically involve survey data (e.g., self-report measures of learner perceptions and self-efficacy), trace data (e.g., digital traces in online learning platforms like clicks and page views), and text data (e.g., forum posts, discourse in intelligence tutoring system, and written assessments) [29]. Beyond these conventional data streams, recent advancements in multimodal LA and game LA also highlighted the educational value contained in learners' physical and physiological traces (e.g., gesture, speech, and heart rate) [70], and game progression, achievements, and decisions [81]. In terms of capturing relevant learner data, while GenAI could potentially introduce complications and novel challenges, such as attributing learner agency, it can also open up new opportunities for transforming and augmenting existing data-capturing processes in LA, including formatting unstructured text data, expanding the modality of learner data, and support the generating of synthetic data for model training and open-science initiatives. The following sections elaborate on each of these challenges and opportunities.

3.2.1 Learner Agency. As the boundaries between AI and human learners blur, the significance of capturing data about the learning processes and accounting for learner agency in LA becomes even more pronounced [38]. Learner agency is defined as learners' capacity to act independently and make decisions regarding their learning [3]. For LA to optimise learning for human learners, it is essential to understand their agency during the learning process as it determines what they choose to learn and how they conduct learning, navigate challenges, and integrate feedback [82]. GenAI poses a significant challenge in identifying when and how a learner exerts agency over their learning process as learners may offload or collaborate with GenAI to conduct various learning tasks [87]. For example, some students may use GenAI tools to improve the grammar and presentation of their reflective writing, while others may ask these tools to generate the entire reflective assessment, undermining the educational purpose of such tasks. LA should ideally be able to capture and distinguish such differences in learner agency as these differences also contain significant insights. For example, based on the social cognitive perspective of self-regulation [123], a learner's decision to use GenAI tools for specific tasks but not others could potentially offer insights into their self-efficacy, values, and strategic decision-making processes. Likewise, from established work in goal orientation and self-regulated learning (e.g., traced goal orientation [122]), learners' interactions with these GenAI tools can also provide some insights into their metacognitive processes, such as when they seek GenAI for assistance, when they disagree, challenge, or question AI-generated results, and how they integrate AI-generated content with their unique understanding and perspectives. As GenAI becomes increasingly prevalent, the failure to capture learner agency could lead to biased and unfair interpretations of learning data and conceal the rich interplay of decision-making, motivation, and self-regulation [100]. Given this context, the LA

community must urgently build upon the established frameworks (e.g., Winne and Hadwin's COPES model [110]) and emerging research on human-AI regulation [40, 66] and student agency analytics [37, 38] to devise robust measures and mechanisms that capture learner agency [21]. This ensures that the essence of learning remains deeply rooted in human-centric principles.

3.2.2 Automated Transcription. GenAI could also support multimodal LA solutions that relied on capturing and analysing learners' verbal communications [11]. A major challenge exists in the accuracy and timeliness of transcribing learners' audio data as this process was often done in a post-hoc manner and mostly relied on manual efforts or external services, limiting the potential of these analytics in supporting real-time feedback and reflection practices [115]. GenAI tools like Whisper and other speech-to-text models have demonstrated the potential to automate the transcription process [80], opening up the opportunity to streamline the process of capturing and transcribing audio data to text data, which can be processed by established LA models or LLMs to extract meaningful insights, such as reflecting different teamwork constructs (e.g., situation awareness and close-loop communication) [121]. While the research on the validity and reliability of GenAI in transcribing speech data still requires further empirical validations, as this line of research matures, it could potentially offer future LA research a less labour-intensive and integrated approach to cover meaningful insights within learners' verbal communication data.

3.2.3 Dynamic Representations. The diversity of learner data that may become available for future LA research may also be enriched by the prevalence of GenAI. Specifically, GenAI's capabilities in generating human-like multimodal content introduce novel opportunities to enrich the modality and interactivity of the data that can be used in LA research and practices, offering a more comprehensive understanding of the learner's learning process [9, 53, 62]. Central to this potential is GenAI's capability to generate dynamic representations of learning [120]. Instead of the conventional reliance on static text or images, learners can now harness GenAI tools to transform their textual reflections or visual illustrations into animated sequences. For example, text-to-image models like Stable Diffusion, Midjourney, and Dall-E [9, 78] can empower educators and learners to convert textual descriptions into detailed visual content. Such advancements could progress LA to diverse learning settings, such as art-focused STEAM education [53], where learners' continued interactions with these tools (e.g., prompts and iterations) may provide the data required to gain a deeper understanding of learners' cognitive processes like their proficiency in translating abstract concepts into tangible products [22, 120]. Likewise, the emergence of text-to-video generation tools, such as Runway Gen-2 and Make-A-Video [91], further enhance the richness of data available for analysis in LA. By transforming textual content into engaging video sequences, these tools can enhance the immersive nature of learning environments, shedding light on learners' reasoning and narrative abilities. Such dynamic representations are especially pivotal for capturing reliable data on learners with specific disabilities, like dyslexia [58], reinforcing LA's commitment to inclusive education [92]. However, capturing learner agency during these novel learning processes is essential for ensuring the efficacy of LA.

3.2.4 Synthetic Data. Another opportunity that GenAI may contribute to the future of LA regarding data is supporting synthetic data initiatives. In LA research, synthetic data can serve as a valuable intermediary while waiting for actual data to become available, especially during bureaucratic processes like gaining approval from multiple ethics boards or navigating data ownership issues [5]. Such synthetic data can offer researchers immediate access, especially those who may not have access to different data sources, enabling them to refine methodologies and evaluate the scalability and adaptivity of their solutions in data-rich environments. In a closely related field, educational data mining, synthetic data generation are already gaining momentum in data-mining contexts [79], suggesting that a similar approach can be

extended to LA [5]. For example, synthetic data can potentially support the development of privacy-preserving LA [31], provide LA training materials [48], examine the inner workings of knowledge tracing models [23], and contribute to the open-science practices in LA research [17]. While most of the previous studies aimed to generate synthetic data that mirrored the statistical attributes of learner data (e.g., demographic, educational data [17]), prior works on simulated learners illustrated its potential in playing the role of a tutoring training system for educators, learning partners for students, test fields for instructional designers to conduct formative evaluations with different learning systems [23, 63]. However, these simulated learners were often limited in their ability to mimic actual human learners due to the complexity of human metacognition, cognition, and affection [23]. GenAI could potentially address these issues, especially with the recent emergence of AI agents, which are autonomous and adaptive AI that work towards a goal without continuous user input (e.g., AgentGPT [83]). For example, these AI agents have also been used to simulate human behaviour at a scale of 25 agents, where they were able to engage in daily interaction, organise parties, and make new friends [76]. Considering the versatility of these GenAI models, using these AI agents to play the role of learners, educators, and specialised roles, such as standard patients for medical students' diagnosis practice training [94], could open up a range of novel learning opportunities, and thus, generating rich data for future LA research to tackle.

3.3 Analytics

The third step of the LA cycle involves extracting analytics from learner data to generate meaningful insights into the learning process. These analytics can be categorised into four different types [93], including *descriptive analytics* that use data aggregation and visualisation to understand historical trends and patterns of learning behaviours, *diagnostic analytics* that use data mining and correlation analysis to understand the reasons behind observed trends and relationships, *predictive analytics* that focus on combining historical data with statistical models to forecast potential trends and behaviours, and *prescriptive analytics* that aim to recommend possible courses of action based on a combination of machine learning algorithms, business rules, and computational modelling. Apart from extracting these analytics from learner data, adequate reporting and presentation in LA dashboards are also essential to ensure their utility and impactfulness [44, 60, 104]. Human-centred LA approaches are essential for designing effective LA that can be adopted, comprehended, and used successfully to support learning and teaching practices [6, 98]. We foresee four major potentials for GenAI to enhance the analysis, presentation, and informativeness of LA and diversify the interactions that educational stakeholders could potentially have with future LA dashboards. These potentials include supporting the analysis of unstructured data, enabling the multimodal illustration of LA, enriching the informativeness of LA with elaborated explanations, and contributing to the development of interactive LA. The following subsection elaborated on each of these potentials.

3.3.1 Unstructured Data Analysis. Using natural language processing techniques to analyse and extract insights from unstructured text data is a rapidly evolving focus of LA research, spanning across educational discourse modelling [18] and writing analytics [1, 28]. These data directly reflect the learners' experience, perceptions, and linguistic behaviours during learning tasks. They can provide different and potentially more enriched insights into their learning process and learning gains [29]. However, transforming unstructured text data to structured metrics that can be digested by LA systems to extract meaningful indicators is challenging. While conventional natural language processing techniques, such as Linguistic Inquiry and Word Count and Latent Dirichlet Allocation [1, 113], enabled opportunities to process unstructured text data, their utility is often limited to a specific task like extracting linguistic features or topic modelling, potentially limiting the practical values of the resulting insights. Prior studies using conventional natural language

processing techniques for human-AI collaboration in qualitative analysis have identified limitations regarding the ability of these solutions to deal with the intricacies in the linguistic and semantic features of different language structures [26, 84]. On the contrary, GenAI, specifically LLMs, may offer more diversity and scalability in processing and analysing unstructured text data. For example, a study that used LLMs (GPT-3) for deductive coding found that with a pre-determined codebook, the model has achieved fair to substantial agreements with expert-coded results [112]. Furthermore, a recent study illustrated that state-of-the-art GenAI tools like ChatGPT demonstrated superior capabilities in capturing a variety of language structures. The ability of ChatGPT to elaborate on its decision was also deemed helpful to improve construct validity, pinpoint potential ambiguity in code definitions, and support human coders to achieve higher inter-rater reliability than prior tools (e.g., nCoder) [117]. This approach of using GenAI to explain its own decisions could potentially enhance the transparency of the decision-making process or at least provide additional information to cross-validate with humans. However, the validity and reliability of GenAI in analysing unstructured text data still require further investigation.

3.3.2 Multimodal Illustration. As discussed in Section 3.2.3, GenAI can potentially enhance the dynamic representations of learning resources and learner data. Similar opportunities may also apply to the illustration and presentation of LA. Prior LA research has demonstrated the benefits of data-driven personalised feedback in improving learners' academic achievement [57, 75]. One of these systems, OnTask [74], offers an interface that allows educators to transform learner data into specific rules for email generation, thereby enabling personalised feedback for each learner and ensuring that the feedback is directly relevant to their learning progress. For such personalised feedback, GenAI offers the opportunity to move beyond mere text or graphical feedback but also enables the generation of audio and potential video feedback. For example, using 3D diffusion model [107] and text-to-speech models (e.g., Meta's Voicebox [52]), educators can generate a digital representation of themselves and deliver the feedback content through narrated voice instead of merely text. Apart from benefiting learners with dyslexia [58] or preferring multimedia learning [61], there is also some evidence showing that audio and video feedback was perceived as more personal, fostering learner comprehension, and more dynamic and engaging than written feedback [64, 72]. Yet, the lack of usage of these feedback modalities was mainly associated with the workload and technical challenges in creation and distribution [64], which can be potentially addressed with the advancement in GenAI and the emergence of low-tech solutions. However, a cautious approach must be taken when using these 'deepfake' technologies to generate personable content as they present unique ethical and cybersecurity challenges (e.g., identity theft and fake information) [108].

3.3.3 Explanatory Analytics. Besides supporting dynamic representations, GenAI can also augment existing LA visualisations by providing elaborated explanations. This potential resonates with the emerging studies that utilised data storytelling to enhance learners' and educators' interpretation of LA [59, 69]. For example, prior studies have illustrated that adding text narrative and a layered approach to LA visualisations has provoked deeper reflection in educators [59] and supports the sensemaking of complex visualisations [69]. Moving away from exploratory and towards *explanatory visual LA* could facilitate learners' and educators' ease of interpretation and provide more actionable insights [20]. However, a current limitation in this line of research is the dependency on structured learning activity, which is required to formulate the rules and protocols that generate the explanations. This limitation undermines the utility of explanatory analytics in open-ended learning tasks where learning behaviours were less predictable [59]. GenAI could potentially contribute to addressing such limitations. One potential approach is to use GenAI (e.g., AgentGPT [83]) to create simulated learners and run multiple learning simulations, in which clustering techniques and reinforcement learning can be applied to identify prevalent learning strategies [23, 63]. These simulated learning strategies can derive rules and

protocols for supporting explanatory analytics in open-ended learning tasks. Another potentially more generalisable approach is to use retrieval-augmented generation [54] with LLMs to generate more accurate and context-sensitive explanations. This approach could potentially be achievable by storing previous knowledge of learning tasks (e.g., learning design and theoretical underpinnings) in an embedding database [32], where LLMs can draw on for more contextual information when generating explanations for visual LA (visualisations can be generated automatically by LLMs with a code interpreter plugin like ChatGPT-4).

3.3.4 Interactive Analytics. GenAI can also help to address several essential issues around LA dashboards, including one-size-does-not-fit-all, actionability, and data literacy. To address these issues, Verbert et al. [104] emphasised the need to move beyond building LA dashboards that serve merely as a one-way information route, where users receive information and potentially get more explanations through interaction (e.g., click on a data point). GenAI could transform these conventional dashboards into dynamic, conversational systems, where LA dashboards become interactive platforms that foster the collaboration between the user and the dashboard in achieving a shared goal like enhancing learning outcomes [8]. For example, using retrieval-augmented generation [54] to connect GenAI with course resources and relevant theoretical frameworks (e.g., self-regulated learning [60, 109]) could support AI-powered dashboards to proactively recommend resources, interventions, or learning strategies to learners based on their current performance and trajectory instead of merely presenting static data. By infusing relevant contextual and theoretical knowledge into AI-powered dashboards, LA dashboards can also become more interactive, where users can engage in conversation with the dashboards (e.g., a chatbot for the dashboard) to enhance their interpretation and internalisation of the analytics (e.g., provoking self-reflection and planning behaviours [43]). However, ensuring that AI-powered dashboards respect data privacy (e.g., transparency in data usage) and avoid potential inaccuracy and biases in their communication and recommendations is crucial [104].

3.4 Interventions

The final step to complete the LA cycle is going beyond methodologies, instruments, and scientific results and applying LA in practice to deliver interventions that optimise the learning process and outcomes [12]. Such interventions could be providing adequate support to students who have been identified as at-risk through predictive LA [36] or utilising multimodal LA to reflect and improve learners' oral presentation skills [71]. While this final step is essential to make LA more impactful and contribute to addressing immediate and long-standing challenges in education [25, 111], it is also the step that is most underdeveloped in existing LA literature [73] with merely a small amount of studies that demonstrated practical benefits on learning outcomes, supports, and teaching [105]. Regarding supporting LA intervention initiatives, we foresee three major opportunities that GenAI may offer to promote such interventions' development, delivery, and evaluation. These opportunities include enhancing the level of personalisation available through LA interventions, improving the adaptivity of these interventions from a longitudinal perspective, and broadening the accessibility of LA across different learners and regions of the world. The following sections elaborate on these opportunities and discuss the associated challenges.

3.4.1 Personalisation. Prior studies on learners' attitudes and expectations of LA have consistently identified personalisation as one of the expected values of LA [75, 85, 88]. For example, learners desired LA dashboards that generate personalised feedback and recommendations to support further improvements [46, 71]. However, personalisation is a deeply human-centred approach, where different learners may have diverse perspectives. For example, some learners may feel uncomfortable and unfair for others to receive additional learning content or support as they could potentially

be disadvantaged during the process of personalisation, others may prefer automated and unidentifiable notifications over personalised and identifiable emails from educators [74, 85]. Likewise, personalisation is also associated with dynamic representations (Section 3.2.3) and multimodal illustration (Section 3.3.2), as different learners may have different preferences over the reporting and presentation of LA [64, 72]. These differences in individual preferences pose challenges in developing LA interventions that could cater to the requirements of different learner groups. To achieve such in-depth levels of personalisation, LA interventions should be developed based on the needs, knowledge, performances, learning process and contexts of individual learners [6, 85]. While this level of personalisation seems almost impossible to achieve, we foresee the potential to move towards this goal by utilising the aforementioned capabilities of GenAI. For example, the ability of GenAI to transform unstructured text into structured data [112], analyse diverse linguistic structures [117], and generate coherent feedback [15] can be integrated into functions that foster the reciprocal communications between learners and LA interventions [104], where these communications (e.g. chat history) can be processed to update learners' current knowledge, performances, and learning process and contexts and informing researchers and practices about any emerging needs in a real-time manner (e.g., automatically summarised report). Additionally, combined with GenAI's ability to generate multimodal illustrations, learners could control what, where, and how they would like to receive LA interventions. However, this take on personalisation falls into the philosophical debate of individual autonomy [34] versus paternalism [19], which demands LA researchers and practitioners to consider the balance between structured guidance and learner autonomy over personalisation for different learner groups (e.g., age and culture).

3.4.2 Adaptive Intervention. Another opportunity that GenAI may bring to the delivery of LA involves its capabilities to support adaptive intervention. Effective LA and technologies should provide learners with adaptive support tailored to their current knowledge, performances, learning process and contexts [24, 33]. Prior LA research has focused on creating learner profiles that capture the essence of a learner's history, preferences, and performance metrics to tailor educational experiences [4, 13, 118]. These profiles, based on demographics and accumulated trace data over time, have been instrumental in informing researchers and educators about learners' learning behaviours (e.g., help-seeking [13]) and providing adequate support to optimise their educational experiences [4]. We foresee an exciting opportunity for integrating GenAI into prior LA research on learner profiles [4, 13, 118]. GenAI's ability to analyse vast and diverse datasets (as discussed in Section 3.3.1) allows it to construct detailed learner profiles, taking into account various competencies, attributes, and experiences (e.g., captured through interactive analytics; Section 3.3.4). Unlike conventional rule-based or static models, which may often be confined to specific disciplinary domains, GenAI can integrate data across multiple domains, offering a more holistic view of a learner's learning trajectory. Additionally, by connecting GenAI with external knowledge (through multiple embedding databases), it could potentially navigate the complexities of various educational spaces (e.g., technical, social, and political [4]), offering insights that are not just confined to individual contexts but span across multiple educational scenarios. This comprehensive profiling can aid in bridging the gap between current assessment practices and a more adaptive, learner-centred approach [34]. The potential of GenAI also lies in its ability to be integrated with previously validated methodologies and instruments. For example, recent system infrastructural progress of the FLORA engine (a LA platform for supporting self-regulated learning) [55] has combined learners' prior knowledge and self-regulated learning process to generate adaptive prompts for ChatGPT to provide learners with adaptive scaffolding during their learning process. Such integration could advance future LA interventions, making them more adaptive and holistic.

3.4.3 Accessibility. Regarding accessibility, GenAI presents a nuanced scenario for future LA interventions. While its capabilities in dynamic representations could indeed be beneficial for learners with specific challenges or disabilities, such as dyslexia [58], there is an imminent concern that the economic implications might overshadow these advantages [101]. For those individuals and institutions with financial capacities, leveraging commercial GenAI tools or deploying open-source versions can potentially enhance the depth, interactivity, and multimodality of their LA interventions, as elaborated in Sections 3.3.2 and 3.3.4. However, the computational resources associated with the development, deployment, and operation of GenAI are concerning, potentially mirroring or even exceeding their capabilities [47, 101]. Such economic barriers could widen the digital divide, placing advanced GenAI-powered LA tools out of reach for many learners, especially those in under-resourced settings [77]. Likewise, while GenAI could potentially support personalised learning pathways, it can amplify the looming risk of disadvantaging large numbers of people and creating a two-tiered education system: one where affluent learners enjoy personalised, dynamic, and interactive learning experiences, and another where less privileged learners are stuck with static, one-size-fits-all content [89]. The ethical implications of this discrepancy are significant. It raises questions about equity in education and the potential for GenAI and LA to heighten existing inequalities rather than alleviate them [92, 115]. Additionally, while the initial investment in GenAI might be high, there is also the ongoing cost of maintenance, updates, and ensuring compatibility with educational methodologies and policies [97]. As LA advances, it is crucial to consider the pedagogical implications of GenAI and its broader socioeconomic impact. A balanced approach, where the benefits of GenAI are made available more equitably while mitigating its challenges, will be essential for the future of inclusive and effective LA interventions [92].

4 FUTURE RESEARCH ENDEAVOURS WITH GENERATIVE AI

We summarised the contextualised opportunities and challenges of GenAI through the LA cycle [12] into six potential research endeavours for future studies to explore and investigate the synergy between LA and GenAI. Through these endeavours, the LA community can effectively and ethically leverage the technological advancement of GenAI and fulfil its foundational purpose of optimising learning for all learners.

Fist Endeavour: Hybrid Human-AI Learning Paradigms. From the learner aspect, as the lines blur between learners and GenAI tools, future LA research should delve deep into frameworks and methodologies that advocate for a collaborative learning experience between humans and AI. Investigating these hybrid models will offer in-depth insights into learners and their learning process in the age of GenAI [40, 42, 68, 119].

Second Endeavour: Advanced Analysis of Unstructured Learning Data. From the data aspect, given the richness of unstructured text data in reflecting learners' cognitive processes, future LA research should focus on leveraging GenAI for a nuanced analysis of such data. This would involve processing discourse, reflections, and other forms of textual data to gain deeper insights into the learning process [18, 28, 112, 117].

Third Endeavour: Enhanced Visual and Explanatory Analytics. From the analytics aspect, GenAI can augment conventional LA dashboards and visualisations, offering more detailed, contextual explanations and data storytelling. Future research should look at integrating GenAI to make LA more actionable, interactive, and user-friendly, ensuring a deeper understanding and reflection for both learners and educators. [8, 43, 59, 69].

Forth Endeavour: Deep Personalisation and Adaptive Interventions. From the intervention aspect, with GenAI's capability to offer real-time updates and create detailed learner profiles, future LA interventions can achieve deeper personalisation. This requires research into how GenAI can be integrated to ensure interventions remain relevant and effective throughout a learner's educational journey [4, 6, 33, 85].

Fifth Endeavour: Enhancing Accessibility in LA with GenAI. From the intervention aspect, future research should ensure that advancements in GenAI-powered interventions are accessible to all learners, regardless of their socio-economic background. This requires addressing economic barriers and ensuring that LA interventions remain inclusive and equitable [58, 92, 115].

Sixth Endeavour: Ethical and Equitable Integration of GenAI in LA. Overall, as GenAI becomes more embedded in LA, addressing ethical concerns becomes essential. Future research should focus on ensuring the integration of GenAI in LA respects data privacy, maintains ethical standards, and does not amplify existing educational inequalities and problematic practices, such as gender bias and discrimination [92, 99, 104, 115].

5 FINAL REMARK

For LA to achieve its promise in optimising learning, the LA community needs to venture beyond merely uncovering insights about learning or developing innovative methodologies and analytics. The promise of LA lies in its actionable application—addressing real-world educational problems and benefiting both teaching and learning processes [62, 73, 111]. This position piece illustrates a promising synergy between GenAI and LA. Such a combination has the potential to support systemic changes in education, making LA more comprehensive, evidence-based, and aligned with the real-world complexities of education [97]. Yet, despite this optimistic vision, we must be vigilant. GenAI also introduces new theoretical complications, such as redefining the very essence of a ‘learner’, and has broader socioeconomic impacts, notably the risk of widening the digital divide. As the field of LA moves forward, these considerations must be at the forefront of a collective endeavour among LA researchers, practitioners, and policymakers, ensuring an inclusive future for both LA and education in the era of GenAI.

ACKNOWLEDGMENTS

This research was at least in part funded by the Australian Research Council (DP210100060) and Jacobs Foundation (Research Fellowship).

REFERENCES

- [1] Laura K. Allen, Sarah C. Creer, and Püren Öncel. 2022. Natural Language Processing as a Tool for Learning Analytics - Towards a Multi-Dimensional View of the Learning Process. In *The Handbook of Learning Analytics* (2 ed.), Charles Lang, George Siemens, Alyssa Friend Wise, Dragan Gašević, and Agathe Merceron (Eds.). SoLAR, 46–53.
- [2] John W Ayers, Adam Poliak, Mark Dredze, Eric C Leas, Zechariah Zhu, Jessica B Kelley, Dennis J Faix, Aaron M Goodman, Christopher A Longhurst, Michael Hogarth, et al. 2023. Comparing physician and artificial intelligence chatbot responses to patient questions posted to a public social media forum. *JAMA internal medicine* (2023).
- [3] Albert Bandura. 1989. Human agency in social cognitive theory. *American psychologist* 44, 9 (1989), 1175.
- [4] Abhinava Barthakur, Shane Dawson, and Vitorim Kovanic. 2023. Advancing leaner profiles with learning analytics: A scoping review of current trends and challenges. In *LAK23: 13th International Learning Analytics and Knowledge Conference*. 606–612.
- [5] Alan Mark Berg, Stefan T Mol, Gábor Kissmihók, and Niall Slater. 2016. The role of a reference synthetic data generator within the field of learning analytics. *Journal of Learning Analytics* 3, 1 (2016), 107–128.
- [6] Simon Buckingham Shum, Rebecca Ferguson, and Roberto Martinez-Maldonado. 2019. Human-centred learning analytics. *Journal of Learning Analytics* 6, 2 (2019), 1–9.
- [7] Jie Cao, Ananya Ganesh, Jon Cai, Rosy Southwell, E Margaret Perkoff, Michael Regan, Katharina Kann, James H Martin, Martha Palmer, and Sidney D’Mello. 2023. A Comparative Analysis of Automatic Speech Recognition Errors in Small Group Classroom Discourse. In *Proceedings of the 31st ACM Conference on User Modeling, Adaptation and Personalization*. 250–262.
- [8] Sven Charleer, Andrew Vande Moere, Joris Klerkx, Katrien Verbert, and Tinne De Laet. 2017. Learning analytics dashboards to support adviser-student dialogue. *IEEE Transactions on Learning Technologies* 11, 3 (2017), 389–399.
- [9] Thomas KF Chiu. 2023. The impact of Generative AI (GenAI) on practices, policies and research direction in education: a case of ChatGPT and Midjourney. *Interactive Learning Environments* (2023), 1–17.

- [10] Jonathan H Choi, Kristin E Hickman, Amy Monahan, and Daniel Schwarcz. 2023. Chatgpt goes to law school. *Available at SSRN* (2023).
- [11] Yi Han Victoria Chua, Justin Dauwels, and Seng Chee Tan. 2019. Technologies for automated analysis of co-located, real-life, physical learning spaces: Where are we now?. In *Proceedings of the 9th International Conference on Learning Analytics and Knowledge*. 11–20.
- [12] Doug Clow. 2012. The learning analytics cycle: closing the loop effectively. In *LAK'12*. 134–138.
- [13] Linda Corrin, Paula G De Barba, and Aneesha Bakharia. 2017. Using learning analytics to explore help-seeking learner profiles in MOOCs. In *Proceedings of the seventh international learning analytics & knowledge conference*. 424–428.
- [14] Jiaxi Cui, Zongjian Li, Yang Yan, Bohua Chen, and Li Yuan. 2023. Chatlaw: Open-source legal large language model with integrated external knowledge bases. *arXiv preprint arXiv:2306.16092* (2023).
- [15] Wei Dai, Jionghao Lin, Flora Jin, Tongguang Li, Yi-Shan Tsai, Dragan Gašević, and Guanliang Chen. 2023. Can large language models provide feedback to students? A case study on ChatGPT. In *Proceedings of the 23rd IEEE International Conference on Advanced Learning Technologies*.
- [16] Sidney D'Mello and Emily Jensen. 2022. Emotional Learning Analytics. In *The Handbook of Learning Analytics* (2 ed.), Charles Lang, George Siemens, Alyssa Friend Wise, Dragan Gašević, and Agathe Merceron (Eds.). SoLAR, 120–129.
- [17] Mohsen Dorodchi, Erfan Al-Hossami, Aileen Benedict, and Elise Demeter. 2019. Using synthetic data generators to promote open science in higher education learning analytics. In *2019 IEEE International Conference on Big Data (Big Data)*. IEEE, 4672–4675.
- [18] Nia Dowell and Vitomir Kovanić. 2022. Modeling Educational Discourse with Natural Language Processing. In *The Handbook of Learning Analytics* (2 ed.), Charles Lang, George Siemens, Alyssa Friend Wise, Dragan Gašević, and Agathe Merceron (Eds.). SoLAR, 105–119.
- [19] Gerald Dworkin. 2014. ‘Paternalism’. In *Reading Political Philosophy*. Routledge, 342–356.
- [20] Vanessa Echeverria, Roberto Martinez-Maldonado, Simon Buckingham Shum, Katherine Chiluiza, Roger Granda, and Cristina Conati. 2018. Exploratory versus explanatory visual learning analytics: Driving teachers’ attention through educational data storytelling. *Journal of Learning Analytics* 5, 3 (2018), 73–97.
- [21] Yizhou Fan, Joep van der Graaf, Lyn Lim, Mladen Raković, Shaveen Singh, Jonathan Kilgour, Johanna Moore, Inge Molenaar, Maria Bannert, and Dragan Gašević. 2022. Towards investigating the validity of measurement of self-regulated learning based on trace data. *Metacognition and Learning* 17, 3 (2022), 949–987.
- [22] Logan Fiorella and Shelbi Kuhlmann. 2020. Creating drawings enhances learning by teaching. *Journal of Educational Psychology* 112, 4 (2020), 811.
- [23] Brendan Flanagan, Rwitajit Majumdar, and Hiroaki Ogata. 2022. Fine grain synthetic educational data: challenges and limitations of collaborative learning analytics. *IEEE Access* 10 (2022), 26230–26241.
- [24] Bryan Forsyth, Carmen Kimble, James Birch, Gary Deel, and Trisha Brauer. 2016. Maximizing the Adaptive Learning Technology Experience. *Journal of Higher Education Theory & Practice* 16, 4 (2016).
- [25] Dragan Gašević, Shane Dawson, and George Siemens. 2015. Let’s not forget: Learning analytics are about learning. *TechTrends* 59 (2015), 64–71.
- [26] Simret Araya Gebregziabher, Zheng Zhang, Xiaohang Tang, Yihao Meng, Elena L Glassman, and Toby Jia-Jun Li. 2023. Papat: Human-ai collaborative qualitative coding with explainable interactive rule synthesis. In *CHI’2023*. 1–19.
- [27] Andrew Gibson, Adam Aitken, Ágnes Sándor, Simon Buckingham Shum, Cherie Tsingos-Lucas, and Simon Knight. 2017. Reflective writing analytics for actionable feedback. In *Proceedings of the seventh international learning analytics & knowledge conference*. 153–162.
- [28] Andrew Gibson and Antonette Shibani. 2022. Natural Language Processing - Writing Analytics. In *The Handbook of Learning Analytics* (2 ed.), Charles Lang, George Siemens, Alyssa Friend Wise, Dragan Gašević, and Agathe Merceron (Eds.). SoLAR, 96–104.
- [29] Geraldine Gray and Yoav Bergner. 2022. A Practitioner’s Guide to Measurement in Learning Analytics: Decisions, Opportunities, and Challenges. In *The Handbook of Learning Analytics* (2 ed.), Charles Lang, George Siemens, Alyssa Friend Wise, Dragan Gašević, and Agathe Merceron (Eds.). SoLAR, 20–28.
- [30] Lucas Rafael Stefanel Gris, Ricardo Marcacini, Arnaldo Candido Junior, Edresson Casanova, Anderson Soares, and Sandra Maria Aluísio. 2023. Evaluating OpenAI’s Whisper ASR for Punctuation Prediction and Topic Modeling of life histories of the Museum of the Person. *arXiv preprint arXiv:2305.14580* (2023).
- [31] Mehmet Emre Gursoy, Ali Inan, Mehmet Ercan Nergiz, and Yucel Saygin. 2016. Privacy-preserving learning analytics: challenges and techniques. *IEEE Transactions on Learning technologies* 10, 1 (2016), 68–81.
- [32] Donghoon Ham, Jeong-Gwan Lee, Youngsoo Jang, and Kee-Eung Kim. 2020. End-to-end neural pipeline for goal-oriented dialogue systems using GPT-2. In *Proceedings of the 58th annual meeting of the association for computational linguistics*. 583–592.
- [33] Jeongyun Han, Kwan Hoon Kim, Wonjong Rhee, and Young Hoan Cho. 2021. Learning analytics dashboards for adaptive support in face-to-face collaborative argumentation. *Computers & Education* 163 (2021), 104041.
- [34] Michael J Hannafin and Susan M Land. 1997. The foundations and assumptions of technology-enhanced student-centered learning environments. *Instructional science* 25 (1997), 167–202.
- [35] Allyson I Hauptman, Beau G Schellble, Nathan J McNeese, and Kapil Chalil Madathil. 2023. Adapt and overcome: Perceptions of adaptive autonomous agents for human-AI teaming. *Computers in Human Behavior* 138 (2023), 107451.
- [36] Christothea Herodotou, Martin Hlostá, Avinash Boroowa, Bart Rienties, Zdenek Zdrahal, and Chrysoula Mangafa. 2019. Empowering online teachers through predictive learning analytics. *British Journal of Educational Technology* 50, 6 (2019), 3064–3079.
- [37] Danial Hooshyar, Kairit Tammets, Tobias Ley, Kati Aus, and Kair Kollom. 2023. Learning Analytics in Supporting Student Agency: A Systematic Review. *Sustainability* 15, 18 (2023), 13662.

- [38] Päivikki Jääskelä, Ville Heilala, Tommi Kärkkäinen, and Päivi Häkkinen. 2021. Student agency analytics: Learning analytics as a tool for analysing student agency in higher education. *Behaviour & Information Technology* 40, 8 (2021), 790–808.
- [39] Sanna Järvelä, Paul A Kirschner, Ernesto Panadero, Jonna Malmberg, Chris Phielix, Jos Jaspers, Marika Koivuniemi, and Hanna Järvenoja. 2015. Enhancing socially shared regulation in collaborative learning groups: Designing for CSCL regulation tools. *Educational Technology Research and Development* 63 (2015), 125–142.
- [40] Sanna Järvelä, Andy Nguyen, and Allyson Hadwin. 2023. Human and artificial intelligence collaboration for socially shared regulation in learning. *British Journal of Educational Technology* (2023).
- [41] Sanna Järvelä, Andy Nguyen, Eija Vuorenmaa, Jonna Malmberg, and Hanna Järvenoja. 2023. Predicting regulatory activities for socially shared regulation to optimize collaborative learning. *Computers in Human Behavior* 144 (2023), 107737.
- [42] Sanna Järvelä, Guoying Zhao, Janne Heikkilä, Hanna Järvenoja, Kristina Mikkonen, and Satu Kaleva. 2023. Hybrid Intelligence–Human-AI Co-Evolution and Learning in Multirealities (HI). In *HHAI 2023: Augmenting Human Intellect*. IOS Press, 392–394.
- [43] Ioana Jivet, Maren Scheffel, Hendrik Drachsler, and Marcus Specht. 2017. Awareness is not enough: Pitfalls of learning analytics dashboards in the educational practice. In *Data Driven Approaches in Digital Education: 12th European Conference on Technology Enhanced Learning, EC-TEL 2017, Tallinn, Estonia, September 12–15, 2017, Proceedings* 12. Springer, 82–96.
- [44] Ioana Jivet, Maren Scheffel, Marcus Specht, and Hendrik Drachsler. 2018. License to evaluate: Preparing learning analytics dashboards for educational practice. In *Proceedings of the 8th international conference on learning analytics and knowledge*. 31–40.
- [45] Srecko Joksimovic, Dirk Ifenthaler, Rebecca Marrone, Maarten De Laat, and George Siemens. 2023. Opportunities of artificial intelligence for supporting complex problem-solving: Findings from a scoping review. *Computers and Education: Artificial Intelligence* (2023), 100138.
- [46] Fatma Gizem Karaoglan Yilmaz and Ramazan Yilmaz. 2020. Student opinions about personalized recommendation and feedback based on learning analytics. *Technology, knowledge and learning* 25 (2020), 753–768.
- [47] Enkelejda Kasneci, Kathrin Seßler, Stefan Küchemann, Maria Bannert, Daryna Dementieva, Frank Fischer, Urs Gasser, Georg Groh, Stephan Günemann, Eyke Hüllermeier, et al. 2023. ChatGPT for good? On opportunities and challenges of large language models for education. *Learning and individual differences* 103 (2023), 102274.
- [48] Brenda Koester. 2015. University of Michigan source code and data associated with practical learning analytics course. <https://www.solaresearch.org/about/what-is-learning-analytics/>. Accessed: 2023-09-24.
- [49] Vitomir Kovanović, Srećko Joksimović, Negin Mirriahi, Ellen Blaine, Dragan Gašević, George Siemens, and Shane Dawson. 2018. Understand students' self-reflections through learning analytics. In *Proceedings of the 8th international conference on learning analytics and knowledge*. 389–398.
- [50] Tiffany H Kung, Morgan Cheatham, Arielle Medenilla, Czarina Sillos, Lorie De Leon, Camille Elepaño, Maria Madriaga, Rimel Aggabao, Giezel Diaz-Candido, James Maningo, et al. 2023. Performance of ChatGPT on USMLE: Potential for AI-assisted medical education using large language models. *PLoS digital health* 2, 2 (2023), e0000198.
- [51] Charles Lang, Alyssa Friend Wise, Agathe Merceron, Dragan Gašević, and George Siemens. [n. d.]. What is Learning Analytics? In *The Handbook of Learning Analytics* (2 ed.), Charles Lang, George Siemens, Alyssa Friend Wise, Dragan Gašević, and Agathe Merceron (Eds.). SoLAR, 8–18.
- [52] Matthew Le, Apoorv Vyas, Bowen Shi, Brian Karrer, Leda Sari, Rashel Moritz, Mary Williamson, Vimal Manohar, Yossi Adi, Jay Mahadeokar, et al. 2023. Voicebox: Text-guided multilingual universal speech generation at scale. *arXiv preprint arXiv:2306.15687* (2023).
- [53] Unggi Lee, Ariel Han, Jeongjin Lee, Eunseo Lee, Jiwon Kim, Hyeoncheol Kim, and Cheolil Lim. 2023. Prompt Aloud!: Incorporating image-generative AI into STEAM class with learning analytics using prompt data. *Education and Information Technologies* (2023), 1–31.
- [54] Patrick Lewis, Ethan Perez, Aleksandra Piktus, Fabio Petroni, Vladimir Karpukhin, Naman Goyal, Heinrich Küttler, Mike Lewis, Wen-tau Yih, Tim Rocktäschel, et al. 2020. Retrieval-augmented generation for knowledge-intensive nlp tasks. *Advances in Neural Information Processing Systems* 33 (2020), 9459–9474.
- [55] Xinyu Li, Yizhou Fan, Mladen Raković, Shaveen Singh, Joep van der Graaf, Lyn Lim, Johanna Moore, Inge Molenaar, Maria Bannert, and Dragan Gašević. 2023. FLoRA: Facilitating Self-Regulated Learning with Personalised Scaffolds on Students' own Regulation Activities. *arXiv preprint arXiv:2303.15687* (2023).
- [56] Yuheng Li, Lele Sha, Lixiang Yan, Jionghao Lin, Mladen Raković, Kirsten Galbraith, Kayley Lyons, Dragan Gašević, and Guanliang Chen. 2023. Can large language models write reflectively. *Computers and Education: Artificial Intelligence* 4 (2023), 100140.
- [57] Lisa-Angelique Lim, Sheridan Gentili, Abelardo Pardo, Vitomir Kovanović, Alexander Whitelock-Wainwright, Dragan Gašević, and Shane Dawson. 2021. What changes, and for whom? A study of the impact of learning analytics-based process feedback in a large course. *Learning and Instruction* 72 (2021), 101202.
- [58] Louise Long, Sean MacBlain, and Martin MacBlain. 2007. Supporting students with dyslexia at the secondary level: An emotional model of literacy. *Journal of Adolescent & Adult Literacy* 51, 2 (2007), 124–134.
- [59] Roberto Martinez-Maldonado, Vanessa Echeverria, Gloria Fernandez Nieto, and Simon Buckingham Shum. 2020. From data to insights: A layered storytelling approach for multimodal learning analytics. In *Proceedings of the 2020 chi conference on human factors in computing systems*. 1–15.
- [60] Wannisa Matcha, Dragan Gašević, Abelardo Pardo, et al. 2019. A systematic review of empirical studies on learning analytics dashboards: A self-regulated learning perspective. *IEEE transactions on learning technologies* 13, 2 (2019), 226–245.
- [61] Richard E Mayer. 2005. Cognitive theory of multimedia learning. *The Cambridge handbook of multimedia learning* 41 (2005), 31–48.
- [62] Carlo Alberto Mazzoli, Federico Semeraro, and Lorenzo Gamberini. 2023. Enhancing Cardiac Arrest Education: Exploring the potential use of MidJourney. *Resuscitation* 189 (2023).

- [63] Gord McCalla and John Champaign. 2013. Simulated learners. *IEEE Intelligent Systems* 28, 04 (2013), 67–71.
- [64] Josh McCarthy. 2015. Evaluating written, audio and video feedback in higher education summative assessment tasks. *Issues in Educational Research* 25, 2 (2015), 153–169.
- [65] Tomas Mikolov, Ilya Sutskever, Kai Chen, Greg S Corrado, and Jeff Dean. 2013. Distributed representations of words and phrases and their compositionality. *Advances in neural information processing systems* 26 (2013).
- [66] Inge Molenaar. 2022. The concept of hybrid human-AI regulation: Exemplifying how to support young learners' self-regulated learning. *Computers and Education: Artificial Intelligence* 3 (2022), 100070.
- [67] Benjamin A Motz, Yoav Bergner, Christopher A Brooks, Anna Gladden, Geraldine Gray, Charles Lang, Warren Li, Fernando Marmolejo-Ramos, and Joshua D Quick. 2023. A LAK of direction: Misalignment between the goals of learning analytics and its research scholarship. *Journal of Learning Analytics* (2023), 1–13.
- [68] National Academies of Sciences, Engineering, and Medicine and others. 2021. Human-AI teaming: State-of-the-art and research needs. (2021).
- [69] Gloria Milena Fernandez Nieto, Kirsty Kitto, Simon Buckingham Shum, and Roberto Martinez-Maldonado. 2022. Beyond the learning analytics dashboard: Alternative ways to communicate student data insights combining visualisation, narrative and storytelling. In *International Conference on Learning Analytics and Knowledge 2022*. Association for Computing Machinery (ACM), 219–229.
- [70] Xavier Ochoa. 2022. Multimodal Learning Analytics - Rationale, Process, Examples, and Direction. In *The Handbook of Learning Analytics* (2 ed.), Charles Lang, George Siemens, Alyssa Friend Wise, Dragan Gašević, and Agathe Merceron (Eds.). SOLAR, 54–65.
- [71] Xavier Ochoa, Federico Domínguez, Bruno Guamán, Ricardo Maya, Gabriel Falcones, and Jaime Castells. 2018. The RAP system: Automatic feedback of oral presentation skills using multimodal analysis and low-cost sensors. In *LAK'18*. 360–364.
- [72] John Orlando. 2016. A comparison of text, voice, and screencasting feedback to online students. *American Journal of Distance Education* 30, 3 (2016), 156–166.
- [73] Zacharoula Papamitsiou, Michail N Giannakos, and Xavier Ochoa. 2020. From childhood to maturity: are we there yet? Mapping the intellectual progress in learning analytics during the past decade. In *Proceedings of the tenth international conference on learning analytics & knowledge*. 559–568.
- [74] Abelardo Pardo, Kathryn Bartimote, Simon Buckingham Shum, Shane Dawson, Jing Gao, Dragan Gašević, Steven Leichtweis, Danny Liu, Roberto Martínez-Maldonado, Negin Mirriahi, et al. 2018. Ontask: Delivering data-informed, personalized learning support actions. (2018).
- [75] Abelardo Pardo, Jelena Jovanovic, Shane Dawson, Dragan Gašević, and Negin Mirriahi. 2019. Using learning analytics to scale the provision of personalised feedback. *British Journal of Educational Technology* 50, 1 (2019), 128–138.
- [76] Joon Sung Park, Joseph C O'Brien, Carrie J Cai, Meredith Ringel Morris, Percy Liang, and Michael S Bernstein. 2023. Generative agents: Interactive simulacra of human behavior. *arXiv preprint arXiv:2304.03442* (2023).
- [77] Taciana Pontual Falcão, Rafael Ferreira Mello, and Rodrigo Lins Rodrigues. 2020. Applications of learning analytics in Latin America. , 871–874 pages.
- [78] Alec Radford, Jong Wook Kim, Chris Hallacy, Aditya Ramesh, Gabriel Goh, Sandhini Agarwal, Girish Sastry, Amanda Askell, Pamela Mishkin, Jack Clark, et al. 2021. Learning transferable visual models from natural language supervision. In *International conference on machine learning*. PMLR, 8748–8763.
- [79] Md Mostafizer Rahman, Yutaka Watanobe, Taku Matsumoto, Rage Uday Kiran, and Keita Nakamura. 2022. Educational data mining to support programming learning using problem-solving data. *IEEE Access* 10 (2022), 26186–26202.
- [80] Ashwin Rao. 2023. Transcribing Educational Videos Using Whisper: A preliminary study on using AI for transcribing educational videos. *arXiv preprint arXiv:2307.03200* (2023).
- [81] Emily Reardon, Vishesh Kumar, and Glenda Revelle. 2022. Game Learning Analytics. In *The Handbook of Learning Analytics* (2 ed.), Charles Lang, George Siemens, Alyssa Friend Wise, Dragan Gašević, and Agathe Merceron (Eds.). SOLAR, 152–162.
- [82] Johnmarshall Reeve. 2013. How students create motivationally supportive learning environments for themselves: The concept of agentic engagement. *Journal of educational psychology* 105, 3 (2013), 579.
- [83] Reworkd. 2023. Introduction to AgentGPT. <https://docs.reworkd.ai/introduction>. Accessed: 2023-09-24.
- [84] Tim Rietz and Alexander Maedche. 2021. Cody: An AI-based system to semi-automate coding for qualitative research. In *CHI'2021*. 1–14.
- [85] Lynne D Roberts, Joel A Howell, and Kristen Seaman. 2017. Give me a customizable dashboard: Personalized learning analytics dashboards in higher education. *Technology, Knowledge and Learning* 22 (2017), 317–333.
- [86] Ido Roll and Philip H Winne. 2015. Understanding, evaluating, and supporting self-regulated learning using learning analytics. *Journal of Learning Analytics* 2, 1 (2015), 7–12.
- [87] Jürgen Rudolph, Samson Tan, and Shannon Tan. 2023. ChatGPT: Bullshit spewer or the end of traditional assessments in higher education? *Journal of Applied Learning and Teaching* 6, 1 (2023).
- [88] Clara Schumacher and Dirk Ifenthaler. 2018. Features students really expect from learning analytics. *Computers in human behavior* 78 (2018), 397–407.
- [89] Neil Selwyn. 2019. What's the problem with learning analytics? *Journal of Learning Analytics* 6, 3 (2019), 11–19.
- [90] George Siemens, Fernando Marmolejo-Ramos, Florence Gabriel, Kelsey Medeiros, Rebecca Marrone, Srecko Joksimovic, and Maarten de Laat. 2022. Human and artificial cognition. *Computers and Education: Artificial Intelligence* 3 (2022), 100107.
- [91] Uriel Singer, Adam Polyak, Thomas Hayes, Xi Yin, Jie An, Songyang Zhang, Qiyuan Hu, Harry Yang, Oron Ashual, Oran Gafni, et al. 2022. Make-a-video: Text-to-video generation without text-video data. *arXiv preprint arXiv:2209.14792* (2022).
- [92] Roger Slee. 2018. Defining the scope of inclusive education. (2018).

- [93] SOLAR. 2021. What is Learning Analytics? <https://www.solaresearch.org/about/what-is-learning-analytics/>. Accessed: 2023-09-25.
- [94] Wenfeng Song, Xia Hou, Shuai Li, Chenglizhao Chen, Danyang Gao, Yuzhe Sun, Jianxia Hou, Aimin Hao, et al. 2022. An Intelligent Virtual Standard Patient for Medical Students Training Based on Oral Knowledge Graph. *IEEE Transactions on Multimedia* (2022).
- [95] Calkin Suero Montero and Jarkko Suhonen. 2014. Emotion analysis meets learning analytics: online learner profiling beyond numerical data. In *Proceedings of the 14th Koli calling international conference on computing education research*. 165–169.
- [96] Christian Terwiesch. 2023. Would chat GPT3 get a Wharton MBA. *A prediction based on its performance in the operations management course. Wharton: Mack Institute for Innovation Management/University of Pennsylvania/School Wharton* (2023).
- [97] Yi-Shan Tsai and Dragan Gašević. 2017. Learning analytics in higher education—challenges and policies: a review of eight learning analytics policies. In *Proceedings of the seventh international learning analytics & knowledge conference*. 233–242.
- [98] Yi-Shan Tsai and Roberto Martinez-Maldonado. 2022. Human-Centered Approaches to Data-Informed Feedback. In *The Handbook of Learning Analytics* (2 ed.), Charles Lang, George Siemens, Alyssa Friend Wise, Dragan Gašević, and Agathe Merceron (Eds.). SoLAR, 213–222.
- [99] Yi-Shan Tsai, Alexander Whitelock-Wainwright, and Dragan Gašević. 2020. The privacy paradox and its implications for learning analytics. In *Proceedings of the tenth international conference on learning analytics & knowledge*. 230–239.
- [100] Julianne C Turner and Helen Patrick. 2008. How does motivation develop and why does it change? Reframing motivation research. *Educational psychologist* 43, 3 (2008), 119–131.
- [101] Eva AM Van Dis, Johan Bollen, Willem Zuidema, Robert van Rooij, and Claudi L Bockting. 2023. ChatGPT: five priorities for research. *Nature* 614, 7947 (2023), 224–226.
- [102] Henriikka Vartiainen and Matti Tedre. 2023. Using artificial intelligence in craft education: crafting with text-to-image generative models. *Digital Creativity* 34, 1 (2023), 1–21.
- [103] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. 2017. Attention is all you need. *Advances in neural information processing systems* 30 (2017).
- [104] Katrien Verbert, Xavier Ochoa, Robin De Croon, Raphael A Dourado, and Tinne De Laet. 2020. Learning analytics dashboards: the past, the present and the future. In *Proceedings of the tenth international conference on learning analytics & knowledge*. 35–40.
- [105] Olga Viberg, Mathias Hatakka, Olof Bälter, and Anna Mavroudi. 2018. The current landscape of learning analytics in higher education. *Computers in human behavior* 89 (2018), 98–110.
- [106] Lei Wang, Xijin Ge, Li Liu, and Gangqing Hu. 2023. Code Interpreter for Bioinformatics: Are We There Yet? *Annals of Biomedical Engineering* (2023), 1–3.
- [107] Tengfei Wang, Bo Zhang, Ting Zhang, Shuyang Gu, Jianmin Bao, Tadas Baltrusaitis, Jingjing Shen, Dong Chen, Fang Wen, Qifeng Chen, et al. 2023. Rodin: A generative model for sculpting 3d digital avatars using diffusion. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 4563–4573.
- [108] Mika Westerlund. 2019. The emergence of deepfake technology: A review. *Technology innovation management review* 9, 11 (2019).
- [109] Philip H. Winne. 2022. Learning Analytics for Self-Regulated Learning. In *The Handbook of Learning Analytics* (2 ed.), Charles Lang, George Siemens, Alyssa Friend Wise, Dragan Gašević, and Agathe Merceron (Eds.). SoLAR, 78–85.
- [110] P. H. Winne and A. F. Hadwin. 1998. Studying as self-regulated learning. In *Metacognition in educational theory and practice*, D. J. Hacker, J. Dunlosky, and A. C. Graesser (Eds.). Lawrence Erlbaum Associates Publishers, 277–304.
- [111] Alyssa F Wise, Simon Knight, and Xavier Ochoa. 2021. What makes learning analytics research matter. *Journal of Learning Analytics* 8, 3 (2021), 1–9.
- [112] Ziang Xiao, Xingdi Yuan, Q Vera Liao, Rania Abdelghani, and Pierre-Yves Oudeyer. 2023. Supporting Qualitative Analysis with Large Language Models: Combining Codebook with GPT-3 for Deductive Coding. In *Companion Proceedings of the 28th International Conference on Intelligent User Interfaces*. 75–78.
- [113] Wanli Xing, Hee-Sun Lee, and Antonette Shibani. 2020. Identifying patterns in students' scientific argumentation: content analysis through text mining using Latent Dirichlet Allocation. *Educational Technology Research and Development* 68 (2020), 2185–2214.
- [114] Lixiang Yan, Lele Sha, Linxuan Zhao, Yuheng Li, Roberto Martinez-Maldonado, Guanliang Chen, Xinyu Li, Yueqiao Jin, and Dragan Gašević. 2023. Practical and ethical challenges of large language models in education: A systematic scoping review. *British Journal of Educational Technology* 00 (2023), 1–23.
- [115] Lixiang Yan, Linxuan Zhao, Dragan Gašević, and Roberto Martinez-Maldonado. 2022. Scalability, sustainability, and ethicality of multimodal learning analytics. In *LAK22: 12th international learning analytics and knowledge conference*. 13–23.
- [116] Seyma N Yildirim-Erbasli and Okan Bulut. 2023. Conversation-based assessment: A novel approach to boosting test-taking effort in digital formative assessment. *Computers and Education: Artificial Intelligence* 4 (2023), 100135.
- [117] Andres Felipe Zambrano, Xiner Liu, Amanda Barany, Ryan Shaun Baker, Juhan Kim, and Nidhi Nasiar. 2023. From nCoder to ChatGPT: From Automated Coding to Refining Human Coding. (2023). <https://doi.org/10.35542/osf.io/grmzh>
- [118] Andrew Zamecník, Vitomír Kováčović, Srećko Joksimović, and Lin Liu. 2022. Exploring non-traditional learner motivations and characteristics in online learning: A learner profile study. *Computers and Education: Artificial Intelligence* 3 (2022), 100051.
- [119] Rui Zhang, Nathan J McNeese, Guo Freeman, and Geoff Musick. 2021. "An ideal human" expectations of AI teammates in human-AI teaming. *Proceedings of the ACM on Human-Computer Interaction* 4, CSCW3 (2021), 1–25.

- [120] Zhihui Helen Zhang and Marcia C Linn. 2011. Can generating representations enhance learning with dynamic visualizations? *Journal of research in science teaching* 48, 10 (2011), 1177–1198.
- [121] Linxuan Zhao, Zachari Swiecki, Dragan Gašević, Lixiang Yan, Samantha Dix, Hollie Jaggard, Rosie Wotherspoon, Abra Osborne, Xinyu Li, Riordan Alfredo, and Roberto Martinez-Maldonado. 2023. METS: Multimodal Learning Analytics of Embodied Teamwork Learning. In *Proceedings of the 13th International Conference on Learning Analytics and Knowledge* (Arlington, TX, USA) (LAK2023). 186–196.
- [122] Mingming Zhou and Philip H Winne. 2012. Modeling academic achievement by self-reported versus traced goal orientation. *Learning and Instruction* 22, 6 (2012), 413–419.
- [123] Barry J Zimmerman. 2000. Attaining self-regulation: A social cognitive perspective. In *Handbook of self-regulation*. Elsevier, 13–39.