

# Learning Analytics in the Era of Large Language Models

Elisabetta Mazzullo <sup>1</sup>, Okan Bulut <sup>2,\*</sup>, Tarid Wongvorachan <sup>1</sup> and Bin Tan <sup>1</sup>

<sup>1</sup> Measurement, Evaluation, and Data Science, University of Alberta, Edmonton, AB T6G 2G5, Canada; mazzullo@ualberta.ca (E.M.); wongvora@ualberta.ca (T.W.); btan4@ualberta.ca (B.T.)

<sup>2</sup> Centre for Research in Applied Measurement and Evaluation, University of Alberta, Edmonton, AB T6G 2G5, Canada

\* Correspondence: bulut@ualberta.ca

**Abstract:** Learning analytics (LA) has the potential to significantly improve teaching and learning, but there are still many areas for improvement in LA research and practice. The literature highlights limitations in every stage of the LA life cycle, including scarce pedagogical grounding and poor design choices in the development of LA, challenges in the implementation of LA with respect to the interpretability of insights, prediction, and actionability of feedback, and lack of generalizability and strong practices in LA evaluation. In this position paper, we advocate for empowering teachers in developing LA solutions. We argue that this would enhance the theoretical basis of LA tools and make them more understandable and practical. We present some instances where process data can be utilized to comprehend learning processes and generate more interpretable LA insights. Additionally, we investigate the potential implementation of large language models (LLMs) in LA to produce comprehensible insights, provide timely and actionable feedback, enhance personalization, and support teachers' tasks more extensively.

**Keywords:** learning analytics; large language model; artificial intelligence; process data



**Citation:** Mazzullo, E.; Bulut, O.; Wongvorachan, T.; Tan, B. Learning Analytics in the Era of Large Language Models. *Analytics* **2023**, *2*, 877–898. <https://doi.org/10.3390/analytics2040046>

Academic Editor: Giuseppe Pirrò

Received: 2 August 2023

Revised: 16 September 2023

Accepted: 14 November 2023

Published: 16 November 2023



**Copyright:** © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

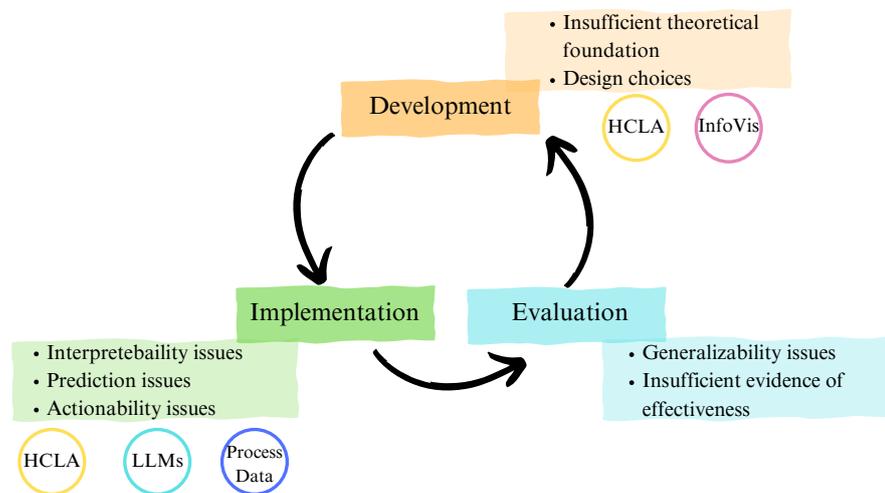
Rapid technological advancements are bringing about significant transformations in every aspect of the education system. The advent of digital learning environments (DLEs) has made large volumes of novel data available. As students interact in the DLEs, *digital traces* about learning, performance, and engagement are recorded [1]. To exploit these new forms of information and make use of computational analysis techniques, learning analytics (LA) has emerged as a new research field at the intersection of student learning, data analytics, and human-centered design. LA is defined as the “measurement, collection, analysis and reporting of data about learners and their contexts, for understanding and optimizing learning and the environment in which it occurs” [2] (p. 4). To date, many efforts in LA have been devoted to information visualization or predicting students' academic performance. The essential utilities of LA, as listed by Society for Learning Analytics Research [SoLAR] [1], include (1) promoting the development of learning skills and strategies; (2) offering personalized and timely feedback; (3) increasing student awareness by supporting self-reflection; and (4) generating empirical evidence on the success of pedagogical innovations. With the growing number of published studies focused on LA each year [3], the field of LA has been recognized for its potential to improve learning outcomes for students and educators.

Researchers have generated a vast literature on LA over the past decade. Systematic reviews of these studies identify various benefits attributed to LA systems, related not only to teaching and learning but also to management aspects and educational research [4]. For example, LA could enhance students' engagement and performance by predicting performance and identifying students at risk of failing, providing personalized feedback and intervention strategies, personalization of learning, curriculum improvement, and

course offering suggestions. In turn, these would favor better management of educational resources, improving enrollment and expense allocation. Furthermore, LA can increase our understanding of learning processes and foster the development of innovative methods for analyzing educational data [4–6].

However, there are still some areas for improvement in the development of LA systems relative to their theoretical grounding and design choices [7–9], challenges in their implementation [10], and issues in the evaluation of their effectiveness [11]. Jivet et al. [12] recognize these as critical moments in the LA life cycle, which should always be informed by learning theories to produce effective LA tools. More recently, scholars have shifted their attention to raising awareness of the importance of making teachers an integral part of the LA design process and improving the usability of LA systems based on learning theories [13]. Moreover, with the release of advanced artificial intelligence (AI) systems that can complete a variety of tasks, from memorizing basic concepts to generating narratives and ideas using human-like language, technology is revolutionizing the way we think about learning and opening up new standards for teaching practices [14]. Considering these areas for improvement in LA practices, this paper offers an overview of the current challenges and limitations and proposes directions for its future development. In particular, we encourage teacher empowerment in developing LA systems and using LA to aid teaching practices. To this end, we reflect on how process data and large language models (LLMs) can be harnessed to improve the development of LA systems and support instructional tasks.

This position paper begins by introducing various types of LA and their applications. Then, we present the challenges that modern LA practices face. Figure 1 provides a visual overview of these limitations, situating them within the LA life cycle, together with their proposed solutions. Insufficient grounding in learning sciences and poor design choices during the development of LA systems exacerbate issues in the interpretability of LA insights, which add to further challenges in their implementation related to prediction and actionability of feedback. Lastly, the evaluation of LA solutions brings forward issues related to their generalizability and scarce evidence of their effectiveness. From the issues presented, we put forward our recommendations based on the existing literature to involve teachers as LA designers for interpretable pedagogy-based LA systems. We also recommend using process data and natural language processing (NLP) to enhance the interpretability of LA. After that, we discuss how natural language models and their larger variants, like ChatGPT, can increase LA personalization and support teaching practices. We conclude the paper by discussing how the posited recommendations can enhance LA practice as a whole.



**Figure 1.** Limitations in the LA life cycle and proposed solutions. Note: Circles represent the proposed solutions. HCLA stands for human-centered learning analytics. InfoVis stands for information visualization. LLM stands for large language models.

## 2. LA: Limitations and Ongoing Challenges

This section briefly presents the different scopes of existing LA systems and illustrates the weaknesses of current research and practices in this field. It is essential to acknowledge that the limitations discussed in this section, while inherently challenging and may sound detrimental, represent invaluable opportunities for investigation and potential influence on the advancement of LA and the broader landscape of modern education.

### 2.1. Descriptive, Predictive, and Prescriptive LA

Insights from LA systems are often communicated to stakeholders through LA dashboards (LADs), which is “a single display that aggregates different indicators about learners, learning processes and learning contexts into one or multiple visualizations” [15] (p. 37). LADs can display multiple types of information. Descriptive analytics show trends and relationships among learning indicators (e.g., grades and engagement compared to peers). Descriptive dashboards typically provide performance visualizations and outcome-focused feedback [16]. Researchers use modern computational techniques to analyze educational data, not only to determine student performance but also to understand why they performed as they did, what their expected performance is, and what they should do next. Predictive LA systems utilize machine learning algorithms to analyze current and past data patterns to predict future outcomes. These systems, or LADs, are mainly used to forecast academic outcomes, such as grades in upcoming assignments and final exams, and the likelihood of non-submission, course failure, or similar results. As further explored below, predictive analytics come with their own set of technical limitations and ethical challenges. More recently, there has been a shift towards creating prescriptive dashboards offering process-oriented feedback: actionable recommendations pointing students to what they should be doing next to reach their learning goals [16–19]. Examples of similar systems can be found in the “call to action” emails employed by Iraj et al. [20], or in a LAD providing students with content recommendations and skill-building activities [11].

### 2.2. Insufficient Grounding in Learning Sciences

Researchers have criticized existing LA systems for their insufficient grounding in the learning sciences and called for a better balance between theory and data-driven approaches [7,8]. Most studies took a data-driven approach at the beginning of LA investigations without utilizing specific learning theories to guide their analysis. While this approach allowed for identifying behavioral patterns, interpreting and understanding them remained problematic [8]. The exact definition of LA identifies measurement and analytics not as the goal itself but as a “means to an end” [21], which is the understanding and optimization of learning and educational environments. This implies that the data are meaningful only to the extent that they support interpretation and guide future actions.

Of the 49 articles included in their literature review, Algayres and Triantafyllou [22] found that only 28 presented a theoretical framework, primarily referring to theories of self-regulated learning. Similarly, a scoping review of LA articles published from 2016 to 2020 revealed that 37 studies utilized the most common theories, namely self-regulated learning and social constructivism [8]. The authors invite researchers to explore behavioral and cognitive theories, going beyond observable behavioral log data and investigating information processing strategies (e.g., problem-solving, memory). Also, they discuss how learning theories should be used to interpret LA data and promote pedagogical advancement by validating learning designs. DLEs make an astounding amount of data available to researchers and educators. Still, without a theory, they are left astray in interpreting them and deciding which variables are valuable and should be selected for their models [23]. Furthermore, sometimes aggregate measures derived from simple indicators from process data are more informative for learning, as they better represent learning behaviors studied by educational theories [24]. Therefore, it is essential to understand the meaning of these new measures generated in the DLE and to remember that engagement does not necessarily equate to learning [25].

### 2.3. Interpretability Challenges

The interpretability of insights derived from LA is not only related to the theory underlying the data but also to the choices being made related to communication and design. LADs are located at the intersection between educational data science and information visualization. Recently, scholars have been reminding LAD researchers and developers that these instruments should not merely display data and ask students and teachers to assume the role of data scientists; instead, their main goal should be communicating the most essential information [26].

Research shows that learners' ability to interpret data may be limited, and to best support cognition, design choices should be founded on the principles of cognitive psychology and information visualization [9]. For example, coherent displays and colors can reduce visual clutter and direct attention to the essential elements for correctly interpreting the data. The usability and interpretability of LA tools are critical. In general, educators feel that LA fosters their professional development [27]; however, even if LA tools are perceived as valuable, teachers sometimes struggle to translate data into actions [28]. According to the Technology-Acceptance Model [29], the intention to use technology is influenced by its perceived usefulness and ease of use of the instrument. Therefore, even though teachers recognize the potential benefits of LA, they might avoid using dashboards if they do not feel comfortable navigating or interpreting them.

### 2.4. Prediction Issues

While they provide richer information than purely descriptive LA systems, predictive models come with their own set of limitations and ethical challenges, such as the risk of stereotyping and biased forecasts. Evidence on teacher use of predictive LA tools is also mixed. Some studies find that teachers who make more intense and consistent use of LA tools can better identify students who need additional support [30], while others do not corroborate these findings [31]. Furthermore, predictions are often generated by black-box models, lacking transparency, interpretability, and explicability [32]. These characteristics favor actionability [33], and in their absence, the utility of the system and users' trust are reduced [34]. Researchers highlight the need to improve prediction accuracy, together with its validity and generalizability [35], and advise that predictions need to be followed by appropriate actions and effective interventions to influence learning outcomes [36].

In the second edition of *The Handbook of Learning Analytics*, SoLAR provides directions for using measurement to transition from predictive models to explanatory models. The goal of LA is optimization, which goes a step further than prediction. At the same time, explanation is neither necessary nor sufficient for optimization; there has to be a causal mechanism on which students and teachers base their decisions if these actions are expected to produce specific desired outcomes [37].

### 2.5. Beyond Prediction: Actionability Issue for Automatically Generated Feedback

Researchers are now advocating for the development of LADs that inform students about how they have performed so far and how they can do better [18,32]. As it is widely recognized, feedback supports learning and academic achievement [38]. Earlier studies on feedback adopted an information paradigm, focusing on the type of information provided to learners, its precision, and the level of cognitive complexity [20,38]. More recently, the focus has shifted to feedback as a dialogic process and its actionability: students (and teachers) are not passive recipients of information. However, it is crucial to develop their abilities to understand feedback and take action [39]. For feedback to be effective, learners need to understand the information, evaluate their own work, manage their emotions related to the feedback, and take appropriate actions [39,40].

Another important characteristic of good feedback is timeliness. Research shows that the effectiveness of feedback is more significant when it is received quickly [41]. LA tools can offer instructors and learners constant access to automatic-generated feedback and real-time performance monitoring. Iraj et al. [20] found early engagement with feedback

to be positively associated with student outcomes when instructors used an LA tool to monitor students' progress and send personalized weekly emails that provided learners with feedback on their activity and highlighted the actions required next in their learning through "call to action" links to task materials. Prescriptive information is appreciated by students [16,17] and seems to support student motivation [42]. However, emerging prescriptive dashboards often rely on human intervention or employ automated algorithms based on hard-core heuristics and thresholds, so some researchers call for developing more sophisticated systems [18].

Moreover, the effectiveness of feedback is influenced by student characteristics [38,43], and it is enhanced when feedback is personalized [20]. LA tools can offer instructors and learners continuous access to automatically generated feedback and real-time monitoring of their performance, and they offer new opportunities to provide individualized feedback to students. Technology-mediated feedback systems have been found to increase students' engagement, satisfaction, and outcomes [44,45]. By favoring personalization and timeliness of feedback, together with the display of adequate and actionable information, student-facing LADs could help reduce the "feedback gap" [20], the difference between the potential and actual use of feedback [44,46]. However, an extensive literature review from Matcha et al. [7] suggests that existing dashboards, with their scarce grounding in theory, are unlikely to follow literature recommendations for best feedback.

#### *2.6. Generalizability Issue*

The development and adoption of LA tools are complex and require intense efforts in terms of time and expertise. Therefore, learning institutions often assume a "one-size-fits-all" approach, creating a single tool applied across every course, discipline, and level. There has been an increase in the offering of LA tool packages that use the same off-the-shelf algorithms for all modules, disciplines, and levels [47]. However, "trace data reflects the instructional context that generated it and validity and reliability in one context is unlikely to generalize to other contexts" [37] (p. 22). The Gašević et al. [48] study demonstrates that LA predictive models must "account for instructional conditions", as generalized models are far less powerful than course-specific models to guide practice and research. The literature review by Joksimović et al. [49] on LA approaches in massive open online courses highlights the lack of generalizability of these studies, as they adopt a widely different range of metrics to model learning. They suggest that a shared conceptualization of engagement by finding generalizable predictors could make results from future research more comparable across different contexts, and they invite a shift from observation to experimental approaches.

#### *2.7. Insufficient Evidence of Effectiveness*

Reviews of the literature highlight the lack of rigorous evaluations of the effects of LA tools [50]. The literature review by Bodily and Verbert [50] on student-facing LADs shows that more research is needed to understand the impact of LADs on student behavior, achievement, and skills, as the studies conducted are few and yielded mixed results. They encourage the adoption of more robust research methodologies, such as quasi-experimental studies and propensity score matching, and the investigation of underdeveloped topics, such as how students engage and interact with LADs and the evaluation of their effectiveness. Quantitative findings supporting the positive effects of LADs on learning outcomes are starting to emerge [18]; however, most of the literature consists of studies that tend to consider few outcome measures and to evaluate usability aspects, using small samples and adopting mainly qualitative strategies of inquiry [18,50]. Jivet et al. [12] advise that, in evaluating LA solutions, usability studies should investigate the tool's perceived ease of use and utility and how users interpret and understand the outputs they receive. However, these aims must remain secondary to assessing whether the intended outcomes were achieved by LA and to evaluating their affective and motivational effects. The authors suggest strengthening the evaluation of LA by triangulating data from validated

self-reported measures, assessments, and tracked data. When assessing the effectiveness of LA systems, it is essential to consider not only the outcomes but also the learning process itself. As explored below, researchers may use diversified data types collected by the LA system to offer valuable insights into learner activities, such as video logs, fine-grained click streams, eye-tracking data, and log files. These types of data allow for extracting meaningful patterns and features that can help understand learners' intermediate states of learning and how they are related to the learning outcomes [51].

### 2.8. Insufficient Teacher Involvement

Teachers are among the most critical stakeholders in integrating LA systems in schools. Thus, the effectiveness of LA systems is very much dependent on the acceptance and involvement of teachers [52]. As mentioned above, although teachers usually hold positive attitudes toward LA [27], they are also identified as a potential source of resistance to the adoption of these new systems [53,54]. Surveys reveal that in 2016, LA initiatives were primarily driven by IT experts and a few dedicated faculty members in Australia and the UK. Still, for the most part, teachers were left "out of the loop" of these novel initiatives [47]. However, teachers may develop a negative attitude toward LA systems and be reluctant to utilize them if they perceive them as lacking usefulness or ease of use [55]. Therefore, it is vital to understand and address teachers' needs and tolerance for complex systems. Moreover, teachers represent not only the end users of LA systems but also content experts in their subjects and classrooms. As educators orchestrate the teaching and learning process, they should be called to take part in designing the learning tools they will be expected to adopt. Involving teachers as designers in the development of LA systems would help create a bridge between data and theory by integrating the teachers' learning design and aid design choices that support the usability and readability of dashboards.

## 3. Moving Forward in LA

The previous sections highlighted the most critical gaps in present LA research. Although LA offers excellent potential to the educational field, clear guidelines for LAD development and robust evaluation procedures are still lacking. Scarce grounding in learning theories, lack of generalizability, and subsequent scalability challenges have generated a rather large body of literature from which it is hard to draw interpretations and conclusions on the effectiveness of LA tools. Moreover, an excessive focus on data and insufficient involvement of teachers and students in the design of these systems created dashboards that are too disconnected from the instructors' learning designs, users' needs, and data literacy abilities, leading to usability and interpretability challenges.

Such limitations must be acknowledged as they lead to venues for improvement that could enhance LA practice in several aspects. This section presents some approaches that could offer valuable guidelines for the future developments of LA and enhance their implementation. Some of these approaches have started to be adopted in the literature; however, as acknowledged as a limitation of our paper, some of the ideas must still be developed and tested to verify their educational effectiveness and their actual value in improving LA.

### 3.1. Involving Teachers as Co-Designers in LA

Human-centered learning analytics (HCLA) [13] proposes to overcome some limitations of LA through the participatory design of LA tools. Engaging stakeholders as co-creators holds the potential to develop more effective tools by transforming LA from something done to learners into something done with learners. This shift could lower ethical concerns and lead to the development of tools that better fit the needs of their users. For example, this perspective is switching the focus from relying entirely on users in data interpretation to giving them the answers they are interested in.

Dimitriadis et al. [56] identified three fundamental principles of HCLA: (1) theoretical grounding for the design and implementation of LA; (2) intensive inter-stakeholder

communication in the design process; and (3) the integration of LA into every phase of the learning design cycle to “support teacher inquiry into student learning and evidence-based decision-making”. During LA design, the target of the LA tools should be derived from the learning design; then, the implementation of the LA tools can provide valuable insights to inform the orchestration of learning and the evaluation of the learning design itself. Finding a way to hear all stakeholders’ voices can be challenging; to facilitate the orchestration Prestigiaco et al. [57] introduce OrLA, which provides a roadmap to guide communication. Through the participatory design and the active involvement of teachers not only as end users but as designers and content experts, HCLA could favor a scalable implementation of LA and lead to the development of instruments that fit teachers’ data abilities and needs [58]. Similar principles remain valid when broadening the discussion from LA to educational AI in general. Cardona et al. [59] identify three instructional loops in which cooperation between AI and teachers should always center on educators: the act of teaching, the planning and evaluation of teaching, and the design and evaluation of tools for teaching and learning.

An increasing number of studies have started to implement methods of participatory and co-design in the development of LA dashboards [60]. Examples of LA developed in cooperation with teachers can be found in the work of Pardo et al. [61] and Martinez-Maldonado et al. [62]. The tools developed for these studies allow educators to set “if-then” rules that reflect their learning design and influence the output returned from the analysis of the various data sources used by the system (i.e., semi-automated emails for process-oriented feedback and data stories, respectively). Interviews with educators revealed that they liked to be able to see the rules and modify them, and some proposed showing them to students during in-class debriefings so that they could understand the difference between their performance and the learning expectations [62]. Conijn et al. [63] present the iterative procedure they used to develop a dashboard that provides interpretable and actionable feedback about students’ writing process. The steps included the cooperation of writing researchers and teachers for the design of the tool and usability tests with new teachers, which pointed to the effectiveness of the approach.

### *3.2. Using Natural Language to Increase Interpretability*

To reduce reliance on users’ data literacy for LADs interpretation and support the inference process, Alhadad [9] suggests integrating textual elements into visualizations, for example, through narrative and storytelling aspects. The incorporation of storytelling in LA visualization was introduced by Echeverria et al. [64]. The authors advocate for the explanatory instead of the exploratory purpose of LA: dashboards should not invite the exploration of data, but rather explain insights. They propose a learning design-driven data storytelling approach, which builds on principles from information visualization and data storytelling and, in accordance with HCLA, connects them to teachers’ intentions (i.e., learning design). Contrary to traditional “one-size-fits-all” data-driven visual analytics approaches, the new method derives rules from the learning design and uses them to construct storytelling visual analytics. Data storytelling principles determine which visual elements should be emphasized, while the learning design determines which events should be the focus of communication.

Fernandez Nieto et al. [65] explored the effectiveness of three visual-narrative interfaces built on three different communication methods: visual data slices, tabular visualizations, and written reports. From interviews with educators, it emerged that different methods are more helpful for different purposes. For example, written reports were perceived as beneficial for teachers’ reflection but not as much to be used in students’ debriefings, for which tabular visualizations were thought more appropriate. Therefore, defining the purpose of the LAD and involving stakeholders in this process seems to be fundamental for developing effective dashboards.

To incorporate textual elements into LADs, Ramos-Soto et al. [66] developed a service that uses natural language templates and data extracted from the DLE to automatically

generate written reports about students' activity. According to the evaluation of an expert teacher, the system was able to generate practical and overall truthful insights, albeit with small divergences and not as complete as those that would have been derived from the data by human experts.

Natural language generation could automate the production of verbal descriptions and data stories to facilitate and guide the interpretation of charts and infographics generated by LA systems. While no system is mature enough to be trusted on its own, research in the field is moving fast. Sultanum and Srinivasan [67] recently developed DataTales, an LLM-powered system to support authoring narratives about any given chart. The system does not simply tell users what is conveyed by the data but also helps them read the chart: when the user hovers over a particular portion of text, an interactive visualization highlights the relevant elements of the graph. The prototype was evaluated through interviews with data experts. Participants found the tool effective in assisting both data explanation ("what to talk about and how" (p. 3)) and exploration (including to "get a high-level summary of the data in natural language form" (p. 4)) and extracting insights ("the why's" (p. 3)) from the data. Although responses were mostly positive, some issues were identified, including style, lengthiness, and wrong or inaccurate interpretations. Even though the technology is not perfect, it can offer us a glimpse into the future; or, even in its flawed state, it could be used in an expert-led environment to support the development of data literacy abilities of teachers and students.

### *3.3. Using Process Data to Increase Interpretability*

In recent years, with the popularity of learning systems, researchers have been interested in the process data; that is, the data generated while students interact with the learning systems. In a learning system, students' interactions with the user interface, including their duration on each screen and actions such as clicking, are often logged and commonly referred to as process data [68]. However, process data encompasses more than log data; it broadly includes empirical data that indicates the process of working on a test item based on cognitive and non-cognitive constructs [69]. This encompasses various data types, such as action sequences, frequency of actions, conversations or interactions within the learning system, and even eye-tracking movements and think-aloud data. In recent years, process data have received extensive research attention within the context of educational data mining, learning analytics, and artificial intelligence. Process data serves as a valuable source of detailed information regarding students' learning process within a learning system, enabling interpretation of both cognitive and behavioral aspects of learning.

As an important aspect of process data, response time has been extensively studied and is commonly regarded as an indicator of students' behaviors and cognitive processes. For example, the response time has been used to identify students presenting abnormal behaviors in assessments. Wise and Ma [70] proposed a normative threshold method that compares an examinee's response time with that of their peers to determine rapid guessers or disengaged test-takers. In addition, Rios and Guo [71] developed a mixture log-normal approach which assumes that, in the presence of low effort, a bimodal response time distribution should be observed, with the lower mode representing non-effortful responding and the upper mode indicating effortful responding. This approach employs an empirical response time distribution, fits a mixed log-normal distribution, and identifies the lowest point between the two modes as the threshold. A more straightforward yet effective method is to visually inspect bimodal response time distributions for a distinctive gap, which can differentiate rapid guessers from other test-takers [72]. These methods can also be extended beyond assessment environments to infer students' motivation, engagement, and learning experiences by analyzing their time spent navigating learning systems.

In addition to response time, clickstream data recorded during test-taking experiences can provide valuable insights into behavior patterns. For example, Su and Chen [73] utilized clustering techniques to group students' clickstream data with similar behavior

usage patterns. Ulitzsch et al. [74] considered both action sequences and timing, employing cluster edge deletion to identify distinct groups of action patterns that represent common response processes. Each pattern describes a typical response process observed among test-takers. Furthermore, Tang et al. [75] introduced the model agreement index as a measure to quantify the typicality or atypicality of an examinee's clickstream behaviors compared to a sequence model of behavior. To achieve this goal, they trained a Long Short-Term Memory network to model student behaviors. This approach allows the model to incorporate various behavior patterns and acquire knowledge about normal behavior patterns across different test-taker archetypes and styles. Gao et al. [76] used fine-grained log data to capture students' progress in a programming class. Using differential sequence mining on data from the first assignment, they could predict the final course outcome with 79% accuracy and capture interpretable behavioral patterns that reflect effective and ineffective strategies that students enact to learn. For example, specific coding patterns frequent among low performers were interpreted by all researchers as indicative of unsystematic actions performed without taking time to think and of uncertainty.

Biometric measures, such as analyzing eye movements, can also offer valuable insights into students' learning and test-taking behaviors. The duration of eye fixation can reflect the level of attention a test-taker pays to specific words in test items, with more challenging items generally requiring longer fixation periods [77]. Pupil size can indicate fatigue levels, interest in specific learning content, and the cognitive workload associated with a particular task [78]. Moreover, blink rates tend to decrease when there is a higher visual demand, indicating the reallocation of cognitive resources [79]. For instance, research studies have demonstrated that when individuals encounter unfamiliar, ambiguous, or complex items, they tend to increase their regression rate, which means they look back at previous parts of the text to reinstate or confirm their cognitive effort [80,81]. Furthermore, such regression has been strongly linked to the level of effort and attention a reader devotes to a reading task. Thus, increased regression is often associated with improved accuracy in processing the content information [82].

The intermediate states of students' problem-solving or writing processes within the learning system can also be analyzed. For example, Adhikari [83] proposed several process visualization practices for writing and coding tasks in learning systems, such as the playback of typing and tracking changes in paragraphs, sentences, or lines over time. By employing these visualization practices, educators can directly see (1) the specific points in the process where students spent the majority of their time, (2) the distribution of time between creating the initial draft and revising and editing it, (3) the paragraphs that underwent editing and revision, and (4) the paragraphs that remained unedited. These visualizations allow educators to explore, review, and analyze students' learning processes and their approach to writing or programming. In addition, students themselves can leverage these visualizations for self-reflection, direction, and improvement. Furthermore, the temporal analysis of keystrokes and backspaces provides insights into learners' engagement [84] and affective states [85]. Allen et al. [86] encourage the exploration of additional aspects of the online language production process, such as pausing typing to check syntax or research the vocabulary.

Another example of where process data proves useful is in identifying and interpreting the patterns of action sequences associated with different learning or testing outcomes. For instance, He and von Davier [87] combined sequence mining with n-gram techniques to pinpoint common patterns leading to either successful or unsuccessful action sequences. Their findings revealed that the patterns of action sequences linked to correct responses are more consistent across countries than those linked to incorrect responses. Extending this line of research, Ulitzsch et al. [88] incorporated graph-based data clustering to identify how, and in which aspects, the patterns of action sequences related to correct responses differ from those related to incorrect responses.

Moreover, NLP techniques can be employed to analyze process data. For example, Guthrie and Chen [89] analyzed log data from an online learning platform and introduced

a novel approach to modeling student interactions. They incorporated information about logged event duration to differentiate between abnormally brief events and normal or extra-long events. These new event records were treated as a form of language, where each word represented a student's interaction with a specific learning module, and each sentence captured the entire sequence of interactions. The authors used second-order Markov chains to identify patterns in this new language of student interactions. By visualizing these Markov chains, the authors found the interaction states associated with either disengagement or high levels of engagement. However, LLMs have been rarely applied for log analysis. To address this gap, Chhabra [90] experimented with several BERT models to establish a system for automatically extracting information (i.e., the events occurring within a system) from log files. In contrast to traditional log parsing approaches that heavily relied on humans constructing regular expressions, rules, or grammars for information extraction, the proposed system significantly reduced the time and human effort required for log analysis. This work demonstrates the potential of using LLMs to extract and analyze the logged events collected through LA systems, thereby improving the ease of interpreting students' learning process.

According to the showcased examples above, process data can increase the interpretability of students' learning process, and including this type of data in LA systems could lead to the generation of more interpretable insights. Identifying the concrete behavioral patterns that underlie learning processes can bring to light the strategies students adopt and prompt teachers and learners to reflect on their effectiveness and what they might need to do differently to improve their performance.

### *3.4. Using Language Models to Increase Personalization*

Our review of the literature identifies timeliness [41], personalization [20], and actionability [39] as attributes of effective feedback, which would support effective implementation of LA. A thematic analysis of learners' attitudes toward LADs reveals that students are interested in features that support learning opportunities: they express a wish for systems that provide everyone with the same opportunities and, at the same time, a desire for customization to deliver meaningful information. They demonstrate awareness of privacy concerns and prefer automated alerts over personalized messages from teachers. This might be because the latter elicits feelings of surveillance [91]. Automatically generated personalized feedback could provide the benefits of customized messages without making students feel monitored by their teachers.

A literature review on automatic feedback generation (AFG) in online learning environments [92] points to the usefulness of this technology, with about half of the studies indicating that AFG enhances student performance (50.79%) and reduces teacher effort (53.96%). The main techniques used in generating feedback were comparison with a desired answer, dashboards, and NLP.

NLP analyzes language in its multi-dimensionality and delivers insights about both texts and learners. Descriptive features of language (e.g., number or frequency of textual elements) can inform about student engagement or be used for predicting task completion or identifying comparable texts. Characteristics of the lexicon employed in a text can be used to classify genres or estimate readability. The syntactical structure of sentences informs about the readability, quality, and complexity of the utterance, and can be used to evaluate linguistic development. Semantic analyses can identify the central message of the text and its affective connotations, or detect overlap between two texts (e.g., original text and summary). NLP analyses can also estimate cohesion and coherence, which inform how learners process and elaborate knowledge. Moreover, NLP can communicate with teachers and learners through natural language, for example, through the generation of reports or personalized feedback [86].

Cavalcanti et al. [92] notice how existing studies on AFG are plagued by two of the limitations that have already been highlighted in this paper: insufficient grounding in educational theories for effective feedback and a lack of consideration for the role

of teachers in the provision of feedback. Therefore, they encourage further research to evaluate feedback quality and develop tools focused on instructors. Moreover, they call for studies on the generalizability of systems for AFG, identifying a possible solution in natural language generation.

LLMs are advanced NLP models that use deep learning techniques to learn patterns and associations between the elements of natural language and capture statistical and contextual information from the training data. The models are usually trained on vast databases encompassing various textual data sources, such as books, articles, and web pages. LLMs are not only able to understand language but also to produce coherent human-like utterances in response to any user-generated prompt. LLMs can translate, summarize and paraphrase a given text, and generate new ones. With the release of ChatGPT in November 2022, LLMs gained huge traction in society and across numerous fields, from medicine to education, as scholars explore the applications of these new systems and warn about their pitfalls. In fact, even though the training corpora is massive, it is not always accurate or up to date, which means that sometimes outputs generated by ChatGPT can be inaccurate or outdated. For example, there are records of LLMs providing links to unrelated sources or citing nonexistent literature [93]. These are examples of hallucinations, which are only one of the unresolved challenges in LLM research [94]. Moreover, LLMs are not (yet) great at solving math problems [95].

Lim et al. [96] invite researchers to develop LA systems that can make feedback more dialogue-based; personalized feedback messages should go a step further and include comments on learning strategies (i.e., metacognitive prompts) to support sense-making, as understanding feedback and interpreting it in relation to one's own learning process is necessary to plan appropriate action in response to the feedback.

Dai et al. [97] provided ChatGPT with a rubric and asked it to produce feedback on student assignments to compare it against instructor-generated feedback. The AI tool produced fluent and coherent feedback, which received a higher average readability rating than the ones written by the teacher. Agreement between the instructor and ChatGPT was high on the evaluation of the topic of the assignment; however, precision was not as satisfactory on the evaluation of other aspects of the rubric (goal and benefit). ChatGPT generated task-focused feedback for all the students and provided process-focused feedback for just over half of the assignments. On the other hand, the AI never gave feedback on self-regulation and self, while the instructor provided similar feedback in 11% and 24% of cases, respectively.

Similarly, Matelsky et al. [98] developed FreeText, a model-agnostic framework that can leverage any specific LLM to provide students with timely and individualized feedback on short answers to open-ended questions. The system does not assign grades but offers textual feedback on the overall answer and specific snippets of the response that might contain errors or inaccuracies. Teachers have the option to set evaluation criteria, which the tool can also use to present them with improved versions of the question prompt. FreeText is intended to support both teachers and students, not to fully automate assessment, and it should soon be tested in a large-scale context.

Yildirim-Erbasli and Bulut [99] discussed the potential of conversational agents in improving students' learning and assessment experiences through continuous and interactive conversations. The authors argue that conversational agents can create an interactive and dynamic learning and assessment system by administering tasks or items and offering feedback to students. The use of NLP enables conversational agents to provide real-time feedback that adapts to students' responses and needs, fostering a more effective and engaging learning environment. Consequently, students' motivation and engagement levels in learning and test-taking can be continuously boosted through personalized conversations and directed feedback.

Hasan et al. [100] recently introduced SAPIEN, a highly customizable, high-fidelity virtual agent powered by LLMs, able to engage in dynamic video-call conversations in 13 languages and to adapt vocal and facial expressions across the range of seven basic emotions.

Users can set the demographic characteristics of the avatar, choose the topic and goal of the conversation, and obtain feedback at the end of the video call. The authors suggest a wide range of applications for the tool, including language learning. They demonstrate awareness about the ethical risks linked to a virtual agent so highly humanized, and in response, they set short limits to the length of the call and the information retention capabilities of the tool. SAPIEN offers an example of what can be achieved when LLMs are coupled with other technologies (i.e., animations, speech-to-text, and text-to-speech models). While conversational agents for education do not necessarily need to be realistic, longer attention spans would likely be more beneficial than customizable humanoid avatars. Educational researchers should explore what could be achieved when LLMs are integrated into educational systems such as LA tools or intelligent tutoring systems.

Another aspect of the personalization potential of LLMs is that they can be utilized to generate learning tasks or assessment items that are optimally tailored to individual student abilities. For example, LLMs have been employed to automatically generate a variety of learning and assessment materials, including reading passages [101,102], programming exercises [103], question stems [104,105], and distractors [106,107]. These examples demonstrate the potential of LLMs to create large item banks. The automatically generated assessment items then can be integrated into the existing framework of computerized adaptive tests, a testing methodology that adapts the selection of the following item based on the student's ability level inferred from their previous responses [108]. As a result, students can engage in a personalized and adaptive learning experience, thereby enhancing their engagement and improving learning outcomes [109].

With a large item bank or a bank of learning tasks created, LLMs can be further used to build recommender systems. Recommender systems in education aim to offer personalized items that match individual student preferences, needs, or ability levels, helping them navigate through educational materials and optimize their learning outcomes. Typically, there are two popular approaches for building recommender systems: collaborative filtering and content-based filtering. The underlying idea of collaborative filtering is to analyze students' past behavior and preferences to generate recommendations, identifying patterns and similarities between users or items. It assumes that students who have exhibited similar interests will continue to do so in the future. On the other hand, content-based filtering examines the content of the items and compares them to students' profiles or past interactions. By identifying similarities between the content of items and students' preferences, needs, or ability levels, the system can generate recommendations that match students. In the era of LLMs, language model recommender systems have been proposed to increase transparency and control for students by enabling them to interact with the learning system using natural language [110]. LLMs can interpret natural language user profiles and use them to modulate learning materials for each session [111]. For example, Zhang et al. [112] proposed a language model recommender system leveraging several language models, including GPT2 and BERT. They converted the user-system interaction logs (items watched: 1193, 661, 914) to text inquiry ("the user watched <item name> of 1193, <item name> of 661, and <item name> of 914") and then used language models to fill in the masks for recommendation ("now the next item the user wants to watch is \_\_\_\_\_").

Therefore, integrating LLMs into LA systems could generate more effective tools by affording students a highly personalized learning experience and providing detailed and timely verbal feedback on their performance and progress. Moreover, by automating feedback generation, LLMs can relieve teachers from demanding and time-consuming tasks, allowing them to devote more time to other aspects of teaching. However, it is always important to keep teachers in the loop in the process of the generation and provision of automatic feedback, as these systems are not (yet) able to touch upon all the dimensions of learning and might not integrate the learning design or take the student history into account. Some existing LA systems offer an instructor-mediated approach to personalized feedback, which offers teachers greater control over the metrics and messages returned to students, for example, by allowing them to set up "if-then" rules for message delivery

based on their specific learning design. A focus group exploring students' perception of a similar system reveals that, even if they knew that the messages were, to some extent, automated, pupils perceived that their instructor cared about their learning. The authors argue that the perception of interpersonal communication favored proactive recipients of feedback and increased motivation for learning [113]. Cardona et al. [59] support the use of AI for AFG but recommend always keeping educators at the center of the feedback loops and invite researchers to create feedback that is not solely deficit-focused but also asset-oriented, able to help students recognize their strengths and build onto them.

### 3.5. Using Language Models to Support Teachers

Feedback generation is only one of the many possible applications of LLMs to support educational practices. Allen et al. [86] suggest that when applying NLP to LA, we should consider both the multi-dimensional nature of language and the multiple ways in which language is part of the learning process. Language permeates every aspect of learning: it is through processing natural language that learners are asked to understand course materials and tasks (input), explain their reasoning (process), and formulate their responses (output). NLP can be leveraged to analyze the learning process in all its different phases. At the input level, NLP can inform teachers about how their communications and the materials they select impact students and also identify the most appropriate materials for each student based on their reading abilities and vocabulary skills. To understand cognitive processes underlying learning, NLP techniques can automate the analysis of think-aloud protocols and open-ended questions in which respondents describe their reasoning. Lastly, NLP can analyze textual outputs produced by students with different objectives, such as automated essay scoring (AES), assessing students' abilities (e.g., vocabulary skills) and understanding of the course content, and providing highly personalized feedback.

Bonner et al. [114] provide examples of practical uses of LLMs to alleviate teachers' workload and free up time to focus on learners while creating engaging lessons and personalized materials. LLMs such as ChatGPT can correct grammar and evaluate cohesion in student-generated texts, summarize texts, generate presentation notes from a script, offer ideas for lessons and classroom activities, create prompts for writing exercises, generate test items, write or modify existing texts into suitable assessment materials based on skill level, and guide teachers in the development of teaching objectives and rubrics. By crafting well-thought-out and specific inputs, teachers can receive outputs that best fit their intent and meet their needs. For example, teachers can specify how many distractors to be included in the multiple choice questions generated by the AI, what writing style should be used, or how difficult the text should be. When asked to provide ideas for classroom activities to introduce a topic, ChatGPT proposed tasks that span across the taxonomy of learning, from analyzing to applying, depending on the students' skill level that the activity was thought for. Through LLMs, teachers can create personalized materials for each student in a fraction of the time it would take them to do so themselves.

AI technologies could enhance the practices of formative assessment by capturing complex competencies, such as teamwork and self-regulation, by promoting accessibility for neurodivergent learners, or by offering students constant support whenever needed, even outside of class times [59]. For example, LLMs can be used to build virtual tutors that can help learners understand concepts, test their knowledge, improve their writing, or solve assignments. Khanmigo is a virtual tutor developed by Khan Academy that uses GPT-4 to support both students and teachers in many of the ways presented above. The system was instructed to tutor students based on the best practices identified by the literature, which means it supports and guides student reasoning processes without doing the assignment for them, even when asked to do so. Chat logs are made available for teachers to access, and inappropriate requests (e.g., cheating) are automatically flagged by the system and brought to the educator's attention [115].

All these applications are anticipated to reduce teachers' workload, either by taking it on themselves (e.g., modifying a text so that it meets the appropriate difficulty level for

learners) or by offering educators guidance and ideas (e.g., planning classroom activities). Users are encouraged to be specific when providing prompts and to keep interacting with the LLMs, giving them further instructions if they are not satisfied with the answer they received, as these systems retain a more or less extensive memory of the conversation (context window). Increasing efforts have been recently focused on enlarging the mnemonic capabilities of LLMs, which would be useful to approach complex tasks, such as summarizing entire books or keeping the memory of each student's background and interests.

#### 4. Discussion

Although the literature has received numerous contributions over the last few years, there are still limitations in the development and design of LA tools and challenges in their implementation. Existing LA applications still suffer from an insufficient grounding in pedagogical theories, leading to difficulties in the valid interpretation and use of the learner data. Moreover, the generalizability of LA models is still sub-optimal in terms of performance, and substantial evidence of LA effectiveness is lacking due to mixed results and the paucity of evaluation studies making use of strong research methodologies. Educators' overall attitude towards LA tends to be positive, but they still face challenges in adopting LA tools. An excessive focus on data, detached from learning theories and the teacher's learning design, together with poor design choices, can create tools that do not meet their end users' needs and data literacy abilities.

Making teachers co-designers in the development of LA seems to be a promising route to integrate pedagogical theories and the teachers' own learning design with the behavioral data collected in the DLE. The collaborative design process proposed by HCLA should yield tools that better meet the context need, better enable teachers to interpret insights, and better meet their data literacy skills.

Another promising way to aid users in interpreting LA data is to integrate visual information with written text. Natural language is central to communication, permeating every aspect of teaching and learning. With the recent and fast evolution of language models, a plethora of new opportunities are opening up in the educational field. LLMs can be used to evaluate assignments, provide personalized feedback on students' essays and their progress, or offer support as an ever-accessible tutor. Furthermore, LLMs can support teachers in various other tasks, from adapting learning materials to their students' language proficiency levels to developing creative activities, learning plans, essay prompts, or questions for testing. LLMs should not be embraced as the solution to all problems in education and LA; they could be effective in increasing interpretability and personalization of LA insights but cannot address the foundational issues related to the development of LA systems and the investigation of their effects.

Integrating LLMs into LA could make insights more interpretable for users, and integrating LA into LLMs could give the language model the context necessary to offer each student highly personalized and better-rounded feedback that takes their history, progress, and interests into account when providing reports and recommendations. Moreover, LLMs can serve educators as support tools to approach complex and time-consuming teaching tasks. The intent is not to use AI to replace teachers but to put technology at the service of teachers. Educators should use LLMs as a resource to reduce workload, stimulate creativity, and offer students tailored materials and more feedback while retaining their role as reference figures and decision-makers in the planning and evaluation of learning. AI is not supposed to strip teachers of the value of their expertise but rather to support it and allow them to focus on tasks in which the human factor cannot be replaced.

Contrary to the interpretation of LA data, LLM outputs are generally as straightforward as possible since the systems communicate directly through natural language. In this regard, one of the barriers to acceptance and usability is removed. However, integrating LLM systems into teaching practices still requires trust in the technology and an adjustment in the ways teachers have been operating until now. As applications of LLMs increasingly take hold in the educational world, we should provide educators with guidelines on how

to interact with these systems, including how to phrase their prompts to obtain the answer that best fits their needs, understanding the limitations of these tools, and being aware of risks. When used responsibly, LLMs such as ChatGPT present opportunities to enhance students' learning experience and mitigate a considerable amount of workload for teachers, for example, through assistance in the formulation of test item writing [116]. However, educators should be aware of potential issues that LLMs entail, such as over-reliance on the LLM, copyright, and cheating [116,117]. In this era of rapid technological development, a new approach to teaching practices may be necessary to revolutionize modern education and reconcile the tension between human teachers and artificial intelligence [117]. In particular, to successfully establish a safe and prolific cooperation with AI in education, we need to find a balance between the contrasting forces of human control and delegation to technology, between collecting more data to better represent students and respecting their privacy, and strive for personalization that does not cross over the line of teacher surveillance [59].

#### *Limitations and Directions for Future Research*

This study has several limitations worth noting. First, we recognize that this paper does not offer any AI- or LA-based solutions to overcome the limitations in the evaluation stage of the LA life cycle. Specifically, the issue of generalizability is an ongoing challenge for LA researchers. Inadequate feature representation, inadequate sample size, and imbalanced class are primary causes that hinder the generalizability of LA models [118]. However, such problems are commonly encountered in real-world datasets. The achievement of a shared conceptualization is hampered by patterns in the population, as both individual factors (e.g., the shift in interest) and societal factors (e.g., trends in education) could change at the sub-group level and, therefore, hinder a common feature representation. The mentioned sample size and imbalanced class issues are also hardly avoidable, as in predictive tasks that target low-occurrence but high-impact situations, such as school dropout, the discrepancy between the minority and the majority class is usually high [119]. Thus, these limitations can usually be addressed only after the fact.

Furthermore, the challenge of insufficient evidence of effectiveness cannot be addressed solely by using AI- or LA-based solutions, but it calls for purposeful choices in the development of LA tools and evaluation studies. To guide the planning of LA evaluations, we encourage future research to follow Jivet et al. [12]'s recommendations outlined above. However, future evaluation studies might employ NLP techniques and LLMs to support the qualitative analysis of teachers' and students' responses to open-ended questions about LA usability and perceived utility.

Lastly, we want to note that the effectiveness of the LLM-based solutions proposed in this paper to improve LA has not been tested yet, as the LLM-based educational tools discussed above are still under development. Furthermore, despite the potential benefits of LLM-based solutions, technology readiness remains a significant challenge. Yan et al. [120] conducted a scoping review on the applications of LLMs in educational tasks, focusing on the practical and ethical limitations of LLM applications. The authors asserted that there was little evidence for the successful implementation of LLM-based innovations in real educational practices. In addition, they noted that existing LLMs applications are still in the early stages of technology readiness and struggle to handle complex educational tasks effectively, despite showing high performance in simple tasks like sentiment analysis of student feedback [121]. Furthermore, the authors pointed out that many reviewed studies lacked sufficient details about their methodologies (e.g., not open-sourcing the data and codes used for analysis), making it challenging for other researchers and practitioners to replicate their proposed LLMs-based innovations. Based on the results, Yan et al. [120] suggest future studies to validate LLM-based education technologies through their deployment and integration in real classrooms and educational settings. Real-world studies would allow researchers to test the models' performance in authentic scenarios, particularly for tasks of prediction and generation, and to evaluate their generalizability. The authors warn

researchers that studies in educational technology tend to suffer from limited replicability. Therefore, they encourage them to open-source their models and share enough details about their datasets.

From an ethical perspective, adopting LLMs and AI-powered learning technologies in education should carefully consider their accountability, explainability, fairness, interpretability, and safety [122]. Data privacy is a primary concern for ethical AI, and information security standards should be followed at all stages of data management. Informed consent for data collection, usage, sharing, and disposal is the first essential step to ensure ethical data treatment [123]; however, often users are not aware of the extent of personal information they agree to share [124], and more concerns about individual freedom of choice arise when the use of AI-based technologies is required by the school [123]. Scholars warn that excessive surveillance can diminish learner agency, and predictive models based on student characteristics can put self-freedom at risk and perpetuate systematic biases embedded in the algorithms [125]. The majority of existing LLMs-based innovations are considered transparent and understandable only by AI researchers and practitioners. At the same time, none are perceived as sufficiently transparent by educational stakeholders, such as teachers and students [120]. To address this issue, future research should incorporate a human-in-the-loop component, actively involving educational stakeholders in the development and evaluation process. This also ensures that the educational stakeholders gain insights into how LLMs and AI-powered learning technologies function and how they can be harnessed effectively for improved learning outcomes.

Future studies could further explore the application of NLP techniques to analyze process data and generate written reports from students' data. Although LLMs have only recently been developed and numerous challenges remain to be solved, researchers both inside and outside of academia are hastily at work to address them and improve these models, and as the capabilities of LLMs expand, so will their applications [94]. For example, expanding LLM context windows would support the provision of feedback that takes into account students' background information, such as individual interests and level of language proficiency [126]. Further, LLM could also be used with an intelligent tutoring system to enhance the quality of feedback provided to students [127]. Moreover, as LLMs will find their way into teaching and learning practices, further consideration should be given to the ethical implications of AI in education. Data privacy and transparency concerns call for higher model explainability and greater involvement of stakeholders in developing and evaluating educational technologies. Moreover, while the high level of personalization that LLMs could offer students might increase equity, the costs currently associated with developing and adopting these technologies raise issues about equality. Additional concerns involve model accuracy, discrimination, and bias [120]. Researchers, policymakers, and other educational stakeholders should consider what they can do to mitigate these threats to fairness and ensure that educational AI will not broaden inequalities instead of reducing them.

## 5. Conclusions

While LA holds many promises to enhance teaching and learning, there is still work to be done to bring them to full fruition. The present paper highlighted the areas for improvement in the development, implementation, and evaluation of LA and offered guidelines and ideas that could be tested to overcome some of these challenges. In particular, there is a need for incorporating data and learning theories, as these would provide a lens to make sense of LA insights. HCLA offers principles to reach this integration through intensive cooperation with educators as co-designers of LA solutions. In addition, using process data in LA systems can enhance our understanding of students' learning processes and increase the interpretability of insights. Furthermore, we explored numerous ways in which LLMs can be deployed to make LA insights more interpretable and customizable, to increase personalization through feedback generation and content recommendation, and to support teachers' tasks more broadly while always maintaining a human-centered approach.

By raising awareness about areas for improvement and highlighting the tools that the literature and recent technological innovations are providing, we hope this paper can inspire further efforts to bring LA closer to fulfilling its potential. Future research should strive to implement human-centered frameworks in LA development, from identifying users' needs to assuring that design choices support usability. Indicators of engagement and learning should not be defined solely by obscure algorithms but be based on shared conceptualizations funded in pedagogical theories and fitted to the instructors' learning design. Only then it becomes possible to strike a balance between generalizability and context-specificity, between prediction accuracy and interpretability. Moreover, as SoLAR reminds researchers, the utility of LA extends beyond prediction, encompassing the development of complex skills and learning strategies and the personalization of feedback. To tap into these potentials, future studies could incorporate process data into LA systems to identify concrete behavioral patterns and the underlying learning processes, giving teachers and students concrete elements to reflect upon to understand their performance and insights that could be linked to learning theories. Examples of valuable data sources are response times and action sequences, such as writing and editing processes. In the wake of the recent innovations in language models, we invite researchers to explore how LLMs can be integrated into LA to support interpretability and personalization: the limited but expanding capabilities of existing LLM-based tools presented in this paper can offer a promising starting point for researchers to improve upon and test in realistic settings. Evaluation studies are crucial to assess the effectiveness of LA and inform the community if we are moving in the right direction.

**Author Contributions:** Conceptualization, E.M., O.B., T.W. and B.T.; Methodology, E.M., O.B., T.W. and B.T.; Investigation, O.B., T.W. and B.T.; Resources, E.M., T.W. and B.T.; Writing—original draft, E.M., O.B., T.W. and B.T.; Visualization, E.M.; Supervision, O.B. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research received no external funding.

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** No new data were created or analyzed in this study. Data sharing is not applicable to this article.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Society for Learning Analytics Research [SoLAR]. What is Learning Analytics? 2021. Available online: <https://www.solaresearch.org/about/what-is-learning-analytics/> (accessed on 31 May 2023).
2. Siemens, G. Learning analytics: Envisioning a research discipline and a domain of practice. In Proceedings of the 2nd International Conference on Learning Analytics and Knowledge, Vancouver, BC, Canada, 29 April–2 May 2012; LAK '12, pp. 4–8. [CrossRef]
3. Lee, L.K.; Cheung, S.K.S.; Kwok, L.F. Learning analytics: Current trends and innovative practices. *J. Comput. Educ.* **2020**, *7*, 1–6. [CrossRef]
4. El Alfy, S.; Marx Gómez, J.; Dani, A. Exploring the benefits and challenges of learning analytics in higher education institutions: A systematic literature review. *Inf. Discov. Deliv.* **2019**, *47*, 25–34. [CrossRef]
5. Banihashem, S.K.; Aliabadi, K.; Pourroostaei Ardakani, S.; Delaver, A.; Nili Ahmadabadi, M. Learning Analytics: A Systematic Literature Review. *Interdiscip. J. Virtual Learn. Med. Sci.* **2018**, *9*. [CrossRef]
6. Avella, J.T.; Kebritchi, M.; Nunn, S.G.; Kanai, T. Learning Analytics Methods, Benefits, and Challenges in Higher Education: A Systematic Literature Review. *Online Learn.* **2016**, *20*, 13–29.
7. Matcha, W.; Uzir, N.A.; Gašević, D.; Pardo, A. A Systematic Review of Empirical Studies on Learning Analytics Dashboards: A Self-Regulated Learning Perspective. *IEEE Trans. Learn. Technol.* **2020**, *13*, 226–245. [CrossRef]
8. Wang, Q.; Mousavi, A.; Lu, C. A scoping review of empirical studies on theory-driven learning analytics. *Distance Educ.* **2022**, *43*, 6–29. [CrossRef]
9. Alhadad, S.S.J. Visualizing Data to Support Judgement, Inference, and Decision Making in Learning Analytics: Insights from Cognitive Psychology and Visualization Science. *J. Learn. Anal.* **2018**, *5*, 60–85. [CrossRef]

10. Wong, B.T.M.; Li, K.C. A review of learning analytics intervention in higher education (2011–2018). *J. Comput. Educ.* **2020**, *7*, 7–28. [[CrossRef](#)]
11. Bodily, R.; Ikaiahifo, T.K.; Mackley, B.; Graham, C.R. The design, development, and implementation of student-facing learning analytics dashboards. *J. Comput. High. Educ.* **2018**, *30*, 572–598. [[CrossRef](#)]
12. Jivet, I.; Scheffel, M.; Specht, M.; Drachler, H. License to evaluate: Preparing learning analytics dashboards for educational practice. In Proceedings of the 8th International Conference on Learning Analytics and Knowledge, Sydney, NSW, Australia, 7–9 March 2018; pp. 31–40. [[CrossRef](#)]
13. Buckingham Shum, S.; Ferguson, R.; Martinez-Maldonado, R. Human-Centred Learning Analytics. *J. Learn. Anal.* **2019**, *6*, 1–9. [[CrossRef](#)]
14. UNESCO. Section 2: Preparing learners to thrive in the future with AI. In *Artificial Intelligence in Education: Challenges and Opportunities for Sustainable Development*; Working Papers on Education Policy; The United Nations Educational, Scientific and Cultural Organization: Paris, France, 2019; pp. 17–24.
15. Schwendimann, B.A.; Rodríguez-Triana, M.J.; Vozniuk, A.; Prieto, L.P.; Boroujeni, M.S.; Holzer, A.; Gillet, D.; Dillenbourg, P. Perceiving Learning at a Glance: A Systematic Literature Review of Learning Dashboard Research. *IEEE Trans. Learn. Technol.* **2017**, *10*, 30–41. [[CrossRef](#)]
16. Sedraky, G.; Malmberg, J.; Verbert, K.; Järvelä, S.; Kirschner, P.A. Linking learning behavior analytics and learning science concepts: Designing a learning analytics dashboard for feedback to support learning regulation. *Comput. Hum. Behav.* **2020**, *107*, 105512. [[CrossRef](#)]
17. Rets, I.; Herodotou, C.; Bayer, V.; Hlosta, M.; Rienties, B. Exploring critical factors of the perceived usefulness of a learning analytics dashboard for distance university students. *Int. J. Educ. Technol. High. Educ.* **2021**, *18*, 46. [[CrossRef](#)] [[PubMed](#)]
18. Susnjak, T.; Ramaswami, G.S.; Mathrani, A. Learning analytics dashboard: A tool for providing actionable insights to learners. *Int. J. Educ. Technol. High. Educ.* **2022**, *19*, 12. [[CrossRef](#)]
19. Valle, N.; Antonenko, P.; Valle, D.; Sommer, M.; Huggins-Manley, A.C.; Dawson, K.; Kim, D.; Baiser, B. Predict or describe? How learning analytics dashboard design influences motivation and statistics anxiety in an online statistics course. *Educ. Technol. Res. Dev.* **2021**, *69*, 1405–1431. [[CrossRef](#)]
20. Iraj, H.; Fudge, A.; Khan, H.; Faulkner, M.; Pardo, A.; Kovanović, V. Narrowing the Feedback Gap: Examining Student Engagement with Personalized and Actionable Feedback Messages. *J. Learn. Anal.* **2021**, *8*, 101–116. [[CrossRef](#)]
21. Wagner, E.; Ice, P. Data changes everything: Delivering on the promise of learning analytics in higher education. *Educ. Rev.* **2012**, *47*, 32.
22. Algayres, M.G.; Triantafyllou, E. Learning Analytics in Flipped Classrooms: A Scoping Review. *Electron. J. E-Learn.* **2020**, *18*, 397–409. [[CrossRef](#)]
23. Wise, A.F.; Shaffer, D.W. Why Theory Matters More than Ever in the Age of Big Data. *J. Learn. Anal.* **2015**, *2*, 5–13. [[CrossRef](#)]
24. You, J.W. Identifying significant indicators using LMS data to predict course achievement in online learning. *Internet High. Educ.* **2016**, *29*, 23–30. [[CrossRef](#)]
25. Caspari-Sadeghi, S. Applying Learning Analytics in Online Environments: Measuring Learners’ Engagement Unobtrusively. *Front. Educ.* **2022**, *7*, 840947. [[CrossRef](#)]
26. Few, S. Dashboard Design: Taking a Metaphor Too Far. *DM Rev.* **2005**, *15*, 18.
27. McKenney, S.; Mor, Y. Supporting teachers in data-informed educational design. *Br. J. Educ. Technol.* **2015**, *46*, 265–279. [[CrossRef](#)]
28. van Leeuwen, A. Teachers’ perceptions of the usability of learning analytics reports in a flipped university course: When and how does information become actionable knowledge? *Educ. Technol. Res. Dev.* **2019**, *67*, 1043–1064. [[CrossRef](#)]
29. Davis, F.D.; Bagozzi, R.P.; Warshaw, P.R. User Acceptance of Computer Technology: A Comparison of Two Theoretical Models. *Manag. Sci.* **1989**, *35*, 982–1003. [[CrossRef](#)]
30. van Leeuwen, A.; Janssen, J.; Erkens, G.; Brekelmans, M. Supporting teachers in guiding collaborating students: Effects of learning analytics in CSCL. *Comput. Educ.* **2014**, *79*, 28–39. [[CrossRef](#)]
31. van Leeuwen, A.; Janssen, J.; Erkens, G.; Brekelmans, M. Teacher regulation of cognitive activities during student collaboration: Effects of learning analytics. *Comput. Educ.* **2015**, *90*, 80–94. [[CrossRef](#)]
32. Ramaswami, G.; Susnjak, T.; Mathrani, A.; Umer, R. Use of Predictive Analytics within Learning Analytics Dashboards: A Review of Case Studies. *Technol. Knowl. Learn.* **2022**, *28*, 959–980 [[CrossRef](#)]
33. Liu, R.; Koedinger, K.R. Closing the Loop: Automated Data-Driven Cognitive Model Discoveries Lead to Improved Instruction and Learning Gains. *J. Educ. Data Min.* **2017**, *9*, 25–41. [[CrossRef](#)]
34. Bañeres, D.; Rodríguez, M.E.; Guerrero-Roldán, A.E.; Karadeniz, A. An Early Warning System to Detect At-Risk Students in Online Higher Education. *Appl. Sci.* **2020**, *10*, 4427. [[CrossRef](#)]
35. Namoun, A.; Alshantiti, A. Predicting Student Performance Using Data Mining and Learning Analytics Techniques: A Systematic Literature Review. *Appl. Sci.* **2021**, *11*, 237. [[CrossRef](#)]
36. Jayaprakash, S.M.; Moody, E.W.; Lauría, E.J.M.; Regan, J.R.; Baron, J.D. Early Alert of Academically At-Risk Students: An Open Source Analytics Initiative. *J. Learn. Anal.* **2014**, *1*, 6–47. [[CrossRef](#)]
37. Gray, G.; Bergner, Y. A Practitioner’s Guide to Measurement in Learning Analytics: Decisions, Opportunities, and Challenges. In *The Handbook of Learning Analytics*, 2nd ed.; Lang, C., Siemens, G., Wise, A.F., Gasevic, D., Merceron, A., Eds.; SoLAR: Vancouver, BC, Canada, 2022; pp. 20–28.

38. Hattie, J.; Timperley, H. The power of feedback. *Rev. Educ. Res.* **2007**, *77*, 81–112. [[CrossRef](#)]
39. Carless, D.; Boud, D. The development of student feedback literacy: Enabling uptake of feedback. *Assess. Eval. High. Educ.* **2018**, *43*, 1315–1325. [[CrossRef](#)]
40. Sutton, P. Conceptualizing feedback literacy: Knowing, being, and acting. *Innov. Educ. Teach. Int.* **2012**, *49*, 31–40. [[CrossRef](#)]
41. Irons, A. *Enhancing Learning through Formative Assessment and Feedback*; Routledge: London, UK, 2007. [[CrossRef](#)]
42. Karaoglan Yilmaz, F.G.; Yilmaz, R. Learning analytics as a metacognitive tool to influence learner transactional distance and motivation in online learning environments. *Innov. Educ. Teach. Int.* **2021**, *58*, 575–585. [[CrossRef](#)]
43. Butler, D.L.; Winne, P.H. Feedback and Self-Regulated Learning: A Theoretical Synthesis. *Rev. Educ. Res.* **1995**, *65*, 245–281. [[CrossRef](#)]
44. Dawson, P.; Henderson, M.; Ryan, T.; Mahoney, P.; Boud, D.; Phillips, M.; Molloy, E. Technology and Feedback Design. In *Learning, Design, and Technology: An International Compendium of Theory, Research, Practice, and Policy*; Spector, M.J., Lockee, B.B., Childress, M.D., Eds.; Springer International Publishing: Cham, Switzerland, 2018; pp. 1–45. [[CrossRef](#)]
45. Pardo, A.; Jovanovic, J.; Dawson, S.; Gašević, D.; Mirriahi, N. Using learning analytics to scale the provision of personalised feedback. *Br. J. Educ. Technol.* **2019**, *50*, 128–138. [[CrossRef](#)]
46. Evans, C. Making Sense of Assessment Feedback in Higher Education. *Rev. Educ. Res.* **2013**, *83*, 70–120. [[CrossRef](#)]
47. Wilson, A.; Watson, C.; Thompson, T.L.; Drew, V.; Doyle, S. Learning analytics: Challenges and limitations. *Teach. High. Educ.* **2017**, *22*, 991–1007. [[CrossRef](#)]
48. Gašević, D.; Dawson, S.; Rogers, T.; Gasevic, D. Learning analytics should not promote one size fits all: The effects of instructional conditions in predicting academic success. *Internet High. Educ.* **2016**, *28*, 68–84. [[CrossRef](#)]
49. Joksimović, S.; Poquet, O.; Kovanović, V.; Dowell, N.; Mills, C.; Gašević, D.; Dawson, S.; Graesser, A.C.; Brooks, C. How Do We Model Learning at Scale? A Systematic Review of Research on MOOCs. *Rev. Educ. Res.* **2018**, *88*, 43–86. [[CrossRef](#)]
50. Bodily, R.; Verbert, K. Review of Research on Student-Facing Learning Analytics Dashboards and Educational Recommender Systems. *IEEE Trans. Learn. Technol.* **2017**, *10*, 405–418. [[CrossRef](#)]
51. Greer, J.; Mark, M. Evaluation methods for intelligent tutoring systems revisited. *Int. J. Artif. Intell. Educ.* **2016**, *26*, 387–392. [[CrossRef](#)]
52. Islahi, F.; Nasrin. Exploring Teacher Attitude towards Information Technology with a Gender Perspective. *Contemp. Educ. Technol.* **2019**, *10*, 37–54. [[CrossRef](#)]
53. Herodotou, C.; Hlosta, M.; Boroowa, A.; Rienties, B.; Zdrahal, Z.; Mangafa, C. Empowering online teachers through predictive learning analytics. *Br. J. Educ. Technol.* **2019**, *50*, 3064–3079. [[CrossRef](#)]
54. Herodotou, C.; Rienties, B.; Boroowa, A.; Zdrahal, Z.; Hlosta, M. A large-scale implementation of predictive learning analytics in higher education: The teachers’ role and perspective. *Educ. Technol. Res. Dev.* **2019**, *67*, 1273–1306. [[CrossRef](#)]
55. Sabraz Nawaz, S.; Thowfeek, M.H.; Rashida, M.F. School Teachers’ intention to use E-Learning systems in Sri Lanka: A modified TAM approach. *Inf. Knowl. Manag.* **2015**, *5*, 54–59.
56. Dimitriadis, Y.; Martínez-Maldonado, R.; Wiley, K. Human-Centered Design Principles for Actionable Learning Analytics. In *Research on E-Learning and ICT in Education: Technological, Pedagogical and Instructional Perspectives*; Tsiatsos, T., Demetriadis, S., Mikropoulos, A., Dagdilelis, V., Eds.; Springer International Publishing: Cham, Switzerland, 2021; pp. 277–296. [[CrossRef](#)]
57. Prestigiacomo, R.; Hadgraft, R.; Hunter, J.; Locker, L.; Knight, S.; Van Den Hoven, E.; Martinez-Maldonado, R. Learning-centred translucence: An approach to understand how teachers talk about classroom data. In Proceedings of the LAK ’20: 10th International Conference on Learning Analytics & Knowledge, Frankfurt, Germany, 23–27 March 2020; pp. 100–105. [[CrossRef](#)]
58. Herodotou, C.; Rienties, B.; Hlosta, M.; Boroowa, A.; Mangafa, C.; Zdrahal, Z. The scalable implementation of predictive learning analytics at a distance learning university: Insights from a longitudinal case study. *Internet High. Educ.* **2020**, *45*, 100725. [[CrossRef](#)]
59. Cardona, M.A.; Rodríguez, R.J.; Ishmael, K. *Artificial Intelligence and Future of Teaching and Learning: Insights and Recommendations*; Technical Report; US Department of Education, Office of Educational Technology: Washington, DC, USA, 2023.
60. Sarmiento, J.P.; Wise, A.F. Participatory and co-design of learning analytics: An initial review of the literature. In Proceedings of the LAK22: 12th International Learning Analytics and Knowledge Conference, Online, 21–25 March 2022; pp. 535–541.
61. Pardo, A.; Bartimote, K.; Shum, S.B.; Dawson, S.; Gao, J.; Gašević, D.; Leichtweis, S.; Liu, D.; Martínez-Maldonado, R.; Mirriahi, N.; et al. OnTask: Delivering Data-Informed, Personalized Learning Support Actions. *J. Learn. Anal.* **2018**, *5*, 235–249. [[CrossRef](#)]
62. Martínez-Maldonado, R.; Echeverria, V.; Fernandez Nieto, G.; Buckingham Shum, S. 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, Honolulu, HI, USA, 25–30 April 2020; pp. 1–15. [[CrossRef](#)]
63. Conijn, R.; Van Waes, L.; van Zaenen, M. Human-Centered Design of a Dashboard on Students’ Revisions During Writing. In *Addressing Global Challenges and Quality Education*; Lecture Notes in Computer Science; Alario-Hoyos, C., Rodríguez-Triana, M.J., Scheffel, M., Arnedillo-Sánchez, I., Dennerlein, S.M., Eds.; Springer: Cham, Switzerland, 2020; pp. 30–44. [[CrossRef](#)]
64. Echeverria, V.; Martínez-Maldonado, R.; Shum, S.B.; Chiluiza, K.; Granda, R.; Conati, C. Exploratory versus Explanatory Visual Learning Analytics: Driving Teachers’ Attention through Educational Data Storytelling. *J. Learn. Anal.* **2018**, *5*, 73–97. [[CrossRef](#)]

65. Fernandez Nieto, G.M.; Kitto, K.; Buckingham Shum, S.; Martinez-Maldonado, R. Beyond the Learning Analytics Dashboard: Alternative Ways to Communicate Student Data Insights Combining Visualisation, Narrative and Storytelling. In Proceedings of the LAK22: 12th International Learning Analytics and Knowledge Conference, New York, NY, USA, 21–25 March 2022; pp. 219–229. [\[CrossRef\]](#)
66. Ramos-Soto, A.; Vazquez-Barreiros, B.; Bugarín, A.; Gewerc, A.; Barro, S. Evaluation of a Data-To-Text System for Verbalizing a Learning Analytics Dashboard. *Int. J. Intell. Syst.* **2017**, *32*, 177–193. [\[CrossRef\]](#)
67. Sultanum, N.; Srinivasan, A. DataTales: Investigating the use of Large Language Models for Authoring Data-Driven Articles. *arXiv* **2023**, arXiv:2308.04076. [\[CrossRef\]](#)
68. He, Q.; von Davier, M. Identifying Feature Sequences from Process Data in Problem-Solving Items with N-Grams. In *Quantitative Psychology Research; Springer Proceedings in Mathematics & Statistics*; van der Ark, L.A., Bolt, D.M., Wang, W.C., Douglas, J.A., Chow, S.M., Eds.; Springer: Cham, Switzerland, 2015; pp. 173–190. [\[CrossRef\]](#)
69. Reis Costa, D.; Leoncio Netto, W. Process Data Analysis in ILSAs. In *International Handbook of Comparative Large-Scale Studies in Education: Perspectives, Methods and Findings*; Springer International Handbooks of Education; Nilsen, T., Stancel-Piatak, A., Gustafsson, J.E., Eds.; Springer International Publishing: Cham, Switzerland, 2022; pp. 1–27. [\[CrossRef\]](#)
70. Wise, S.L.; Ma, L. Setting response time thresholds for a CAT item pool: The normative threshold method. In Proceedings of the Annual Meeting of the National Council on Measurement in Education, Vancouver, BC, Canada, 13–17 April 2012; pp. 163–183.
71. Rios, J.A.; Guo, H. Can Culture Be a Salient Predictor of Test-Taking Engagement? An Analysis of Differential Noneffortful Responding on an International College-Level Assessment of Critical Thinking. *Appl. Meas. Educ.* **2020**, *33*, 263–279. [\[CrossRef\]](#)
72. Wise, S.L.; Kong, X. Response Time Effort: A New Measure of Examinee Motivation in Computer-Based Tests. *Appl. Meas. Educ.* **2005**, *18*, 163–183. [\[CrossRef\]](#)
73. Su, Q.; Chen, L. A method for discovering clusters of e-commerce interest patterns using click-stream data. *Electron. Commer. Res. Appl.* **2015**, *14*, 1–13. [\[CrossRef\]](#)
74. Ulitzsch, E.; He, Q.; Ulitzsch, V.; Molter, H.; Nichterlein, A.; Niedermeier, R.; Pohl, S. Combining Clickstream Analyses and Graph-Modeled Data Clustering for Identifying Common Response Processes. *Psychometrika* **2021**, *86*, 190–214. [\[CrossRef\]](#)
75. Tang, S.; Samuel, S.; Li, Z. Detecting atypical test-taking behaviors with behavior prediction using LSTM. *Psychol. Test Assess. Model.* **2023**, *65*, 76–124.
76. Gao, G.; Marwan, S.; Price, T.W. Early Performance Prediction using Interpretable Patterns in Programming Process Data. In Proceedings of the 52nd ACM Technical Symposium on Computer Science Education, Virtual Event, 13–20 March 2021; pp. 342–348. [\[CrossRef\]](#)
77. Rayner, K. Eye movements in reading and information processing: 20 years of research. *Psychol. Bull.* **1998**, *124*, 372. [\[CrossRef\]](#)
78. Morad, Y.; Lemberg, H.; Yofe, N.; Dagan, Y. Pupillography as an objective indicator of fatigue. *Curr. Eye Res.* **2000**, *21*, 535–542. [\[CrossRef\]](#)
79. Benedetto, S.; Pedrotti, M.; Minin, L.; Baccino, T.; Re, A.; Montanari, R. Driver workload and eye blink duration. *Transp. Res. Part Traffic Psychol. Behav.* **2011**, *14*, 199–208. [\[CrossRef\]](#)
80. Booth, R.W.; Weger, U.W. The function of regressions in reading: Backward eye movements allow rereading. *Mem. Cogn.* **2013**, *41*, 82–97. [\[CrossRef\]](#)
81. Inhoff, A.W.; Greenberg, S.N.; Solomon, M.; Wang, C.A. Word integration and regression programming during reading: A test of the E-Z reader 10 model. *J. Exp. Psychol. Hum. Percept. Perform.* **2009**, *35*, 1571–1584. [\[CrossRef\]](#)
82. Coëffé, C.; O’regan, J.K. Reducing the influence of non-target stimuli on saccade accuracy: Predictability and latency effects. *Vis. Res.* **1987**, *27*, 227–240. [\[CrossRef\]](#) [\[PubMed\]](#)
83. Adhikari, B. Thinking beyond chatbots’ threat to education: Visualizations to elucidate the writing and coding process. *arXiv* **2023**, arXiv:2304.14342. [\[CrossRef\]](#)
84. Allen, L.K.; Mills, C.; Jacovina, M.E.; Crossley, S.; D’Mello, S.; McNamara, D.S. Investigating boredom and engagement during writing using multiple sources of information: The essay, the writer, and keystrokes. In Proceedings of the Sixth International Conference on Learning Analytics & Knowledge, Edinburgh, UK, 25–29 April 2016; LAK ’16, pp. 114–123. [\[CrossRef\]](#)
85. Bixler, R.; D’Mello, S. Detecting boredom and engagement during writing with keystroke analysis, task appraisals, and stable traits. In Proceedings of the 2013 International Conference on Intelligent User Interfaces, Santa Monica, CA, USA, 19–22 March 2013; IUI ’13, pp. 225–234. [\[CrossRef\]](#)
86. Allen, L.; Creer, S.; Oncel, P. Chapter 5: Natural Language Processing as a Tool for Learning Analytics—Towards a Multi-Dimensional View of the Learning Process. In *The Handbook of Learning Analytics*; Society of Learning Analytics Research: Vancouver, BC, Canada, 2022. [\[CrossRef\]](#)
87. He, Q.; von Davier, M. Analyzing process data from problem-solving items with n-grams: Insights from a computer-based large-scale assessment. In *Handbook of Research on Technology Tools for Real-World Skill Development*; IGI Global: Hershey, PA, UAS 2016; pp. 750–777. [\[CrossRef\]](#)
88. Ulitzsch, E.; He, Q.; Pohl, S. Using sequence mining techniques for understanding incorrect behavioral patterns on interactive tasks. *J. Educ. Behav. Stat.* **2022**, *47*, 3–35. [\[CrossRef\]](#)
89. Guthrie, M.; Chen, Z. Adding duration-based quality labels to learning events for improved description of students’ online learning behavior. In Proceedings of the 12th International Conference on Educational Data Mining, Montreal, QC, Canada, 2–5 July 2019.

90. Chhabra, A. A System for Automatic Information Extraction from Log Files. Ph.D. Thesis, University of Ottawa, Ottawa, ON, Canada, 2022.
91. Roberts, L.D.; Howell, J.A.; Seaman, K. Give Me a Customizable Dashboard: Personalized Learning Analytics Dashboards in Higher Education. *Technol. Knowl. Learn.* **2017**, *22*, 317–333. [CrossRef]
92. Cavalcanti, A.P.; Barbosa, A.; Carvalho, R.; Freitas, F.; Tsai, Y.S.; Gašević, D.; Mello, R.F. Automatic feedback in online learning environments: A systematic literature review. *Comput. Educ. Artif. Intell.* **2021**, *2*, 100027. [CrossRef]
93. McGee, R.W. Is ESG a Bad Idea? The Chatgpt Response. *SSRN* **2023**. [CrossRef]
94. Kaddour, J.; Harris, J.; Mozes, M.; Bradley, H.; Raileanu, R.; McHardy, R. Challenges and Applications of Large Language Models. *arXiv* **2023**, arXiv:2307.10169. [CrossRef]
95. Frieder, S.; Pinchetti, L.; Griffiths, R.R.; Salvatori, T.; Lukasiewicz, T.; Petersen, P.C.; Chevalier, A.; Berner, J. Mathematical Capabilities of ChatGPT. *arXiv* **2023**, arXiv:2301.13867. [CrossRef]
96. Lim, L.A.; Dawson, S.; Gašević, D.; Joksimović, S.; Fudge, A.; Pardo, A.; Gentili, S. Students' sense-making of personalised feedback based on learning analytics. *Australas. J. Educ. Technol.* **2020**, *36*, 15–33. [CrossRef]
97. Dai, W.; Lin, J.; Jin, F.; Li, T.; Tsai, Y.S.; Gasevic, D.; Chen, G. Can Large Language Models Provide Feedback to Students? Case Study on ChatGPT. *EdArXiv* **2023**. [CrossRef]
98. Matelsky, J.K.; Parodi, F.; Liu, T.; Lange, R.D.; Kording, K.P. A large language model-assisted education tool to provide feedback on open-ended responses. *arXiv* **2023**, arXiv:2308.02439.
99. Yildirim-Erbasli, S.N.; Bulut, O. Conversation-based assessment: A novel approach to boosting test-taking effort in digital formative assessment. *Comput. Educ. Artif. Intell.* **2023**, *4*, 100135. [CrossRef]
100. Hasan, M.; Ozel, C.; Potter, S.; Hoque, E. SAPIEN: Affective Virtual Agents Powered by Large Language Models. *arXiv* **2023**, arXiv:2308.03022.
101. Bulut, O.; Yildirim-Erbasli, S.N. Automatic story and item generation for reading comprehension assessments with transformers. *Int. J. Assess. Tools Educ.* **2022**, *9*, 72–87. [CrossRef]
102. Attali, Y.; Runge, A.; LaFlair, G.T.; Yancey, K.; Goodwin, S.; Park, Y.; von Davier, A.A. The interactive reading task: Transformer-based automatic item generation. *Front. Artif. Intell.* **2022**, *5*, 903077. [CrossRef]
103. Sarsa, S.; Denny, P.; Hellas, A.; Leinonen, J. Automatic generation of programming exercises and code explanations using large language models. In Proceedings of the 2022 ACM Conference on International Computing Education Research, Virtual Event, 7–11 August 2022; Volume 1, pp. 27–43.
104. Tsai, D.; Chang, W.; Yang, S. Short Answer Questions Generation by Fine-Tuning BERT and GPT-2. In *Proceedings of the 29th International Conference on Computers in Education Conference, ICCE 2021—Proceedings*; Rodrigo, M., Iyer, S., Mitrovic, A., Cheng, H., Kohen-Vacs, D., Matuk, C., Palalas, A., Rajenran, R., Seta, K., Wang, J., Eds.; Asia-Pacific Society for Computers in Education: Taoyuan, Taiwan, 2021; pp. 509–515.
105. Shan, J.; Nishihara, Y.; Yamanishi, R.; Maeda, A. Question Generation for Reading Comprehension of Language Learning Test: -A Method using Seq2Seq Approach with Transformer Model-. In Proceedings of the 2019 International Conference on Technologies and Applications of Artificial Intelligence (TAAI), Kaohsiung, Taiwan, 21–23 November 2019; pp. 1–6. [CrossRef]
106. Offerijns, J.; Verberne, S.; Verhoef, T. Better Distractions: Transformer-based Distractor Generation and Multiple Choice Question Filtering. *arXiv* **2020**, arXiv:2010.09598. [CrossRef]
107. Zu, J.; Choi, I.; Hao, J. Automated distractor generation for fill-in-the-blank items using a prompt-based learning approach. *Psychol. Test. Assess. Model.* **2023**, *65*, 55–75.
108. Wainer, H.; Dorans, N.J.; Flaugher, R.; Green, B.F.; Mislevy, R.J. *Computerized Adaptive Testing: A Primer*; Routledge: Oxfordshire, UK, 2000.
109. Woolf, B.P. *Building Intelligent Interactive Tutors: Student-Centered Strategies for Revolutionizing E-Learning*; Morgan Kaufmann: Cambridge, MA, USA, 2010.
110. Friedman, L.; Ahuja, S.; Allen, D.; Tan, T.; Sidahmed, H.; Long, C.; Xie, J.; Schubiner, G.; Patel, A.; Lara, H.; et al. Leveraging Large Language Models in Conversational Recommender Systems. *arXiv* **2023**, arXiv:2305.07961.
111. Patel, N.; Nagpal, P.; Shah, T.; Sharma, A.; Malvi, S.; Lomas, D. Improving mathematics assessment readability: Do large language models help? *J. Comput. Assist. Learn.* **2023**, *39*, 804–822. [CrossRef]
112. Zhang, Y.; Ding, H.; Shui, Z.; Ma, Y.; Zou, J.; Deoras, A.; Wang, H. Language models as recommender systems: Evaluations and limitations. In Proceedings of the NeurIPS 2021 Workshop on I (Still) Can't Believe It's Not Better, Virtual, 13 December 2021.
113. Lim, L.A.; Dawson, S.; Gašević, D.; Joksimović, S.; Pardo, A.; Fudge, A.; Gentili, S. Students' perceptions of, and emotional responses to, personalised learning analytics-based feedback: An exploratory study of four courses. *Assess. Eval. High. Educ.* **2021**, *46*, 339–359. [CrossRef]
114. Bonner, E.; Lege, R.; Frazier, E. Large Language Model-Based Artificial Intelligence in the Language Classroom: Practical Ideas For Teaching. *J. Teach. Engl. Technol.* **2023**, *2023*. [CrossRef]
115. DiCerbo, K. Building AI Applications Based on Learning Research [Webinar]. 2023. Available online: <https://www.youtube.com/watch?v=ugyfdjI9NEk> (accessed on 14 July 2023).
116. Kasneci, E.; Seßler, K.; Küchemann, S.; Bannert, M.; Dementieva, D.; Fischer, F.; Gasser, U.; Groh, G.; Günemann, S.; Hüllermeier, E.; et al. ChatGPT for good? On opportunities and challenges of large language models for education. *Learn. Individ. Differ.* **2023**, *103*, 102274. [CrossRef]

117. Tlili, A.; Shehata, B.; Adarkwah, M.A.; Bozkurt, A.; Hickey, D.T.; Huang, R.; Agyemang, B. What if the devil is my guardian angel: ChatGPT as a case study of using chatbots in education. *Smart Learn. Environ.* **2023**, *10*, 15. [[CrossRef](#)]
118. Mathrani, A.; Susnjak, T.; Ramaswami, G.; Barczak, A. Perspectives on the challenges of generalizability, transparency and ethics in predictive learning analytics. *Comput. Educ. Open* **2021**, *2*, 100060. [[CrossRef](#)]
119. Barros, T.M.; Souza Neto, P.A.; Silva, I.; Guedes, L.A. Predictive models for imbalanced data: A school dropout perspective. *Educ. Sci.* **2019**, *9*, 275. [[CrossRef](#)]
120. Yan, L.; Sha, L.; Zhao, L.; Li, Y.; Martinez-Maldonado, R.; Chen, G.; Li, X.; Jin, Y.; Gašević, D. Practical and Ethical Challenges of Large Language Models in Education: A Systematic Literature Review. *arXiv* **2023**, arXiv:2303.13379.
121. Truong, T.L.; Le, H.L.; Le-Dang, T.P. Sentiment analysis implementing BERT-based pre-trained language model for Vietnamese. In Proceedings of the 2020 7th NAFOSTED Conference on Information and Computer Science (NICS), Ho Chi Minh City, Vietnam, 26–27 November 2020; pp. 362–367. [[CrossRef](#)]
122. Khosravi, H.; Shum, S.B.; Chen, G.; Conati, C.; Tsai, Y.S.; Kay, J.; Knight, S.; Martinez-Maldonado, R.; Sadiq, S.; Gašević, D. Explainable artificial intelligence in education. *Comput. Educ. Artif. Intell.* **2022**, *3*, 100074. [[CrossRef](#)]
123. Huang, L. Ethics of Artificial Intelligence in Education: Student Privacy and Data Protection. *Sci. Insights Educ. Front.* **2023**, *16*, 2577–2587. [[CrossRef](#)]
124. Remian, D. Augmenting Education: Ethical Considerations for Incorporating Artificial Intelligence in Education. In *Instructional Design Capstones Collection*; University of Massachusetts Boston: Boston, MA, USA, 2019.
125. Akgun, S.; Greenhow, C. Artificial intelligence in education: Addressing ethical challenges in K-12 settings. *AI Ethics* **2022**, *2*, 431–440. [[CrossRef](#)] [[PubMed](#)]
126. Knight, S.; Shibani, A.; Abel, S.; Gibson, A.; Ryan, P.; Sutton, N.; Wight, R.; Lucas, C.; Sandor, A.; Kitto, K.; et al. AcaWriter: A learning analytics tool for formative feedback on academic writing. *J. Writ. Res.* **2020**, *12*, 141–186. [[CrossRef](#)]
127. Kochmar, E.; Vu, D.D.; Belfer, R.; Gupta, V.; Serban, I.V.; Pineau, J. Automated personalized feedback improves learning gains in an intelligent tutoring system. In Proceedings of the Artificial Intelligence in Education: 21st International Conference, AIED 2020, Proceedings, Part II, Ifrane, Morocco, 6–10 July 2020; pp. 140–146.

**Disclaimer/Publisher’s Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.