Science Advanced
ISSN: 3028-5216
Science Advanced
Journal homepage: www.scienceadvanced.org

Big Data in Education: Real-Time Personalization of Learning

Received: 01 11 2024, Accepted: 12 11 2024, Available online: 16 12 2024

Juan David Méndez Niño
Msc. Ciencias Geomáticas. Esp. Gerencia, Gobierno y asuntos públicos. Ing. Topográfico.
Universidad Distrital Francisco judameni@gmail.com
https://orcid.org/0009-0008-4601-6709

Gustavo Adolfo Gonzalez Roys
Ingeniero Agroindustrial - Investigación y desarrollo - Gerencia y Calidad Educativa
Dr. Educación UBC https://orcid.org/0000-0001-5274-8182

Bellanith Aguilar Vazquez
Licenciada en Matemáticas de la Universidad del Tolima. Especialista tecnológica en gestión de proyectos del SENA. Magister en educación de la Corporación Universitaria Minuto de Dios. Maestra en educación con acentuación en procesos de enseñanza y aprendizaje del Tecnológico de Monterrey Doctora en educación de la Universidad Baja California.
tutorabeagva@gmail.com
https://orcid.org/0009-0003-6563-5840

Resumen

El Big data, como una rama de la ciencia de datos, se ha visto como una fuerza transformadora en el ámbito educativo, a pesar de los beneficios potenciales, como mejorar la experiencia de aprendizaje de los estudiantes y permitir decisiones educativas más informadas, también existen preocupaciones éticas y de privacidad relacionadas con la recolección y uso de datos estudiantiles. La personalización del aprendizaje implica adaptar el currículo y las estrategias de enseñanza a las necesidades individuales de los estudiantes, un proceso tradicionalmente impulsado por el juicio profesional de los educadores, sin embargo, el uso de análisis de datos y sistemas basados en algoritmos tiene el potencial de revolucionar cómo se entiende y se aplica la educación personalizada. A medida que los entornos educativos se convierten en ecosistemas ricos en datos, es crucial que las instituciones desarrollen un enfoque basado en datos para mejorar los resultados de aprendizaje. Se enfatiza la importancia de analizar los datos en tiempo real para adaptar la enseñanza y aumentar la equidad educativa.
Palabras clave: Big data, educación, personalización, aprendizaje en tiempo real, ética.

Abstract

Big data, as a branch of data science, has been seen as a transformative force in education, despite the potential benefits, such as improving the learning experience for students and enabling more informed educational decisions, there are also ethical, and privacy concerns related to the collection and use of student data. Personalization of learning involves adapting curriculum and teaching strategies to the individual needs of students, a process traditionally driven by the professional judgment of educators, however, the use of data analytics and algorithm-based systems has the potential to revolutionize how personalized education is understood and applied. As educational environments become data-rich ecosystems, it is crucial for institutions to develop a data-driven approach to improving learning outcomes. The importance of analyzing data in real time to adapt teaching and increase educational equity is emphasized.
Keywords: Big data, education, personalization, real-time learning, ethics.

Introducción

As technology and digital media have become ubiquitous in society, there have been increasing concerns regarding their influence on education, necessitating an examination of their effects on student teaching and learning. Learning analytics has emerged as a relatively new research area, investigating how data can be collected, analyzed, and used to better understand and optimize student learning in educational environments. Over the years, growing attention has been given to the collection and analysis of educational data to improve educational practices. A variety of terms such as “educational data mining,” “learning analytics,” and “big data in education” have been used to describe similar themes (Luan et al., 2020).

In this context, big data, as one branch of data science, has been viewed as a transformative force in education by some scholars. Institutions and researchers have been investigating how data analytics could be leveraged to develop personalized learning experiences for students.

Meanwhile, there is also skepticism over the implications of big data for education, particularly with regard to privacy and ethical considerations in the collection and use of student data. With the rapid rise of big data, digital education tools and resources have proliferated in schools, and there is growing pressure on educators to adopt these technologies. Therefore, it is crucial to understand how big data could transform educational practices and learning outcomes.

Educational personalization refers to the customization of the curriculum and instructional strategies to meet the needs of individual students, a process that traditionally relied on the subjective professional judgment of teachers. However, the involvement of big data in education has the potential to reshape such personalization through data analytics and algorithm-driven decisions. On one hand, big data could empower educators by uncovering hidden insights through analytics and fostering informed decision-making. On the other hand, the reliance on data analytics could render educators powerless, as mechanized systems take control over data interpretation and decision-making processes, similar to the debate surrounding Artificial Intelligence (AI), which could either augment or threaten human intelligence.

Background and Significance

Contemporary education systems are data-rich environments. Student characteristics, behaviors, and academic outcomes are increasingly captured through digital channels. An evidence-based approach to education has the potential to enhance learning and institutional effectiveness (Luan et al., 2020). However, across education systems, data use remains predominantly traditional. A greater focus on sophisticated analytics utilizing big data is needed to realize data-driven improvements in learning outcomes. Personalization of learning and the learning environment is viewed as the most critical application of big data in education. A refined understanding of big data in education, and its potential applications, is necessary to assist institutional leaders in capitalizing on big data to enhance student engagement, retention, and academic success.

Educational contexts have utilized data for decision making for over a century. Beginning with traditional forms of assessment, student data were used to evaluate student performance, summarize academic achievement, and identify learning needs. Reports and summary statistics informed decisions on curriculum design and the allocation of pedagogical resources. During the 1990s, the convergence of web-based learning environments and online assessment systems introduced new forms of data regarding student interactions with learning activities and online content. This data fed into reporting systems that produced a wider range of descriptive statistics on student performance. At this point, almost all education data was numerical and summarized in static reports.

The emergence of a more complex education data ecosystem, incorporating diverse data types, channels, and sources, characterizes the contemporary education landscape. Big data frameworks that analyze complex data can be employed to greatly augment current reporting and “flagging” systems. With algorithms that detect patterns and proactively make recommendations, education data analytics can drive “real-time” personalization of learning and the learning environment. Education systems need to rapidly transition towards a data-driven approach in response to changing student demographics, learning spaces, and learning needs. The data currently collected across education systems could have a profound impact if used to develop a deep understanding of how students learn. By addressing institutional leaders, this review advances the discussions of big data in education from an abstract theoretical level to a consideration of concrete and specific applications that can be implemented using the data currently collected.

This purpose and scope section articulates the study focused on big data in education. In particular, it examines how big data can inform real-time personalization of learning. Grounded in a review of recent literature, it outlines the research questions guiding this inquiry into an agenda for further research and practice. 1) What aspects of big data are to be considered? 2) What is real-time personalization of learning? 3) What is the significance of focusing on the analysis of data as it is generated? 4) What is the scope of this study? 5) Who is the intended audience? 6) What are the expected outcomes? By specifying what big data and real-time personalization of learning mean in the context of this study, these questions frame a clear analysis of trends in education research over the past five years. Given the rapid increase in the volume and variety of data collected in educational contexts, the significance of focusing on real-time data analysis is established first. 1) Data generated about learning as it happens has the most immediate capacity to inform learning, teaching, and assessment; 2) The learning, teaching, and assessment processes are always designed and enacted with some measurable intention, and therefore always have associated data trails. The emphasis here is on data directly and automatically generated by systems used to support learning, teaching, and planning rather than on human-generated data such as logs of observed actions or textual accounts of educational experiences (Luan et al., 2020). The discussion of the significance of focusing on real-time data analysis also clearly establishes the scope of the study. Only data captured by educational technologies will be considered in-discipline and relevant to the educational context explored. Within that scope, the analysis examines what is meant by real-time personalization of learning in education research published between 20152019. The audience is primarily educators, educational

administrators, and education policymakers. These stakeholders are best placed to innovate their teaching methodologies in light of insights driven by big data (Nitkin, 2018). In discussing the significance of this study, the expected outcomes of a clear framework for analysis are set and hope is expressed that findings will motivate stakeholders to better engage with research.

Foundations of Big Data in Education

The purpose of this section is to establish a base-level understanding of big data in educational settings, the definition, characteristics, and capabilities of this technology as it relates to learning environments and other pertinent discussions. Starting with the technical side of things, big data is precisely defined, including the four defining characteristics. Each attribute is examined closely, illustrating the complexities and capabilities of big data systems for educational environments. Possible applications of big data in education are also discussed, including how it can help with teaching, learning, and other administrative tasks. The need for data-driven strategies to support individualized learning is made clear, especially in terms of their potential impact on student outcomes (Wang, 2017). Addressing these issues provides readers with the necessary technical vocabulary and settings to follow along with the discussions on real-time personalization in educational contexts. Ultimately, this section serves as an introduction to big data in education and its role in real-time learning personalization. Foundational ideas and understanding will help future adjustments and educational considerations. The group settings will be addressed as a whole, but the nuance of each edge's role and impact will be examined as much as possible (Luan et al., 2020). As a relatively new field, big data in education isn't thoroughly researched or understood, particularly in policy and public education. Still, the object considerations are confirmed and defined as best as possible. In-depth education personalization through big data or even just data in general is an important future focus, particularly for public educational settings.

The term "big data" continues to gain traction and is often at the forefront of discussions across numerous disciplines and professions. Although it is a much-debated issue, big data presently lacks a universally accepted definition. However, a thorough comprehension of big data is possible by understanding its fundamental characteristics. A variety of definitions for big data exist, with some tendency to group them into four essential traits or attributes. Big data is most often characterized by its four defining characteristics; volume, variety, velocity, and veracity. Beyond these four defining characteristics, big data may also be described by its four descriptive characteristics; relational, scalable, heterogeneous, and complex (Muhammad Nda & Bin Tasmin, 2019).

The ability to collect, store, and analyze massive quantities of data has grown dramatically in recent years and will continue to do so. The data volumes created are so huge that they exceed the processing capacity of current database systems. Organizations must handle large volumes of data regularly and around the clock, whether through online transactions, customer interactions, applications, sensors, or other means. Additionally, many organizations have large quantities of data already stored but unprocessed, which can yield significant insights if analyzed. Various sources generate a wide range of data currently and will continue to do so in the future, including public records, social networks, online transactions, audio/video files, scientific data, machine logs, data from sensors, and mobile devices. Educational institutions, in particular, generate vast volumes of data daily, from student records to examination results, attendance, and curriculum transaction details, and beyond.

The data produced within an organization often originates from a wide assortment of sources and comes in many different formats. In addition to the standard structured data input through relational database systems, organizations typically have numerous unstructured data sources, such as machine logs, emails, images, audio/video files, documents, online transactions, social networking data, and web pages. Furthermore, organizations must also consider semi-structured data, like XML files or web pages with embedded tags. The processing speed required to keep up with the data production is often a challenge. Data may need to be processed in real-time as it arrives. For example, an organization may wish to monitor online transactions as they occur and immediately detect and respond to fraudulent behavior. Educational institutions offer numerous online services, from admissions and registration to class scheduling, attendance, online assessment, and many student services. A substantially large volume of data must be processed in real-time as these transactions take place, in addition to the need for batch processing nightly of other daily transactions.

Data is not necessarily reliable, and an organization must consider how much trust to place in the information. Data may be corrupt, misleading, or overly imprecise. For example, companies collecting public information from the web must consider whether a particular web page is trustworthy. Banks and credit agencies routinely use financial data to assess whether to grant a loan, and an applicant's financial data can significantly affect the outcome. In education, results must be verified to ensure their validity and reliability before making critical decisions affecting students, such as outcomes on assessments, disciplinary actions, and recommendations for scholarships, academic standing, and program dismissal.

Institutions have long collected and maintained data, yet commensurately extracting useful information from this data has proven challenging. As daily transactions at data-

producing centers occur in the millions or even billions currently, and with the data explosion taking place on the web, focused efforts must be made to ensure that good data management is a priority. The four big data capabilities, namely storage, search, transfer, analysis, must be in place, along with considering the relative costs of each. It is essential to further examine how educational institutions will keep control of the data they generate, and ensuring good data management is a priority for these institutions.

The growing prominence of big data across industries, including education, is unmissable. Opportunities abound for educators and education systems to maximize efficiency and effectiveness. Understanding and utilizing the range of educational data generated daily can inform and transform curriculum practices, classroom cultures, and learning environments. Institutions are beginning to leverage big data to improve educational achievement and equity by utilizing data to shape future pedagogies, learning frameworks, and curriculum design. A range of applied practices is emerging, showcasing the power of data analytics to reshape and redefine teaching. Personal learning experiences come from the data and insights available, having a direct impact on student achievement. Emerging applied practices include learning analytics, adaptive learning technologies, and data-informed decisions regarding pedagogy, content, and learning environments. Educators must integrate real-time data within curriculum design and classroom practices to fully capitalize on the potential of big data. Optimizing learning approaches and educational outcomes relies on data being an integral part of the development cycle of curriculum practices (Luan et al., 2020). Examples of institutions employing used data to enhance service levels, learning experiences, and outcomes are included. Data processing, learning analytics, and pedagogical design need to align. Data and educational technology work together. Data can mechanize and optimize the cycle of design, enactment, reflection, and refinement, typical of learning or pedagogical design, thereby augmenting decisions made by educators. Linking theory to practice shows how big data can resolve key educational issues. Education systems must consider the reasons for and rationale behind big data's use, while policymakers need clarity on big data's role in pedagogy, learning, and education. Despite concerns, there are compelling reasons for continued educational policy, practice, and research regarding big data (Ashrafimoghari, 2022).

Numerous applications of big data have emerged within the educational domain. Attempts at exploring the good, utilitarian, and possible within the educational landscape converge using a big data framework. Education is awash with data daily, yet the pedagogical impact of this data is often overlooked. As a result, many schools and educators become passive recipients of data, limiting the potential benefits it could offer. Pedagogical applications of data can be classified as macro, meso, or micro. Data and technology working synergistically can be seen as macro and applied at the whole-network level. Educational technology working together with data impacts the large-scale implementation of good pedagogical practices and classroom approaches. Aggregate data and wider systemic considerations influence how technology and data operate. The systemic view focuses on the whole administrative network, data as infrastructure, technology designing educational freedom, and educational technology. Meanwhile, large technology corporations involve themselves in the education sector's macro-data applications that reshape its meaning and function. Complex economic relations arise due to transparency, accountability, and pedagogical concerns over outsourcing responsibility to the effects of data and technology. There is a growing intersection between big data and educational technology concerns and divergent ideological positions on the impact of technology and data on pedagogy, epistemology, language, and conception of learning.

Real-Time Personalization in Education

The emergence of Big Data has prompted an exploration of its potential applications within educational contexts, particularly in relation to learning personalization. Given the myriad preferences, paces, and needs of learners, the focus of research endeavors has centered on how data can be harnessed to create personalized experiences for students. Within this discourse, the concept of "real-time personalization" is introduced and its growing significance in education outlined. As terms such as learning personalization and adaptive learning have gained traction, the emphasis is on how learning experiences can be adjusted based on individual student needs, as revealed in the data (Fu et al., 2022). Personalization can take place in real-time, with a learning environment capable of timely data collection and processing ensuring that appropriate content and support are provided as students engage with material. In education, this concept relates to the focus on immediate data collection and processing, enabling effective personalization through the identification of students' needs as they learn and the provision of immediate adjustments. The power of personalized experiences is heightened when feedback occurs immediately, thus promoting better learning results.

Real-time personalization is also viewed as a means to overcome the fixed pace of education. With a concern for inclusivity, this approach is seen as desirable for addressing diverse preferred paces and learning styles. Efforts to personalize education usually require a significant amount of data to understand each student's individual characteristics; however, this personalization can only be realized progressively after the data has been collected. In contrast, real-time educational personalization seeks to enrich learning experiences on the grounds of currently captured data, providing suitable

content and directing support right at the moment students interact with the learning environment. A review of research in this area reveals current practices and highlights the technologies that make these personalized experiences possible across learning environments. Most developments center on the use of various analytics tools to uncover the meaning of data for each student concerning the learning context. However, it is argued that uncovering the meaning of data alone is insufficient. As data reveals new insights about each student, it becomes necessary for educators to transform their practices in response to those insights. This calls for the implementation of new ways of teaching and learning in light of what the data indicates about student performance. While there are examples of developing new practices based on data, such efforts are found to be limited and seldom address personalization at the level of content or support. Although real-time educational personalization is generally viewed as a means to offer optimal learning experiences, there is an acknowledgment that such systems might not always be feasible, particularly amid resource limitations. Nevertheless, this personalization is deemed the most desirable state of education to be pursued as technology develops.

Real-time personalization refers to the personalization of instructional strategies based on continuous data streams generated by student interactions with learning technologies. This approach is crucial for the achievement of learning goals (Fu et al., 2022), as it is increasingly common to find educational technologies that provide users with access to a variety of learning resources and activities, which makes the use of these technologies a promising context for the setup of personalized learning interventions. Learning should not be seen as a one-size-fits-all approach, but rather as a process that takes into consideration the individual needs of students. In this sense, different students should be offered alternative learning opportunities, based on their prior knowledge, interests, efforts, and learning goals. This perspective highlights the importance of personalization in learning environments, as a means to increase the engagement and motivation of students, two key factors associated with successful learning experiences.

Personalized learning environments are believed to be a more effective means to achieve the desired learning outcomes. In these environments, students are offered different learning opportunities based on the ongoing assessment of their degree of comprehension. For instance, if a student does not grasp a particular concept, the student should be immediately presented with an alternative explanation, rather than proceeding to a new topic. Pedagogical strategies that employ this mechanism are generally deemed more effective than those that do not. The continuous stream of data generated by student interactions with learning technologies can be used to conduct real-time assessments of the students' degree of comprehension, which, in turn, can inform the adaptation of the instructional strategies being used. Consequently, there is a wide interest in the development of educational technologies capable of this form of personalization. The use of continuous streams of data to inform instructional strategies can promote the effectiveness of learning environments. Several educational technologies have been developed that collect data on student interactions with the system and use it to adapt the learning path followed by the student. Most of these systems employ pre-defined rules that determine how the pedagogy should adapt based on the data collected. In this sense, there is a growing interest in the application of data mining techniques that automate the discovery of the most effective pedagogical interventions, given a specific learning context, based on an analysis of the data collected. When pedagogical strategies are designed from scratch, it is essential to consider how the continuous data collected will be used to assess the effectiveness of the strategies. Thus, a more agile approach to pedagogy design can be effective, where the strategies are initially simple, and the collected data are used to progressively enhance their effectiveness. Educational technologies that collect data on student performance analytics can empower educators to make informed decisions regarding the design of learning activities. In this sense, the design of learning activities should be considered a coevolution process between tech and pedagogy, where data collected from the student interactions with the technology are used to inform adjustments in the learning activities.

Shifting focus to real-time personalization in education, challenges to implementation are noted first: concerns over data privacy; the uneven availability of resources across education systems; and a lack of training for educators in big data and AI technologies. Although these hurdles are daunting, proactive approaches can harness opportunities – more effective and engaging learning environments (Luan et al., 2020). With success, personalization could mean students are more engaged with their educational pathways and are co-designers of learning environments, rather than passive learners. Furthermore, their inputs could engender adaptive pathways through education, underpinned by the co-evolution of learning narratives and analytics. Addressing challenges using a collaborative model would pool expertise across the education ecosystem – educators, policymakers, and technology providers would work together to define problems, outline requirements, and develop solutions using technology. The emphasis would be on how to overcome problems rather than delineating them. There is a current focus on the need for regulations over student/data privacy which, while important, may distract from seizing opportunities. The potential for personalization to reshape education radically and beneficially should be prioritized.

Technologies and Tools for Real-Time Personalization

Focusing on the context of education, this section endeavors an exploratory discussion on some specific technologies and tools that facilitate real-time personalization. It addresses the current cutting-edge advancements on technologies, tool and approaches in the machine learning, artificial intelligence, swarm intelligence, web semantic through simple and easy to follow examples to demonstrate how student data can be analyzed by these technologies. By applying them, educators can get fast and actionable insights of students in order to shape the personal or individualized learning experiences based on student's needs addressed. Along with that, the importance of data mining and predictive analytics to anticipate the performance and trends of students is discussed. The advanced technologies and tools to mine and analyze educational data are proposed in order to not only explore hidden patterns and trends on data but also create predictive models for future outcomes. Thus, by harnessing these advanced technologies and tools, the educational institutes will have the ability to gather actionable insights from data in order to improve the decision-making and instructional methodologies taken place in classrooms (Ashrafimoghari, 2022). Furthermore, it is discussed that by adopting such technologies and tools, the possibilities on learning environment to continuously adapt will be greatly enhanced. Detailed discussions on several successfully implemented or on-going technology integrated projects are provided in real classroom contexts along with the implications of these technologies and tools are presented (Luan et al., 2020). Ultimately, the discussion summed that by having access to innovative and cutting-edge technologies, tools and approaches, the education industry can reap the benefits of education data-mining in order to foster effective real-time personalization on education.

Machine-learning technology employs algorithms that expedite learning by interpreting data in multiple formats. This technology empowers computers to perform tasks typically necessitating human intelligence (R. Kshirsagar et al., 2022). Artificial machine intelligence employs hardware and software that mimic cerebral functions for decision-making. This intelligence achieves superior task performance, operating quicker than the human brain while enhancing effectiveness through automated task completion.

With student input and idealized feedback data, models transform raw states to goal states. In education, student input classifies levels attained through input data and expected goals (E. August & Tsaima, 2021). Students learn at varying rates and times, creating distinct learning paths. Models consider instructional style, student proficiency, time, and learning path, yielding diverse approaches to achieve educational goals. Given models, learning plans target a specific goal, deploying various actions based on time, state, and data to manipulate the student's environment.

Data mining and predictive analytics are essential to facilitate real-time personalization of students´ learning. Data mining is defined as the process of extracting meaningful information from large amounts of data. Data mining collects big data in education and uses it to find trends, important patterns, and common needs among students. This generally involves the use of software and tools that aggregate data from several sources, such as online quizzes, discussion forums, and homework submissions. Once collected, data mining processes this information to create visualization dashboards that help educators monitor their students (Ashrafimoghari, 2022). Predictive analytics uses statistical algorithms to anticipate certain outcomes based on historical patterns found in data. In education, predictive analytics takes mined data to create statistical models that allow educators to estimate the performance of certain students based on past behavior. The combination of both techniques is powerful, as schools can use data mining to identify problems in student learning and apply predictive models to anticipate those same problems before they worsen. Educational institutions can implement various tools and software that help with data mining and predictive analytics. For instance, the programming language R offers several libraries for conducting data mining and predictive analysis. Orange is another excellent open-source solution where educators can drag and drop visual programming widgets to conduct data mining and predictive analysis. Finally, RapidMiner is a subscription-based platform predominantly used in education and research that supports everything from harvesting data to modeling predictions. The extensive analysis that these programs can conduct highlights the importance of data-driven decision-making to create tailored learning experiences. Data mining and predictive analytics can efficiently facilitate the personalization of education and the necessary tools to accomplish it in a real educational setting. Starting from the simple implementation of programs that aggregate data on student performance, educators can leverage data mining and predictive analysis to ensure that every student meets their full potential.

Implementing real-time personalization in educational settings

The rapid development of the big data ecosystem presents an exciting opportunity to craft a new educational paradigm, where each student has a direct connection with an individualized learning pathway to success. Incorporating learning analytics systems in education can harness the potential of big data, providing educators with immediate access to quality data. Such systems can enable

them to quickly adapt their pedagogy to real-time changes in learner behavior, preference, and knowledge. The goal becomes one of learning equity, ensuring all learners have the same opportunity to reach their desired education outcomes through personalization (Nitkin, 2018). Educational institutions must therefore consider how to transition from a historic focus on data collection and retrospective analysis to providing personalized learning through pedagogical intervention based on real-time learning analytics. While researching learning analytics methods and their application to education pedagogy is essential, there are practical considerations to address for successful implementation in educational settings. A structured approach must blend technological capability with pedagogy and curriculum design, ensuring education practitioners are trained in effectively employing real-time data in intervention design—a skill more pivotal than data analytics expertise for successful implementation. By outlining case studies of educational institutions with learning analytics systems successfully leveraging effective strategies, potential stumbling blocks in transitioning to personalized learning environments are discussed. Implementers should also be aware of possible challenges to success, such as institutional inertia or the lack of adequate infrastructure backed by learning analytics systems. Best practices are articulated to overcome these challenges and facilitate a smooth adoption process, ensuring stakeholders have a roadmap for successful implementation based on the most current understanding of pedagogical intervention employing real-time learning analytics.

Case Studies

This subsection presents various case studies that illustrate successful implementations of real-time personalization in educational settings. Institutions considered in this review represent different levels of education (K-12, university) and diverse geographic and cultural contexts. After a brief description of each case study, the approaches taken by the institutions are summarized. Each institution adopted a very different strategy, highlighting the diversity in context. The challenges encountered in these cases are analyzed, as well as the lessons learned during the implementation process. The case studies provide real-world evidence that demonstrates how data-driven techniques have positively impacted student learning and engagement. Each case study also highlights the role of technology and the involvement of a few educators as the main drivers that pushed these initiatives forward (Nitkin, 2018). By comparing different scenarios, insights are drawn that are likely to be applicable to other educational contexts. An emphasis is placed on how strategies adopted by a specific institution can be scaled to other contexts, encouraging broader adoption of successful strategies across different types of institutions. In summary, this subsection aims to validate the potential of real-time personalization through evidence-based examples. Five diverse case studies that have implemented data-driven personalized learning approaches are presented in detail.

The following section outlines a series of best practices and key strategies that can support the successful implementation of real-time personalization in education. To begin, it is crucial that educators, technology experts, and school administrators collaborate to design effective frameworks for the personalization of learning (Fu et al., 2022). Working closely together helps ensure that each effort is curriculum-driven, pedagogically sound, and aligned with the overarching educational vision of the institution. Initial investments in real-time personalization are unlikely to lead to significant outcomes without ongoing training and professional development for educators (Ballard & Butler, 2011). Therefore, it is vital to develop training programs that help educators understand and be equipped to leverage real-time data effectively. Professional development should continue after initial training to accommodate the diverse needs of educators. Integrating technology into the learning process is not a one-time event; it requires an environment and culture that foster adaptability. Technology integration success often stems from exemplary educators who explore new technologies and develop ways to integrate them into their teaching, inspiring colleagues to adopt new practices. Therefore, it is helpful to build a technology integration plan that provides adequate support and flexibility for educators. In such plans, time for experimentation is essential. A successful element of the iEducator project has been the focus on continuous assessment and establishing feedback loops to monitor student progress. Personalized learning approaches should be refined based on what the collected data reveal about students' needs. However, given the vast number of available applications and tools, it is essential to consider some questions when selecting tools for data collection. Additionally, it is crucial to measure students' progress adequately; otherwise, it is difficult to evaluate whether a particular strategy is effective. For instance, standardized tests may not be appropriate for measuring the efficacy of personalized learning approaches based on real-time data. Finally, it is essential to recognize that ensuring real-time personalization is not merely a technological endeavor; it is about educational approaches and strategies that ensure the technology enriches learning. Therefore, the strategies outlined are primarily pedagogical, and stakeholders are encouraged to consider beyond the technological aspect involved.

The increasing utilization of big data in educational settings raises important ethical and privacy considerations. The collection and analysis of data related to students, teachers, and educational institutions have the potential to improve learning outcomes and efficiency. However, concerns about safeguarding sensitive

information and promoting data security are paramount. To address these concerns, a thorough understanding of the potential risks associated with big data is essential. These risks include data breaches, misuse of information, and concerns about how collected data might be used or shared with third parties.

Transparency in data practices is crucial for educators, students, and parents to understand how data is collected, used, or shared. Transparency allows stakeholders to assess whether data is being used responsibly and for the intended purposes. In educational settings, informed consent is vital for building trust with students and parents regarding how their data will be used (Zeide, 2017). Items that need to be considered and discussed include the sensitivity of information being collected, the potential harm if that information falls into the wrong hands, and whether individuals have the right to opt-out of data collection. Balancing the need to leverage data for personalized learning while respecting individuals' privacy rights is of utmost importance. Without proper safeguards in place, there is a risk that the benefits of big data will accrue to a few, while the most vulnerable students remain unprotected.

Concerns about privacy and the ethical use of data are not new; they have emerged throughout each stage of technological advancement. The rise of big data creates new challenges for educational institutions, particularly in developing policies for the ethical use of data. To leverage the potential of data while mitigating risks, considerations must be given to data ownership, storage, informed consent, and the impact of data use. It is essential to explore policy options that promote ethical practices in data use for customized learning experiences. Ultimately, there is a need for an ethical framework that ensures data is used responsibly, respects privacy rights, and minimizes harm.

This section homes in on data security and the anonymity of student information within educational systems. As schools gather more and more data on students, protecting that data from breaches or unauthorized access becomes vital. Because so much sensitive data is stored, it is crucial for schools to establish trust with parents and the public—without that trust, schools cannot function as they are intended. To help schools uphold that trust, a number of best practices can be pursued (Zeide, 2017). First and foremost, educational systems should undergo a security assessment, which includes a thorough check for potential vulnerabilities and the implementation of recommendations from that audit. It should also develop a security plan that explains how data security risks will be mitigated. This security plan should have strict access controls in place that dictate who has access to what data, and those controls should be enforced through security tools and regular access check audits. Finally, it is recommended that systems implement encryption techniques, preserving data in an unreadable format unless an authorized user applies a cryptographic key to unlock it. Educational systems can apply these protections which comply with both legal and regulatory standards.

Another factor to consider is that data security is not the same thing as data anonymity. A data breach may cause real problems under both scenarios, but the nature of those problems differs. Educational systems may face a greater challenge ensuring data security than consideration of educational data involves a complex combination of technology and ethics (Rubel & M. L. Jones, 2017). Once a school data system is in place, it becomes much harder to prevent misuse of data. Having a school system collecting big data on students may enable good things, but it also opens the door to data being used for nefarious purposes. This concern is amplified by the fact that data security is not something that schools can just set up once and forget about. Data security protocols should be in place, but data risks can also arise due to factors beyond a school's control. Hackers may come after data, but data security oversight must account not only for intentional security breaches but also for unintentional breaches. Schools may wish to turn to outside vendors to monitor data security, but bringing in a third party to oversee security checks may raise ethical concerns about who then has access to that data. It may be difficult to ensure that sensitive data never leaves the school or school system. Finally, it would help if systems were continuously monitored for potential vulnerabilities. Data breaches have often been tied to failure to recognize an easy opportunity for exploitation.

Moving on to the ethical dimension, the third chapter delves into the importance of informed consent and transparency when it comes to big data usage in education. Informed consent refers to students (and parents) having a comprehensive understanding of what data is being collected, how, and for what purpose, before giving consent to educational institutions to collect that data. Meanwhile, transparency concerns how well educational institutions inform students (and parents) regarding their data collection practices. The significance of transparency is highlighted, arguing that schools can gain the trust of students and parents by fully informing them about data collection practices. On the other hand, a lack of transparency creates potential for educational institutions to misuse the data, causing harm to students. Thus, this should be a central focus of educational institutions implementing big data (Zeide, 2017).

Vigilance concerning informed consent is especially necessary because educational institutions differ from other organizations that collect big data. Many such organizations collect data from young individuals who are active participants able to give consent. In contrast, educational institutions generally collect data from minors who may not have the capacity to give informed consent. Therefore, parents often need to give consent on behalf of their children. In many cases, the need to acquire parental

consent complicates the situation. In K-12 education, for instance, a single school district may have numerous schools that cater to students from different cultural and socio-economic backgrounds and speak various native languages — each of which brings its own challenges concerning how to appropriately inform parents about data collection. Moreover, even with parental consent, the ability to fully comprehend the implications of data collection may be lacking, similarly due to varying cultural and socio-economic backgrounds. In fact, it is not uncommon for parents to blindly consent simply because they lack the ability to fully understand documents explaining data collection practices, which are often lengthy and complex legal texts. Hence, in educational settings, the process of acquiring informed consent is inherently complex.

Such complexity does not excuse educational institutions from the ethical responsibility of ensuring informed consent, as this responsibility is even greater in educational contexts where institutions have considerable power over students. Institutions must clearly state to students how the data they generate will be used, who will have access to it, and what the implications are for them as students. In light of potential harm that could befall students, such as data being used against them, transparency is crucial. The chapter outlines best practices for educational institutions to establish policies concerning the collection and usage of big data, with the goal of ensuring informed consent and transparency as much as possible. Prioritizing these principles is essential for educational institutions, as it is the ethical dimension that must be considered first when adopting a new technology. In fact, discussions surrounding the ethical dimension are necessary for a data culture to develop in the first place. Therefore, educational institutions should proactively strive to ensure ethical data management as they begin to collect and utilize data in informative ways.

Future Trends and Directions in Big Data and Personalized Learning

To conclude, attention is drawn to future trends and directions surrounding big data and personalized learning in education. After briefly outlining some of the technologies currently on the horizon that are likely to shape the educational landscape, consideration is given to their potential to further enhance personalized learning experiences. Emerging technologies—such as artificial intelligence, machine learning, and augmented reality—have the exciting potential to create even more personalized learning experiences than what is achievable with big data analytics alone (Luan et al., 2020). Learning designers and education providers will need to stay attuned to these emerging technologies and how they might be integrated into the learning ecosystem to provide novel personalized learning pathways.

The role of educators in a data-driven landscape will also evolve. Undoubtedly, big data and predictive analytics will change how educational institutions, and their data systems operate, shifting pedagogical decisions away from educators and putting control of the learning environment into the hands of learning analytics and education designers. With this in mind, it is crucial for educators to remain at the heart of the learning ecosystem, informed about how data are captured. More importantly, there is a need for ongoing training and professional development for educators to adapt to these data-driven systems. Data literacy is an important competency for educators to implement new learning designs and understand the possible limits of learning analytics systems. Continuous learning will need to be prioritized as learning designer roles shift and new technologies are integrated into the learning landscape.

Challenges arise in educating future professionals to be responsive to rapid technological advancements. Who shapes the direction of future education so that it continues to support individual learning needs? What are the implications for educational ethics and equity? Learning designers and education providers will need to foster collaboration among all stakeholders, including software developers, data scientists, educators, and learners, so that the needs of each stakeholder group are taken into account when shaping future learning ecosystems. Ultimately, despite the challenges, there is optimism about the future trends of big data and personalized learning. Educational practice continues to take on board emerging technologies, and with this continuous evolution, there are opportunities for big data to enhance the efficacy of education. Finally, educators and educational institutions are encouraged to think about how they might embrace this evolution in the context of big data and personalized learning.

Bibliographic References

Ashrafimoghari, V. (2022). Big Data and Education: using big data analytics in language learning. https://arxiv.org/pdf/2207.10572

Ballard, J. & Butler, P. (2011). Personalised Learning: Developing a Vygotskian Framework for E-learning. https://core.ac.uk/download/pdf/2732182.pdf

E. August, S. & Tsaima, A. (2021). Artificial Intelligence and Machine Learning: An Instructor's Exoskeleton in the Future of Education. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7948000/

Fu, R., Tian, M., & Tang, Q. (2022). The Design of Personalized Education Resource Recommendation System under Big Data. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC9256372/

Luan, H., Geczy, P., Lai, H., Gobert, J., J. H. Yang, S., Ogata, H., Baltes, J., Guerra, R., Li, P., & Tsai, C. C. (2020). Challenges and Future Directions of Big

Data and Artificial Intelligence in Education. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7604529/

Muhammad Nda, R. & Bin Tasmin, R. (2019). Big Data Management in Education Sector: an Overview. https://core.ac.uk/download/220112831.pdf

Nitkin, D. (2018). Technology-Based Personalization: Instructional Reform in Five Public Schools. https://core.ac.uk/download/161459315.pdf

R. Kshirsagar, P., B. V. Jagannadham, D., Alqahtani, H., Noorulhasan Naveed, Q., Islam, S., Thangamani, M., & Dejene, M. (2022). Human Intelligence Analysis through Perception of AI in Teaching and Learning. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC9206552/

Rubel, A. & M. L. Jones, K. (2017). Data Analytics in Higher Education: Key Concerns and Open Questions. https://core.ac.uk/download/187116087.pdf

Wang, Y. (2017). Education Policy Research in the Big Data Era: Methodological Frontiers, Misconceptions, and Challenges. https://core.ac.uk/download/214057778.pdf

Zeide, E. (2017). The Limits of Education Purpose Limitations. https://core.ac.uk/download/214392814.pdf