Revista Brasileira de Informática na Educação – RBIE
Brazilian Journal of Computers in Education
(ISSN online: 2317-6121; print: 1414-5685)
https://sol.sbc.org.br/journals/index.php/rbie

Submission: 09/Feb/2023; Camera ready: 19/Jun/2023; 1st round notif.: 16/Mar/2023; Edition review: 25/Jun/2023; New version: 10/May/2023; Available online: 25/Jun/2023; 2nd round notif.: 10/Jun/2023; Published: 25/Jun/2023;

Offline Artificial Intelligence for Education: a path to a more inclusive field

Elyda Laisa Soares Xavier Freitas
CESAR School — UPE
ORCID: 0000-0001-7439-9040
elyda.freitas@upe.br

Ig Ibert Bittencourt
Harvard University — UFAL
ORCID: 0000-0001-5676-2280
ig.ibert@gmail.com

Seiji Isotani
Harvard University — USP
ORCID: 0000-0003-1574-0784
sisotani@icmc.usp.br

Leonardo Marques
UFAL
ORCID: 0000-0003-4993-4587
leonardo.marques@cedu.ufal.br

Diego Dermeval
UFAL
ORCID: 0000-0002-8415-6955
diego.matos@famed.ufal.br

Alan Silva
UFAL
ORCID: 0000-0002-1319-6992
alanpedro@ic.ufal.br

Rafael Ferreira Mello
CESAR School — UFRPE
ORCID: 0000-0003-3548-9670
rflm@cesar.org.br

Abstract
The field of Artificial Intelligence (AI) has the potential to improve teaching and learning, for example, by analyzing data produced in educational environments. Furthermore, it can also worsen inequality, as it requires students and instructors to have access to the infrastructure (smartphones or computers) required by the majority of those tools to generate and analyse data. However, access to such infrastructure is not a reality for many students around the world. To shine a light on this problem, this paper investigates, through a Systematic Mapping Study (SMS), initiatives that enable a more inclusive data analysis using AI in education, especially in scenarios with few connectivity resources. We identified that these initiatives are scarce and they are focused in the first phase of the data analysis task: the data collection. Based on the SMS results, we propose a set of recommendations for researchers to offer directions toward a more inclusive analysis of educational data using AI.
Keywords: *Educational Data Analysis; Digital Equity; Offline*

Cite as: Freitas, . E. L. S. X., Bittencourt, I. I., Isotani, S., Marques, L., Dermeval, D., Silva, A. & Mello, R. F. (2023). Offline Artificial Intelligence for Education: a path to a more inclusive field. Revista Brasileira de Informática na Educação – RBIE, 31, 307-322. DOI: 10.5753/rbie.2023.3156.

1 Introduction

Since the end of the 1970s, studies focusing on the investigation of educational data can be listed [5]. A factor that contributed to the growth of research in the area was the popularization of the Internet in the 1990s, allowing online education to become an important branch of educational technology [2] and producing an unprecedented amount of educational data. In the seventies, emerged the application of AI for education [8]. The analysis of educational data using Artificial Intelligence (AI) techniques is the main focus in areas such as Learning Analytics, Educational Data Mining, and Intelligent Tutoring Systems.

The literature contains a vast list of benefits in analyzing educational data, such as: improving students' and instructors' performance, improving the courses' curriculum, and understanding learner behaviour [18]. These applications allow the development of the full potential of their users, aiming for the improvement of education - both in students' and instructors' views. A key aspect of enabling the data analysis is collecting digital "footprints" from students' and instructors' interactions with online environments and providing timely feedback to allow the identification of difficulties and then promoting changes in the attitude aiming to reach better results. However, it requires Internet access and devices such as computers or smartphones, which is not a reality in many places around the world.

A recent UNESCO report [22] details that two-thirds of the world's school-age children have no internet access at home. Even in a developed country such as the United States of America, most families report "insufficient and unreliable access to the internet and internet-connecting devices" [11]. This phenomenon is also known as Digital Divide [3]: a division between people who have access to and use digital media and those who do not. Through digital media, we can understand information and communication technology. These difficulties prevent the students from having access to technologies that have the potential to improve their educational outcomes, increasing the distance to the 2030 United Nations goal for education: "Ensure inclusive and equitable quality education and promote lifelong learning opportunities for all" [23].

Although the inequality in access to education is a well-known fact, the discussion about the digital divide caused by the lack of access to the analysis of educational data using AI is still in its infancy. Therefore, this paper has two main goals: (i) to identify initiatives that allow the analysis of educational data using AI in contexts with few connectivity resources[1] through a Systematic Mapping Study; and (ii) to propose a set of recommendations for researchers, aiming to offer directions for the development of applications that allow the analysis of educational data using AI techniques to improve teaching or learning in the context where users have connectivity difficulties. In a nutshell, it includes efforts to collect data in regions with connectivity limitations, the exploitation of handwritten activities (since the paper is accessible for students with low access to digital learning facilities [19]), and also the adaptation of AI techniques to lower the use of bandwidth.

Our ultimate goal is to understand the means to foster opportunities for all the students to participate in projects that apply AI to education, as mentioned in [6]: "it is important to consider student participation in relation to the opportunities to participate being made available to students".

[1]As "few connectivity resources," we consider those with bad or no internet connection.

2 Towards Digital Equity

With respect to the aforementioned problem of lack of access, the other side of the coin is Digital Equity: “the social-justice goal of ensuring that everyone in our society has equal access to technologies” [21]. The authors [21] also mention that it refers not only to access to software and equipment, but it also includes questions such as the existence of educators skilled in using these resources effectively for teaching and learning.

Equity in education, in general, refers to the completion rates, the transition from one educational level to another, and to overall educational achievement across different groups, based on factors such as gender, income, geographic location, minority status, and disabilities [7]. It is worth mentioning that online learning, as well as the analysis of educational data using AI, has the potential to contribute to reducing these inequalities, if it is used carefully.

However, even the processes associated with the analysis of educational data using AI suffer from inequalities: Gašević [7] mention, for example, the lack of models[2] that contemplate the reality of those who cannot produce data because they do not have access. Cell phones may have helped close the gap to some degree [21] but a large portion of the population is still disconnected. And although using cell phones can be a path to include more groups, this is not the only relevant aspect.

Based on the above considerations, the effort to promote opportunities for those with connectivity difficulties due to the lack of financial or social support, to have their educational data analyzed by AI techniques is an important step toward Digital Equity. This study is an effort to contribute to this theme, aiming to encourage other researchers to engage in this specific problem through the development of inclusive applications for educational data analysis using AI.

Our research goal is to identify initiatives reported in the literature that allow the application of AI techniques on educational data aiming to reach learners in places with few connectivity resources. Then, we intend to propose a set of recommendations for researchers, aiming to offer directions toward a more inclusive analysis of educational data using AI.

2.1 Related Works

As we intend to discuss equity in the access to the analysis of educational data using AI where Internet connectivity is a problem and also identify initiatives that allow a more inclusive use of AI in this context, other studies have similar goals. One of these is the work of Holstein and Doroudi [10]. In this paper, the authors make a large discussion about the risks of amplifying existing inequities with the use of Artificial Intelligence in Education (AIED). They cite, as an example of error, the use of datasets that reflect historical inequities. The paper ends by listing pathways to help reduce inequalities as well as recommending the use of AIED to support filling the gaps of education inequalities.

However, the most related to our work is the one of Prinsloo and Kaliisa [20]. Assuming that Learning Analytics (LA) “has been highlighted as a field aiming to address systemic equity” and also “in line with a commitment from SoLAR (Society for Learning Analytics Research), as

[2] A model can identify patterns, make predictions, or detect associations in data [7]. In this sense, these models work better with the specifications of data they were supplied.

well as the International Conference on Learning Analytics and Knowledge (LAK), to increase participation from marginalized groups and communities", the authors aim to identify LA research developed in the African continent. The study employs a scoping review and found only 15 studies. Although the authors could not conclude how is the real state of LA adoption in the African continent, they contribute by reflecting on the need to support Global South with LA initiatives considering their reality and context and avoiding the data colonialism problem. Furthermore, they discussed the challenges that need immediate attention to the adoption of LA in the African continent. This research is especially different from ours since they focus on Learning Analytics in the African continent when we are interested in any initiative that employs AI for education.

3 Method

We selected the Systematic Mapping Study (SMS) to support reaching our research goal. As [13] states about the development of reviews: "The first phase addresses the task of designing how the study is to be performed, with this being documented through the review protocol."

A review protocol includes, among other things, the research questions, which guide the research development, and the search strategy that specifies how to get the most relevant set of papers. Lastly, it informs how the obtained data will be analyzed. This section presents in detail the review protocol, as suggested in [13].

3.1 Research Questions

Based on the contextualization described in the previous sections, the following research question was formulated:

> *What initiatives reported in the literature allow the application of AI techniques on educational data in places with few connectivity resources?*

The research question above follows the examples of research questions for mapping studies, available at [13]. The question encompasses Artificial Intelligence areas whose focus is the analysis of educational data as well as disconnected scenarios. We consider "disconnected scenarios" those places with bad or no internet connection.

To facilitate data extraction from the resulting papers, the following sub-questions were also formulated [13]:

- Which techniques/algorithms were applied?
- In which countries did these initiatives take place?
- Which vulnerable groups were reached?
- What benefits were reported?
- What challenges were reported regarding the development of the proposal?

- Did such initiatives help to promote greater equity (inclusion) in obtaining benefits from applying these data analysis techniques?

3.2 Search Strategy

3.2.1 Search string

According to the research question, the search string was defined to reach the most complete set of primary studies. We included in the string the different areas that use educational data. The search string follows.

> (Learning Analytics OR Educational Data Mining OR Intelligent Tutoring Systems OR artificial intelligence in education) AND (Offline OR Unplugged OR Disconnected)

3.2.2 Source of information and search strategy

An automatic search [13] was performed in the following digital libraries: Scopus, Web of Science, IEEE, ACM, Science Direct, Springer, Taylor and Francis online, and Wiley Online Library. These sources were selected since they store the proceedings of important conferences on the target theme of this research, such as Artificial Intelligence in Education (AIED), and Learning Analytics & Knowledge (LAK). The search did not limit the period of time. We also performed a manual search on the 2022 proceedings of "Artificial Intelligence in Education," "Educational Data Mining," and "Intelligent Tutoring Systems" conferences. Although the proceedings of these conferences are inserted in the results of the search string, this decision was made since these important international conferences on the theme occurred after the automatic search. The snowball strategy (backward and forward) was also adopted in this study.

3.2.3 Search procedure

Two researchers participated in the procedure. First, the search string was performed on each search engine. All the results were included on Rayyan[3], a tool to facilitate the development of systematic reviews. Rayyan automatically identified possible repeated entries, and researchers could verify and remove them.

Both researchers read the papers' titles and abstracts. During this phase, conflicts regarding the inclusion or not of a paper were decided by its inclusion. Then, both researchers evaluated this result by reading the introduction and conclusion of each paper, as well as applying the inclusion and exclusion criteria (presented below). At this time, conflicts cannot persist.

Both researchers also manually searched the selected conferences, following the same procedures. Lastly, the snowball strategy was performed by two researchers based on the list of references cited in the included papers.

[3] https://rayyan.ai/

3.2.4 Selection criteria

The selection criteria determine which study should be included or excluded from the review. The inclusion criterion was: (i) Studies that included theoretical and/or practical applications that allow the analysis of educational data (at any phase) in environments with few connectivity resources. The exclusion criteria are: (i) Papers that are not in English; (ii) Studies not available in digital format; (iii) Duplicate studies; (iv) Publication venues other than conferences and journals (periodicals and magazines); and (v) Secondary studies.

3.2.5 Data analysis

Data were analyzed by using thematic synthesis [1]. According to [1], thematic synthesis "identifies the recurring themes or issues from multiple studies, interprets and explains these themes, and draws conclusions in systematic reviews." It has the following steps: extract data, codify data, translate codes into themes, create a model of higher-order themes, and assess the trustworthiness of the synthesis.

4 Results

This section presents the results obtained through the Systematic Mapping Study, aiming to answer the research question.

4.1 Performing the Review

The search string was performed on the digital libraries in June 2022, selecting the search only in metadata (title, abstract, and keywords - when it was available in the search engine). Table 1 shows the number of documents retrieved in each search engine. Table 2 shows the number of papers in each research phase. It is possible to verify that this search retrieved many books (especially from Springer). Also, the phase of reading titles and abstracts greatly reduced the number of articles, since most of the works did not mention the topics of interest for this research. In this phase, reviewers had only two conflicts, that were resolved through discussion. In the end, we included nine papers that fully meet the developed criteria.

Except for the duplicated studies and the books retrieved in the search, most of the articles were excluded for not meeting the inclusion criteria (72%) or because they were identified as secondary study (17%).

4.2 Papers' overview

The selected studies propose various solutions to use data analysis in education in scenarios with low connectivity or resources. As shown in Table 3, although some authors have been interested in this theme since 2001 (11%), most papers were published after 2014 (89%). This could be due to the influence of the 2030 ONU Sustainable Development Goals, which were released in 2015. Eight papers (89%) were published in conferences, and only one (11%) in a journal (#3). Besides,

Table 1: Number of documents per search engine.

Search Engine	Quantity of Documents
Springer	821
Scopus	66
Web of Science	46
IEEEXplore	21
ACM	8
T and F	3
Science Direct	2
Wiley	1
Total	968
Total of unique entries	881

Table 2: Quantity of papers in each phase of the research.

Phase	**Quantity of papers**
Initial search (automatic search only)	881
After excluding duplicates and books	431
After reading the title and abstract	16
After manual search	19
After applying inclusion and exclusion criteria	7
After snowball	9

the authors were repeated in two papers: #5 and #7.

#1 [12]: This short paper describes part of a system called TILE (a web-based system for distance education). According to the authors, they were "developing a suitable way to capture interactions over the Internet and to provide a continuous interaction pattern for a given student, even in offline mode." So, TILE allowed students to have an adaptive profile according to their history and preferences. Offline adaptivity occurred due to the maintenance of a group student model based on students' behaviours (in addition to an individual student model) whose updates were managed by a mobile agent.

#2 [4]: Since many educational activities occur in offline classrooms, in this paper, authors are concerned about "collecting learning analytics data in the offline settings." They propose an approach to collecting data from textbooks taking advantage of classroom observation apps (such as LessonNote) and video recording. They also present the results of an intervention study in K-12 education. However, they reported that "'offline' interactions with LessonNote app did not offer satisfactory results." The authors list some reasons: LessonNote has interface problems, it does not allow quick documentation of activities, it does not show the dyadic interactions, and others.

#3 [9]: Hillier describes a portable platform (called MOLEAP) for digitally disconnected students and educators that allow using an offline Learning Management System. The platform includes an offline Moodle and all the resources (such as an operational system, office suite, and graphical editor) to use it on legacy computers. When connected to the internet, MOLEAP communicates responses and updates data.

#4 [24]: Authors aim to use learning analytics by collecting users' metadata (such as filename, timestamp, last modified, icons, collaboration, etc) from the "One Laptop per Children" program in India, Nepal, and Jamaica and then reporting their behaviour when using these computers. Authors get the requirements for their application by interviewing stakeholders. Some report requirements were described: "For instance, the parents would ask for reports on their child's activities across the day, the week, and the academic term. However, the teacher would want similar reports, but aggregated across the entire class." The authors consider scenarios with poor Internet connectivity. The school could be considered a point for sending data to the central location for processing (it is possible even if the Internet connectivity is intermittent).

#5 [15]: This paper proposes a platform that supports collecting educational data from learners without Internet connectivity. The proposal uses crowdsensing techniques[4] and active learning (a method in machine learning) to connect users in regions with low connectivity, minimize battery consumption, predict their performance, and then provide appropriate learning contents and feedback.

#6 [25]: This paper adopts deep learning to recognize off-line handwritten chemical structure formulas aiming to allow automatic correction. The authors composed their own dataset (since "there is no public off-line handwritten cyclic compound structure formula data set available yet"). Their proposal achieved 89.5% of accuracy in interpreting chemical cyclic compound structure formulas with single ring.

[4] "Crowdsensing, also referred to as mobile crowdsensing, is a technique that is enabled by a large group of individuals having mobile devices. These devices must be capable of collecting, computing and sharing information that can be used to analyse, measure, impute, or predict in relation to a data collection task" [17]

Table 3: Papers included in the SMS.

#	Title	Authors	Year
1	Student Adaptivity in TILE: A Client-Server Approach	Kinshuk, Binglan Han, Hong Hong, and Ashok Patel [12]	2001
2	Observing the use of e-textbooks in the classroom: towards “Offline” Learning Analytics	Maka Eradze, Terje Väljataga, and Mart Laanpere [4]	2014
3	Bridging the digital divide with an off-line e-learning and e-assessment platform	Mathew Hillier [9]	2018
4	An Architectural Design for Learning Analytics in Remote Education Environments	Sameer Verma, Andreas Gros, and Martin Dluhos [24]	2018
5	An Intelligent Platform for Offine Learners Based on Model-Driven Crowdsensing Over Intermittent Networks	Shin’ichi Konomi, Lulu Gao, and Doreen Mushi [15]	2020
6	A Component-detection-based Approach for Interpreting Off-line Handwritten Chemical Cyclic Compound Structures	Yifei Wang, Ting Zhang, and Xinguo Yu [25]	2021
7	Designing a Distributed Cooperative Data Substrate for Learners without Internet Access	Shin’ichi Konomi, Xiangyuan Hu, Chenghao Gu, and Doreen Mushi [14]	2022
8	Combining Artificial Intelligence and Edge Computing to Reshape Distance Education (Case Study: K-12 Learners)	Chahrazed Labba, Rabie Ben Atitallah, and Anne Boyer [16]	2022
9	Equitable Access to Intelligent Tutoring Systems Through Paper-Digital Integration	Nirmal Patel, Mithilesh Thakkar, Bansri Rabadiya, Darshan Patel, Shrey Malvi, Aditya Sharma, and Derek Lomas [19]	2022

#7 [14]: The authors propose a system to collect educational data in places with low connectivity. The strategy includes: (i) mobility-based data sharing to collect data; (ii) spatial analytics to map a geographic region; (iii) blockchain for secure and private data exchange; and (iv) InterPlanetary File System for storing and sharing data. It is possible to exchange data with the help of a person who carries a mobile phone through the covered areas. Data may be used to improve the satisfaction and performance of learners.

#8 [16]: This paper proposes an architecture for distance education that allows the use of Machine Learning to predict students' performance. Using Edge Computing, the approach maintains the quality of data analysis on devices with limited memory capacity and provides better use of network resources. The authors also present a case study for predicting the failure of K-12 learners.

#9 [19]: This paper uses commercially available handwriting recognition systems to digitalize and recognize offline algebra studies made on paper. The ultimate goal is to provide access to Intelligent Tutoring Systems capabilities: "Based on the student response, the ITS can produce feedback in a printable document form that we can deliver to offline students by printing it where the internet is available".

The next section presents the answers to the drawn-up Research Question.

4.3 Answering Research Question

The papers included in this SMS approach the problem from very different perspectives. First of all, there is no uniformity in using the term "offline." While some studies consider the Digital Divide problem when planning their solution [19] [9] – and their focus is on reducing this inequality – there is a paper that considers the classroom itself as "offline" [4], disregarding the access conditions of the students or instructors. We provide answers to the research question raised in this paper in the following sections.

4.3.1 Which techniques/algorithms were applied?

Although it could be expected that data mining algorithms were mentioned to answer this question, the included papers brought a very diverse list of techniques to support the use of data analysis in education when the internet connection is not easily accessible. In the paper #1 [12], the oldest one, whose focus was to adapt to each student a system for distance education, the adopted techniques are related to traditional web development, and there is no mention of Artificial Intelligence at all. The term "offline" here refers to the use of information collected about a group with common behaviour to keep the function of adaptation working even when it is not possible to get individual information. In Paper #2 [4], the offline scenario is the classroom itself, and the authors use classroom observation apps to collect data. In turn, paper #3 [9] is heavily concerned about the Digital Divide. The author uses as a strategy the gathering of all the components necessary to use the Learning Management System inside a USB stick. Data from students' activities are sent when the internet is available. It is also what happens in the paper #4 [24] that proposes a solution to solve the connectivity problem and analyze data using the CouchDB platform[5]. On the other hand, other researchers are interested in promoting more opportunities for an internet

[5] https://couchdb.apache.org/

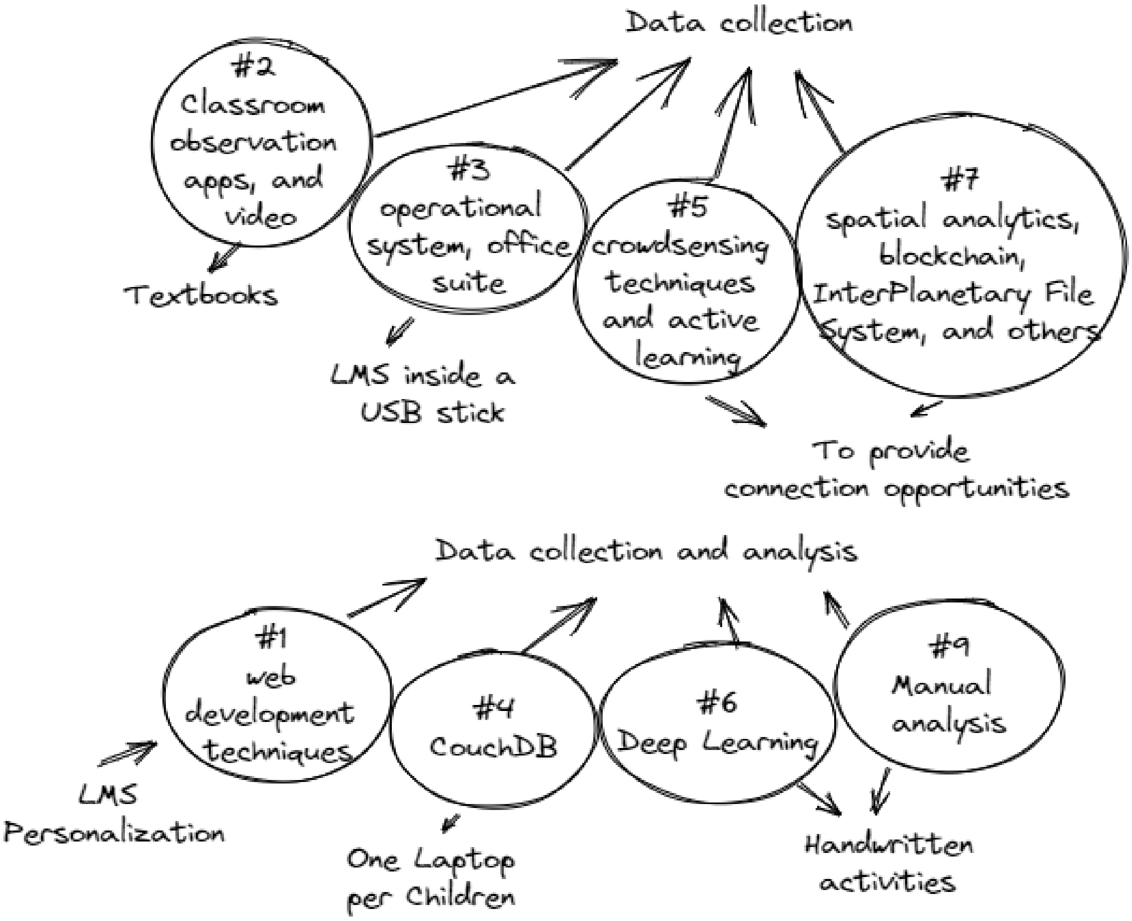

Figure 1: Techniques applied in the included papers

connection to allow data capturing: Papers #5 [15] and #7 [14], that share some authors, focus on the infrastructure to include those without internet connectivity employing technologies such as crowdsensing, blockchain, and InterPlanetary File System. Another strategy shared by the papers #6 [25] and #9 [19] is to focus on handwritten student activity. The former analyses data using deep learning, the latter mainly manually. Finally, in paper #8 [16], regarding the better use of networks, the authors use Edge Computing and a federated machine learning model for data analysis. Figure 1 summarize this information.

4.3.2 *In which countries did these initiatives take place?*

Most of the initiatives have not mentioned this information. However, some of them are specifically focused on a group of countries. Paper #4 [24] cites India, Nepal, and Jamaica. Paper #5 [15], Japan, and Tanzania. Paper #7 focuses on Tanzania. And paper #9 [19] reported a focus on India. Lastly, paper #8 [16] exploits the paper as a system with learners dispersed over 173 countries.

4.3.3 *Which vulnerable groups were reached?*

This is also a piece of information not commonly available in the papers included. Paper #3 [9] concerns students in remote and developing regions. Paper #5 [15] mentions "offline learners including people in developing communities and older adults." Paper #8 [16] focuses on "K-12 learners adopting 100% online education (...) these learners have different demographic profiles and are unable to attend regular schools for many reasons".

4.3.4 *What benefits were reported?*

Since the papers included report research that was still in the prototyping phase at the time of publication, it is not possible to discuss the benefits they already demonstrated. In general, the papers discuss possible advantages that can be achieved when using the solutions. It is the case of the paper #3 [9], that lists as benefits the possibility for their users to use the Moodle platform in a fully disconnected environment. It is possible to print the activities, and use MOLEAP on computers with limited resources, even on computers where the hard disk drive may have been damaged or removed. Regarding paper #4 [24], authors explain that their platform is "able to take the measured data and collect it from the OLPC XO laptops, analyze it to provide aggregate data, and report the information through appropriate visualizations." As a Learning Analytics solution, paper #5 [15] explains that it is possible to improve learning materials and tools and provide relevant feedback to learners, instructors, and other stakeholders. Paper #8 [16] reported good results in the experiments aiming to predict at real-time student failure. Lastly, paper #9 [19], in turn, claims that "paper-digital integration can help us preserve the benefits of writing while keeping digital systems informed about student learning."

4.3.5 *What challenges were reported regarding the development of the proposal?*

The authors of the included papers, in general, were not concerned with discussing the challenges found during the development of their research. Although some challenges are reported, they are not related to the development or execution of their proposal. For example, paper #5 [15] has a specific section called "challenges" where the authors discuss, as an example, the difficulties of e-learning implementation in Tanzania.

4.3.6 *Did such initiatives help to promote greater equity (inclusion) in obtaining benefits from applying these data analysis techniques?*

Another restriction caused by the lack of solutions already running in the real world is the impossibility of answering the above question. Although all the included papers describe solutions that have the potential to promote Digital Inclusion, their results will only be available in the future.

5 Discussion

AI applicability, its implications, as well as the ethics and equity in its use, have been a concern for a diverse range of researchers and institutions worldwide. For example, the Globalpolicy.AI, an international consortium that aims "to help policy makers and the public navigate the international AI governance landscape and access the necessary knowledge, tools, data, and best practices to inform AI policy development (...) in accordance with human rights and democratic values"[6]. World Bank, one of the Globalpolicy.AI members, explains that the development of AI technologies remains concentrated in advanced economies with a strong local talent pool, a robust research and innovation base and access to capital [26]. The report [26] also explain that "countries that lack these conditions and resources are at risk of missing out on the economic and developmental benefits that could be derived from this technology", highlighting the problems associated with the unequal use of the AI.

This paper is inserted in this discussion, focusing on the educational field. We emphasize the losses those without access to the analysis of educational data using AI are subjected to. Also, we corroborate that there is still an important journey to promote equity in the use of AI in the educational field. Some reasons can be listed, based on the results of this SMS.

First of all, there are few papers proposing theoretical or practical initiatives aiming to allow the application of AI techniques on educational data in places with few connectivity resources: only 9 papers attended our inclusion criteria. Also, this area is at a very early stage since most papers report they are still in the prototyping phase. Furthermore, most of the included papers are still working to solve the problem related to the data collection phase so that these solutions will be able to be used in AI solutions in the future.

6 Recommendations

Knowing the problem of the Digital Divide in using AI and reflecting on how to make our applications more inclusive could be an initial step toward the solution of this problem. To support this reflection, and based on the SMS results, we propose a set of recommendations for researchers aiming to promote a more inclusive analysis of educational data using AI. These recommendations are listed below:

- As well as the effort that was undertaken in the works of [15] and [14] to collect data in places with low connectivity from people who would not be included in the regular formats of connectivity, try to include diverse groups in the dataset. Although it would be impossible to include data from people in all different contexts, including one more group will make the application increasingly inclusive.
- To promote more inclusive solutions in scenarios with analysis of handwritten productions, as in [25] and [19], provide tools for students to capture data offline. It could be taking pictures (or scanning) of the paper, for example. Even if sophisticated techniques such as those reported in the articles included in this research are not employed, the use of simple techniques can help to minimize the lack of access to high-capacity communication networks. Also, consider using convolutional neural networks to support the analysis of poor-quality images.

[6] https://globalpolicy.ai/en/about/

- Use compression or aggregation techniques to minimize data traffic, as in [24]. Also, consider that sending devices may be old or of poor quality, so such applications should not require large processing capacities. Beyond this, Software Engineers and Developers must consider poor connectivity as a system requirement.
- Finally, in all the scenarios that employ Learning Analytics, consider providing feedback in templates that are also readable and understandable in print formats and can be received asynchronously.

7 Conclusions

This paper presented a Systematic Mapping Study focusing on identifying initiatives that allow the use of Artificial Intelligence for Education in scenarios where users have connection difficulties. We bring an overview of the research that uses Artificial Intelligence in Education to reach those with limited internet access.

Although research in this field is taking its first steps, it is possible to envision a future with greater Digital Equity. To collaborate with this vision of the future, we have also proposed a set of recommendations for researchers to offer directions toward a more inclusive analysis of educational data using AI. It is worth mentioning, however, that Digital Equity is a special challenge for those who live in countries with difficulties related to accessing basic infrastructure, such as in the Global South.

In future work, we intend to apply these recommendations to turn inclusive a software tool to correct handwritten essays automatically that is under development by our research group.

Acknowledgments

The authors would like to thank "Fundação Parque Tecnológico da Paraíba – Fundação PaqTcPB" for funding this research.

References

[1] Cruzes, D.S., Dyba, T.: Recommended steps for thematic synthesis in software engineering. In: 2011 International Symposium on Empirical Software Engineering and Measurement. pp. 275–284 (2011). 10.1109/ESEM.2011.36, [GS Search]

[2] Devedžić, V.: Introduction to web-based education. Semantic Web and Education pp. 1–28 (2006). 10.1007/978-0-387-35417-0_1, [GS Search]

[3] van Dijk, J.: The Digital Divide. Polity Press (2020)

[4] Eradze, M., Väljataga, T., Laanpere, M.: Observing the use of e-textbooks in the classroom: Towards "offline" learning analytics. In: Cao, Y., Väljataga, T., Tang, J.K., Leung, H., Laan-

pere, M. (eds.) New Horizons in Web Based Learning. pp. 254–263. Springer International Publishing, Cham (2014). 10.1007/978-3-319-13296-9_28, [GS Search]

[5] Ferguson, R.: Learning analytics: Drivers, developments and challenges. Int. J. Technol. Enhanc. Learn. **4**(5/6), 304–317 (jan 2012). 10.1504/IJTEL.2012.051816, [GS Search]

[6] Fincher, S., Robins, A.: The Cambridge Handbook of Computing Education Research. Cambridge Handbooks in Psychology, Cambridge University Press (2019)

[7] Gašević, D.: Include us all! directions for adoption of learning analytics in the global south. Lim, C. P., Tinio, V. L. (Eds.). (2018). Learning analytics for the global south. Quezon City, Philippines: Foundation for Information Technology Education and Development. pp. 1–22 (2018), available at [link]

[8] Guan, C., Mou, J., Jiang, Z.: Artificial intelligence innovation in education: A twenty-year data-driven historical analysis. International Journal of Innovation Studies **4**(4), 134–147 (2020). 10.1016/j.ijis.2020.09.001, [GS Search]

[9] Hillier, M.: Bridging the digital divide with off-line e-learning. Distance Education **39**(1), 110–121 (2018). 10.1080/01587919.2017.1418627, [GS Search]

[10] Holstein, K., Doroudi, S.: Equity and artificial intelligence in education: Will ”aied” amplify or alleviate inequities in education? (2021). 10.48550/arXiv.2104.12920, [GS Search]

[11] Katz, V., Rideout, V.: Learning at home while under-connected (2021), available at [link]

[12] Kinshuk, Han, B., Hong, H., Patel, A.: Student adaptivity in tile: a client-server approach. In: Proceedings IEEE International Conference on Advanced Learning Technologies. pp. 297–300 (2001). 10.1109/ICALT.2001.943927, [GS Search]

[13] Kitchenham, B., Budgen, D., Brereton, P.: Evidence-based software engineering and systematic reviews (1st ed.) (2015). 10.1201/b19467, [GS Search]

[14] Konomi, S., Hu, X., Gu, C., Mushi, D.: Designing a distributed cooperative data substrate for learners without internet access. In: Streitz, N.A., Konomi, S. (eds.) Distributed, Ambient and Pervasive Interactions. Smart Living, Learning, Well-being and Health, Art and Creativity. pp. 137–147. Springer International Publishing, Cham (2022). 10.1007/978-3-031-05431-0_10, [GS Search]

[15] Konomi, S., Gao, L., Mushi, D.: An intelligent platform for offline learners based on model-driven crowdsensing over intermittent networks. In: International Conference on Human-Computer Interaction. pp. 300–314. Springer (2020). 10.1007/978-3-030-49913-6_26, [GS Search]

[16] Labba, C., Ben Atitallah, R., Boyer, A.: Combining artificial intelligence and edge computing to reshape distance education (case study: K-12 learners). In: Rodrigo, M.M., Matsuda, N., Cristea, A.I., Dimitrova, V. (eds.) Artificial Intelligence in Education. pp. 218–230. Springer International Publishing, Cham (2022). 10.1007/978-3-031-11644-5_18, [GS Search]

[17] Livingston, J., Steele, R.: A crowdsensing algorithm for imputing zika outbreak location data. In: 2017 IEEE 8th Annual Ubiquitous Computing, Electronics and Mobile Communication Conference (UEMCON). pp. 334–340 (2017). 10.1109/UEMCON.2017.8249065, [GS Search]

[18] Nunn, S.G., Avella, J.T., Kanai, T., Kebritchi, M.: Learning analytics methods, benefits, and challenges in higher education: A systematic literature review. Online Learning **20**(2), 13–29 (2016). 10.24059/olj.v20i2.790, [GS Search]

[19] Patel, N., Thakkar, M., Rabadiya, B., Patel, D., Malvi, S., Sharma, A., Lomas, D.: Equitable access to intelligent tutoring systems through paper-digital integration. In: Crossley, S., Popescu, E. (eds.) Intelligent Tutoring Systems. pp. 255–263. Springer International Publishing, Cham (2022). 10.1007/978-3-031-09680-8_24, [GS Search]

[20] Prinsloo, P., Kaliisa, R.: Learning analytics on the african continent: An emerging research focus and practice. Journal of Learning Analytics **9**(2), 218–235 (Jun 2022). 10.18608/jla.2022.7539, [GS Search]

[21] Puigjaner, R.: Progressing toward digital equity. In: Mata, F.J., Pont, A. (eds.) ICT for Promoting Human Development and Protecting the Environment. pp. 109–120. Springer International Publishing, Cham (2016). 10.1007/978-3-319-44447-5_11, [GS Search]

[22] Thompson, G.: Two thirds of the world's school-age children have no internet access at home, new unicef-itu report says (2020), available at [link]

[23] United Nations: The sustainable development goals report 2022 (2022), available at [link]

[24] Verma, S., Gros, A., Dluhos, M.: An architectural design for learning analytics in remote education environments. In: Twenty-fourth Americas Conference on Information Systems AMCIS, New Orleans (2018), available at [link]

[25] Wang, Y., Zhang, T., Yu, X.: A component-detection-based approach for interpreting off-line handwritten chemical cyclic compound structures. In: 2021 IEEE International Conference on Engineering, Technology Education (TALE). pp. 785–791 (2021). 10.1109/TALE52509.2021.9678874, [GS Search]

[26] World Bank: Harnessing artificial intelligence for development in the post-covid-19 era: A review of national ai strategies and policies. World Bank Group p. 47 (2021), available at [link]