2024

I don't know who you are, but I know what you need: Guidelines for Federated Learning in Educational Recommender Systems

Enrico Kochon
Universität Osnabrück, Osnabrück, Germany, enrico.kochon@uni-osnabrueck.de

Daniel Stattkus
German Research Center for Artificial Intelligence, Osnabrück, Germany, daniel.stattkus@dfki.de

Simon Binz
German Research Center for Artificial Intelligence, Osnabrück, Germany, simon.binz@dfki.de

Marian Eleks
Strategion GmbH, Osnabrück, Germany, marian.eleks@strategion.de

Nils Lauinger
Forschungsinstitut Bildung Digital, Universität des Saarlandes, Saarbrücken, Germany, nils.lauinger@uni-saarland.de

See next page for additional authors

I don't know who you are, but I know what you need: Guidelines for Federated Learning in Educational Recommender Systems

Research Paper

Enrico Kochon[1], Daniel Stattkus[2], Simon Binz[2], Marian Eleks[3], Nils Lauinger[4], Philipp Fukas[1,3], Ann-Kristin Claudia Müller[4], Julia Knopf[4], and Oliver Thomas[1,2]

[1] Osnabrück University, Information Management and Information Systems, Osnabrück, Germany
{enrico.kochon,philipp.fukas,oliver.thomas}@uos.de
[2] German Research Center for Artificial Intelligence, Smart Enterprise Engineering, Osnabrück, Germany
{daniel.stattkus,simon.binz}@dfki.de
[3] Strategion GmbH, Osnabrück, Germany
{marian.eleks,philipp.fukas}@strategion.de
[4] Saarland University, Forschungsinstitut Bildung Digital, Saarbrücken, Germany
{nils.lauinger,annkristin.mueller}@uni-saarland.de, julia.knopf@mx.uni-saarland.de

Abstract. The ongoing digitalization of the education sector yields great potential through the use of Artificial Intelligence but is decelerated by a necessity for privacy and security. This paper investigates the potential of Federated Recommender Systems in school education as a solution to this problem within a two-cycle design science research approach. Meta-requirements for Federated Recommender Systems are extracted from the literature and evaluated through an educational prototype. To balance the technical evaluation, practical design guidelines are articulated and evaluated by a focus group of experts resulting in tangible guidelines for practitioners and educational stakeholders.

Keywords: Federated Learning, Education, Recommender Systems, Design Guidelines, Design Science Research, Information Systems

1 Introduction

Digital technologies offer great potential for the improvement of education which is yet untapped due to regulatory specifications regarding data protection. An example is the 'Right to be Forgotten', which requires the deletion of user data upon request, impacting the availability of data for educational data mining (Hutt et al., 2023). Actions to secure data protection are critical to ensure the success and ethicality of electronic systems, especially with educational systems, which are processing highly sensitive information such as the personal data of adolescents. This core privacy requirement seems to contradict approaches to create modern artificial intelligence-based solutions where large

19th International Conference on Wirtschaftsinformatik,
September 2024, Würzburg, Germany

quantities of data must get processed to build reliable models. For instance, kinds of data are records of completed activities, tracked learning data, time spent on tasks, and learning outcomes (Chen et al., 2020). A possible solution to this issue is Federated Learning (FL) as a Privacy Preserving Machine Learning (PPML) technique to keep data local and build a global model (Truex et al., 2019). Data protection is built in by definition and the model quality reaches a comparable quality to a scenario where all information is shared. There has been limited research in the field of educational Recommender Systems (RS) applying FL. This work aims to close this gap by answering the following research questions:

RQ1: Which meta-requirements does a Federated Learning Recommender System need to fulfill to be applicable to an educational context?

RQ2: Do Federated Learning systems achieve sufficient performance to be feasibly applied in an educational context?

RQ3: Which aspects need to be considered when implementing Federated Learning Recommender Systems for education in practice?

To build from a solid foundation, preexisting research into Federated RS is aggregated from a literature review into objective meta-requirements. The educational context is intentionally left out at this point due to the limited existing research into educational federated RS and is instead added through RQ2 and RQ3 to make up a large part of this paper's addition to the knowledge base.

To answer the research questions, meta-requirements for FL RS are collected based on a literature review. Then, the FL approach gets prototypically applied in an educational context and technically evaluated. Afterwards, practical aspects for the system's implementation are collected, aggregated into design guidelines and finally confirmed through a focus group interview.

2 Theoretical Foundations

2.1 Federated Learning

A common approach to achieve AI applications with strong privacy protection in the context of Decentralized Machine Learning is FL (McMahan et al., 2017). It enables distributed computing nodes to collaboratively train Machine Learning (ML) models without exposing their own data. This work focuses on Horizontal Federated Learning (HFL). In HFL, unlike Vertical Federated Learning (VFL), the participating clients share a common or largely overlapping feature space but differ in their individual data samples (Q. Yang et al., 2019). In VFL, the samples are either shared or largely overlapping and the feature space over the different clients is different. In HFL, each participating device has its own local data set that is not shared with other participants. Through this technique, multiple participants are enabled to build a common robust model while keeping their individual data private, thus satisfying critical issues such as data privacy, security, access rights, and access to heterogeneous data (Mammen, 2021; L. Yang et al., 2020).

2.2 Needs of the German technological infrastructure in education

Since the OECD report of 2015, the German educational system has acknowledged the deficiencies in its digital infrastructure within schools (OECD, 2015). These shortcomings became glaringly evident during the COVID-19 pandemic and have since been addressed through financial initiatives like the "Digitalpakt Schule" and a series of statements and guidelines that aim to shape the school of tomorrow into a more digitalized form (European Commission, 2020; KMK, 2016, 2021). Moreover, the last decade has witnessed extensive research that informs the latest strategy paper issued by the (KMK, 2021). In this context, (Rohde & Wrase, 2022) summarize the essential factors for the successful digitalization of schools, with the first factor being: The digital infrastructure must be implemented. Data security is crucial while implementing digital infrastructure, yet there is no centralized institution for verification, leaving schools to decide. Consequently, principals are forced to take responsibility for any legal concerns and teachers along with educators need to be certain that the software they employ in the classroom is compliant with data security and privacy requirements. This is a situation where the teachers and principals must be informed to decide whether an app is data safe or not. In summary there are several key factors in the context of the challenges discussed above:

1. Security Concerns: There are major data security uncertainties in education (GEW, 2020), aligning with (Think Tank iRights.Lab, 2021) call for verified apps.

2. Data Privacy as a Priority: As highlighted by the (mmb Institut, 2021), data privacy is a paramount concern when introducing AI in education.

3. Data secure "Data-Lakes": Performant models must be trained with large quantities of data, which is difficult in the educational environment. One of the recommendations of the (mmb Institut, 2021), is to access “Data-Lakes” with secure and anonymized data to enhance the models of AI.

4. Data Literacy: There is a lack of deeper understanding of data management by users and instructors in the educational system (Bock et al., 2023).

Incorporating FL into the educational ecosystem can be a transformative step towards addressing the security and privacy concerns prevalent in modern education. By ensuring that sensitive data remains under the control of schools and educators, while also fostering data literacy and responsible technology use, FL paves the way for a more secure and privacy-respecting digital education landscape. One specific use-case this study focuses on is the implementation of FL in RS. Diverse projects address different needs in the field of education, such as academic advising recommender (Obeid et al., 2018) or a combination of RS and learning management systems (Medio et al., 2020). Effective recommendation systems such as the five reviewed studies in (Ouyang et al., 2022) generate high-quality recommendations that lead to a significant difference in the academic performance. The use of FL in education would address the mentioned key factors. The remaining questions are about the specific technological needs of FL when implementing it in the technological infrastructure of educational systems which are addressed in this paper.

3 Research Design

The challenge for developing an appropriate solution for the digitalization in German schools is approached by applying the methodology of design science research as a common paradigm of research in the Information Systems (IS) (Hevner et al., 2004) to prototypically implement FL in an educational RS.

3.1 Design Science Research

The DSR-efforts result in two design cycles as depicted in Figure 1 to reach a technical prototype and design guidelines within a clearly structured and iterative approach.

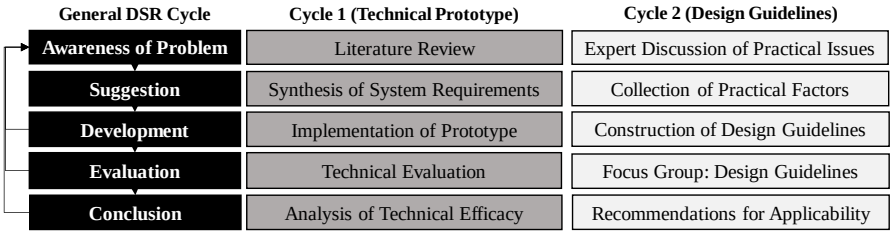

Figure 1: Design Science Research Process based on (Kuechler & Vaishnavi, 2008)

A literature review builds the foundation of the first design cycle, establishing meta-requirements for the designed artifact. It is worth noting that educational requirements are purposefully left out to ensure an optimal application architecture for FL-based RS before evaluating it in an educational context. The meta-requirements are then instantiated through an implementation of a prototype. The use case of the prototype is to recommend to the teacher which students to intervene on because they are at risk of failing their respective classes, allowing for faster and more precise interventions. This prototype then undergoes a technical evaluation to achieve an assessment of the technical efficacy of the system for the education domain. For the second cycle, an expert panel held over multiple sessions where AI, education, and software experts discussed various aspects of digitalization in education provides the initial push to collect factors regarding the practical implementation of the designed system. These factors are synthesized into design guidelines when implementing the system in practice. Following this, a focus group of experts in the fields of education, data science and security discuss and evaluate the design guidelines to provide evaluated, actionable knowledge aimed to be used as guidelines for practice.

3.2 Literature Review

The knowledge of an extensive literature review, guided by (Brocke et al., 2009; Webster & Watson, 2002) serves as source of information in the first DSR-cycle. The search term is made up of *"federated learning" AND "recommender systems" AND "requirement"* over IEEE, SpringerLink, ScienceDirect, AISeL, ACM, Wiley and JSTOR. From

the initial 352 unique results, 304 were removed after screening by title. Of the 48 remaining articles, the abstract was screened for relevance, leaving 37 articles. Finally, a full-text screening was performed for the remaining articles, resulting in 13 articles which are joined with 7 articles from forward/backward searches to make up 20 articles used as one of the informational foundations of the paper. References with certain requirements for RS employing FL were included and poor quality led to an exclusion.

4 Results

4.1 Meta-Requirements for FL-based RS in the context of Flower framework

From the literature review resulted 43 requirements, which were condensed to the nine meta-requirements depicted in Figure 2 marked with filled boxes.

Meta-Requirements	(Alamgir, Khan and Karim, 2022)	(Beutel et al., 2022)	(Cui et al., 2022)	(Harasic et al., 2023)	(Imran et al., 2023)	(Jalalirad et al., 2019)	(Jiang et al., 2020)	(Kairouz et al., 2021)	(Kalloori and Klingler, 2021)	(Liu et al., 2022)	(Luo et al., 2023)	(Muhammad et al., 2020)	(Neumann et al., 2023)	(Perifanis and Efraimidis, 2022)	(Qin et al., 2023)	(Ribero et al., 2022)	(Wang et al., 2022)	(Wen et al., 2023)	(Yang et al., 2020)	(Zhang et al., 2022)	Frequency
MR1 - Computational and Network Resources	■	■	■	■	■		■						■		■	■	■	■	■	■	13
MR2 - Model Management and Personalization	■					■	■	■		■	■	■	■				■		■		10
MR3 - Privacy and Security	■	■		■						■	■			■	■			■	■		9
MR4 - Network Architecture and Support			■	■				■				■		■			■			■	7
MR5 - Performance and Efficiency				■				■			■				■				■		5
MR6 - Data Management and Quality	■								■	■								■			4
MR7 - System Robustness and Scalability				■									■	■							3
MR8 - Adaptability and Evolution	■												■					■			3
MR9 - Device&Infrastructure								■					■								2

References

Figure 2: Meta-Requirements from the Literature Review

The Flower framework is specifically chosen, because it offers an out of the box implementations of server-client FL architecture, different methods for the aggregation of the given parameters and methods for simulation. The user then defines the model which runs in the backend and methods for accessing the parameters, training, and evaluation. The Flower framework inherently fulfils most of the meta-requirements from the literature; these are listed below according to their frequency:

MR1: As Flower operates independently of the underlying machine learning framework, the allocation of computational resources primarily depends on the selected model. Minimizing network resources is feasible by serializing transferred parameters, and smaller models typically require relatively few iterations to converge. Communication efficiency is detailed in Neumann et al. (2023).

MR2: Given that FL is generally agnostic to the underlying machine learning model, any model whose parameters can be aggregated can be trained. Personalization as Requirement for RS is mentioned in Wang et al. (2022) as another aspect of FL RS.

MR3: Flower integrates differential privacy, a technique that introduces statistical noise to confidential data, rendering it indistinguishable and unidentifiable. Data privacy and security needs are met through Flower's implementation of methods such as SSL connections. Perifanis & Efraimidis (2022) evaluate privacy preserving mechanisms in the context of FL RS.
MR4: Flower seamlessly integrates with various machine learning libraries. Examples of these in Python include Keras, scikit-learn, and PyTorch, facilitating compatibility and ease of use. Communication between clients and servers, a core assumption for FL (Kairouz et al., 2021), is a fundamental aspect of the Flower framework.
MR5: L. Yang et al. (2020) describe performance requirements, for instance real-time requirements are relevant.
MR6: RS usually rely on large quantities of data (Kalloori & Klingler, 2021). FL works with models rather than the data itself. It does not hinder data quality but needs additional steps when the need to deal with data understanding and processing arises. This is further elaborated on in the discussion.
MR7: Robustness is a goal for FL systems (Harasic et al., 2023). It relies on the underlying machine learning model, while its scalability is contingent upon the resources available to the clients, as training is distributed across different devices.
MR8: After training, a model can be distributed to clients, and upon collecting new data batches, FL can be applied to the updated data, utilizing the prior model as the starting point for the new FL iteration. Even readjusting for different requirements and constraints happens (Wen et al., 2023).
MR9: There are scenarios where a large client population has to be supported by the infrastructure (Neumann et al., 2023).

4.2 Architecture of the FL-based educational RS

As mentioned in 2.1, this paper focuses on HFL. In the context of education, different clients could for example be different courses or different schools prohibited from sharing data about the performance of their students. These clients possess data that can be utilized to train a model for predicting a student's future performance and the likelihood of passing a particular course. This data may include features such as past examination performance, class attendance or even demographic data about the students. All clients engaged in the FL framework are aware of the type of model to be trained in advance. In the case of this paper, the model predicts which students are at risk of failing the class, recommending these students to the teacher for potential interventions. For the actual FL approach, a client-server architecture is used. The process is as follows:

1. The model type to be used is predefined, and each client is informed about the specific model details (i.e., features used and underlying model architecture), along with the initial set model weights, which can be taken from a random client.
2. Subsequently, until the model converges or reaches satisfactory performance:
    a. Clients compute their model weights locally in a training iteration.
    b. Clients send their gradients or model weights to the server, with the option to mask them for secure communication.

c. The server aggregates the received results using methods like weighted averaging to create a global model.
d. The server sends back the aggregated results to the clients.

Given the relatively short duration of school courses and the diverse nature of required teacher actions based on the learning environment, it becomes crucial to directly involve the teacher in the ML loop rather than solely relying on model predictions. At its current state, the prototype is a support system for teachers making decisions.

4.3 Implementation

Using Flower this work implements the described architecture as a FL-based educational RS prototype trained on the OULAD (Open University Learning Analytics Dataset) (Kuzilek et al., 2017). It is chosen as the primary data source because of its substantial size and close relation to virtual courses and information on how the students interact with the provided material as well as student profiles, and their corresponding assessments. It is worth noting that the data preprocessing is not a primary focus of this work and will only be touched on briefly. The goal of the prototype is to recommend students who are likely to fail their class, hence two cases are considered: pass and fail. Since the students can withdraw from the course at any time, a deadline for the final withdrawal is set, after which, every withdrawal is considered as a failure. On the other hand, the label Distinction provided by the dataset is interpreted as a pass. The dataset containing information on the student's interaction with each course (i.e., the sum of clicks on each type of material until a specific date) is merged with the students' demographic data including the final result. After dropping columns with too many missing values, 30 features and 23860 data points are left (Duke, 2021).
Table 1 lists all features used for training, broken down into features regarding the course, the interaction with the virtual learning environment (VLE) and demographic data about the students.

Neural networks are employed as the underlying machine learning model and implemented through the Keras (Chollet, 2015) library in Python. In the context of FL, neural networks present an attractive choice due to their ability to easily aggregate the model updates by averaging the model weights across all clients. For similar reasons, the decision to average over the model weights directly as opposed to the model gradients is made, as they are easily accessible through the Keras library.

Table 1: All features used in training

Course	code_module, code_presentation, mean_score_day180, num_of_prev_attempts
VLE Interactions	dataplus, dualpane, externalquiz, folder, forumng, glossary, homepage, htmlactivity, oucollaborate, oucontent, ouelluminate, ouwiki, page, questionnaire, quiz, repeatactivity, resource, sharedsubpage, subpage, url
Demographic Data	gender, region, highest_education, age_band, studied_credits, disability

4.4 Technical Evaluation

The global test set used for evaluation is constructed by randomly sampling from the entire dataset. Within this context, 3 different model training scenarios are examined. In the global simulation, a single model is trained on all the available training data and is then evaluated on the global test set. In the local simulation, the data is split between *n* clients and a model is then trained for each local dataset. Each local model is then evaluated on the global test set. In the federated simulation the same data split occurs, however a single model is then trained using the FL architecture described above. The resulting model is then again evaluated on the global test set. By comparing the global and local simulation to each other, it becomes apparent how the reduction of data points affects the performance of the underlying Machine Learning model, as in the local simulation each model only trains on a fraction of the data. Similarly, comparing the federated simulation to the other simulations yields a comparison of how the proposed FL architecture performs under the same reduction in data as the local simulation.
Table 2 and Table 3 show the results of the scenario using the global test set and splitting it into 10 or 100 smaller datasets respectively. The average metrics using cross validation (k=10) are depicted with an average being taken for each model in each run of the cross validation.

Table 2: Results of the simulation (*n*=10)

Metric (averaged)	**Global Simulation**	**Local Simulation**	**Federated Simulation**
Binary Cross Entropy	**0.498 (±1.1×10^{-3})**	0.5422 **(**±5×10^{-4})	0.5094 **(**±3×10^{-4})
Accuracy	**0.77 (±1×10^{-4})**	0.7365 **(**±2×10^{-4})	0.7668 **(**±1×10^{-4})
F1 Score	**0.8262 (±1×10^{-4})**	0.8053 **(**±2×10^{-4})	0.8261 (±1×10^{-4})
Precision	**0.7926 (±3×10^{-4})**	0.7592 **(**±7×10^{-4})	0.7828 **(**±1×10^{-4})
Recall	0.8709 (±7×10^{-4})	0.8702 **(**±3.2×10^{-3})	**0.8824 (±3×10^{-3})**
Area Under Curve	**0.8278 (±1×10^{-3})**	0.7854 **(**±1×10^{-4})	0.823 **(**±1×10^{-4})

Using 10 clients (Table 2), the global model yields the best performances in terms of Binary Cross Entropy, Accuracy, F1-Score Precision and Area Under Curve, while the federated model yields the best performance in terms of Recall. For each metric, there is no large difference between the global and the federated model with both models performing significantly better than the local model, except for the Recall metric.

When splitting the data between 100 different clients, the difference between the local model and the other two models further increases. Table 3 shows that for each metric except Recall, the global model performs best followed by the federated model. The federated model outperforms the other models slightly in the Recall metric.

Table 3: Results of the simulation (n=100)

Metric (averaged)	Global Simulation	Local Simulation	Federated Simulation
Binary Cross Entropy	**0.498 (±1.1×10^{-3})**	0.6349 (±2.1×10^{-3})	0.5363 (±2×10^{-4})
Accuracy	**0.77 (±1×10^{-4})**	0.6730 (±5×10^{-4})	0.7497 (±1×10^{-4})
F1 Score	**0.8262 (±1×10^{-4})**	0.7722 (±7×10^{-4})	0.8210 (±1×10^{-4})
Precision	**0.7926 (±3×10^{-4})**	0.6941 (±6×10^{-4})	0.7502 (±1×10^{-4})
Recall	0.8709 (±7×10^{-4})	0.8873 (±6×10^{-4})	**0.915 (±2×10^{-4})**
Area Under Curve	**0.8278 (±1×10^{-3})**	0.6841 (±1.2×10^{-3})	0.8023 (±1×10^{-4})

5 Design Guidelines

To start the second design cycle, panel discussions between experts in education (2 teachers), ML (2 researchers) and IS (3 researchers) raise awareness for problems regarding the practical implementation of the technically designed and evaluated system and resulted in four design guidelines for the domain of education. These are presented in the following paragraphs.

DG1 - Achieve additional security with hybrid approaches to PPML. While FL as an approach to PPML provides sufficient privacy guarantees for a host of different attacks, it is best to combine it with other approaches like secure multiparty computation, differential privacy, or homomorphic encryption to close off other angles of attack like model or data reconstruction in a hybrid approach (Fang & Qian, 2021; Truex et al., 2019).

DG2 - Implement the system with consideration for the preconditions of schools. While digitalization initiatives in education do push schools to new heights of technical hardware and know-how, the FL system should not overburden their hardware with laborious or unoptimized machine learning algorithms. In the same way, the FL process should be implemented to be as automated as possible to not strain the already scarce personnel resources more than necessary.

DG3 – Scalability and Connectivity first. Enforcing decentralization and keeping ownership of data is no longer a challenge for ML. Thus, it is time to optimistically rethink concepts and create data lakes while preserving data sovereignty and even pushing data autonomy forward at the same time. This will enable better performing recommendation algorithms and will keep control of their data at the hand of all participants.

DG4 – Ensure that Federated Learning fits your problem. FL does not fit every use case: On one hand, if the data used is not personal and does not have a high need for protection, the added complexity of FL should be avoided. On the other hand, there are scenarios with such a high data sensitivity and ethical questionability that FL alone does not provide sufficient protection the combination with other approaches would be necessary.

The focus group meant to evaluate and extend the above design guidelines was held as an online workshop with experts from the domains of education (1 researcher), psychology (1 researcher), data science (3 researcher) and software development (3 researcher). Experts voted on the usefulness of the guidelines and gave feedback in terms of their strengths and weaknesses. In the voting, **DG2** emerges as the clear favorite with all votes fully agreeing with its usefulness and statements underlining it as a must have in the implementation of the system because of heavy time and knowledge constraints found in education.

DG1 is met with strong, albeit not unanimous, support due to its enhancement of data security and trust within the critical domain of education. However, it also brings about increased system complexity and slightly diminishes explainability.

DG3 is important, as its fulfillment ensures a higher quality standard which helps gaining a higher level of acceptance. The main drawback voiced by the experts is a bigger effort for a huge rollout, not only from a technical, but also from an organizational standpoint. Larger rollouts result in higher risks.

DG4 ensures appropriate effort for the individual problem. Experts agree that this recommendation is useful in reducing unnecessary effort and use of data. Unfortunately, it also promotes heterogeneity, thus reducing comparability between schools. It also collides with recommendation three, as the inspection of all individual cases does not scale well. Both are important in their own regard. This friction shows the field of tension when it comes to the actual implementation of the educational RS.

In addition to evaluating the presented guidelines, the experts recommended that ethicality of a use case must be evaluated before during and after the implementation for a holistic ethical view. Also, even though a wide rollout is desirable, a preceding model implementation can drastically reduce the imposed risk.

6 Discussion

This article focuses on data security and data protection in the context of RS in education. The evaluation results show how the distribution of training data and the use of different training architectures can impact model performance. As anticipated, training with a complete global dataset yields the best utility but incurs a heavy cost in data privacy due to the requirement of centralizing all data for training. The local models, on the other hand, still perform reasonably well when training data is fragmented into 10 datasets, but their performance drops significantly when fragmented into 100 datasets. This decrease in performance is due to the reduced amount and diversity of data available to each local model, effectively reducing the information available to learn overarching patterns. Local models offer the best data protection, but a question remains: Where should the line be drawn for a local model? Training these models at the individual student level would result in excessive fragmentation, effectively making the resulting models useless while training at a higher level, such as the class or school level, would again result in centralized data storage. A compromise between the two approaches is offered by FL. The federated model shows improved performance compared to the local models with the same level of privacy. In some aspects, such as F1

score and recall, it even achieves similar or slightly better results than the global model (RQ2). This demonstrates the effectiveness of the FL architecture. A joint model is trained from the contributions of multiple clients without direct data exchange. Hence it is possible to learn from the diversity of distributed data while addressing privacy concerns. FL therefore offers a solution to data protection concerns, particularly in educational settings (GEW, 2020; mmb Institut, 2021), while also leveraging the potential of AI technologies (RQ1). However, when compared to other models, FL presents a new challenge. The creation and composition of the model makes data understanding and preparation more difficult, and feature engineering impossible (Eleks et al., 2023). Therefore, it is important to develop methods for continuous evaluation of the model during operation such as the inclusion of explainable artificial intelligence methods in the federated process from the beginning to provide information about which variables have a particularly strong influence on the forecast. In this context, it is important to consider a human-in-the-loop approach: When receiving recommendations along with explanations, teachers or students can incorporate feedback into the training process. This can lead to a semi-automatic improvement of the system in the long term. The use case described in this article predicts learning performance which raises ethical concerns regarding the use of AI. This concern can also be addressed by human involvement to prevent the potential negative impact of false-positive or false-negative predictions on a student's future. Therefore, it is imperative for teachers to be involved in the process. Other usage scenarios such as assigning teaching materials or entire courses to specific students face similar issues where misjudgments can also have detrimental long-term consequences.

The evaluation with experts further highlights the importance of considering both meta-requirements and practical constraints in implementing FL in education. The focus group's preference for **DG2** highlights the importance of creating FL systems that align with the diverse technological landscapes of schools. This insight directly addresses RQ3, emphasizing the need for user-friendly and low-maintenance FL solutions tailored to the technical capabilities and resources of educational institutions. Furthermore, the group's support for hybrid privacy approaches (**DG1**) reinforces the meta-requirements for FL systems in education (RQ1), advocating for enhanced data security through advanced privacy-preserving methods. This strategy not only increases data protection but also strengthens system trust, addressing both technical and ethical considerations crucial for FL's application.

The discussion on scalability and fit-for-purpose (**DG3** and **DG4**) brings to light the challenges and considerations for FL's practical deployment (RQ2). Scalability ensures that FL systems can accommodate growing educational demands without compromising performance or privacy. Conversely, the need to tailor FL deployment to specific educational scenarios underscores the importance of discerning when FL's complexity is warranted, based on the sensitivity of the data and the educational impact. These considerations highlight that the effective implementation of FL in education requires a balanced approach, considering technical feasibility, data security, and the educational context's unique needs. In the context of IS, this article demonstrates how FL can

leverage the benefits of AI without compromising data protection which is especially important in the context of user acceptance, fostering trust into the system. This, in turn, enables the human-in-the-loop approach and ensures that the system can reach its full potential. For practitioners, the presented approach provides guidance on implementing such a system in practice that is also anticipated to be applicable to educational areas beyond school education, e.g. corporate contexts such as assigning targeted training measures to specific employees.

7 Conclusion

Based on its research questions, this article has delved into the utilization of a FL-based RS within the educational domain, particularly in schools. It sought to uncover meta-requirements essential for such a system's operation, assesses its performance against these criteria in an educational setting, and identifies essential design guidelines critical for its deployment in the domain. The four design guidelines lay the foundation for the successful integration of FL-based RS in educational environments.
The main contributions of this research are threefold: First, we identify meta-requirements essential for FL RS, ensuring privacy and performance. Second, we develop and evaluate a practical prototype that demonstrates the feasibility and effectiveness of FL in this domain, using real-world educational data. Third, we derive and validate practical design guidelines through expert focus groups, offering actionable insights for practitioners.

Like any research endeavor, this study has limitations. The technical evaluation relied on a single dataset, limiting deeper investigation. This may affect the generalizability of the findings, emphasizing the need for research with more diverse and realistic datasets. Additionally, the design guidelines, while evaluated with experts, have not yet been practically implemented. Future research should focus on implementing and evaluating these systems in live educational environments to assess their practicality, effectiveness, and their long-term impact on educational outcomes. Moreover, while the design guidelines provide a strategic framework for implementing FL in education, they also highlight the nuanced balance required between technological sophistication and practical usability. Ensuring that FL systems are both robust in privacy preservation and accessible to educational practitioners remains a key challenge. Addressing this challenge calls for ongoing collaboration between technologists, educators, and policymakers, aiming to refine current approaches to the evolving needs of the educational sector. This paper contributes to the broader discourse on the intersection of artificial intelligence, data privacy, and education by foregrounding the importance of privacy-preserving technologies like FL and showing their transformational potential for the education domain. This transformation, however, must be navigated with careful consideration of technical, ethical, and practical dimensions, ensuring that technological advancements empower educators and students alike, without compromising on the core values of security and privacy.

References

Alamgir, Z., Khan, F. K., & Karim, S. (2022). Federated recommenders: methods, challenges and future. *Cluster Computing*, *25*(6), 4075–4096.

Beutel, D. J., Topal, T., Mathur, A., Qiu, X., Fernandez-Marques, J., Gao, Y., Sani, L., Li, K. H., Parcollet, T., & de Gusmão, P. P. B. (2022). *Flower: A friendly federated learning framework.*

Bock, A., Breiter, A., Hartong, S., Jarke, J., Jornitz, S., Lange, A., & Macgilchrist, F. (2023). *Die datafizierte Schule* (A. Bock, A. Breiter, S. Hartong, J. Jarke, S. Jornitz, A. Lange, & F. Macgilchrist, Eds.). Springer Fachmedien Wiesbaden. https://doi.org/10.1007/978-3-658-38651-1

Brocke, J. vom, Simons, A., Niehaves, B., Niehaves, B., Reimer, K., Plattfaut, R., & Cleven, A. (2009). *Reconstructing the giant: On the importance of rigour in documenting the literature search process*.

Chen, L., Chen, P., & Lin, Z. (2020). Artificial intelligence in education: A review. *Ieee Access*, *8*, 75264–75278.

Chollet, F. (2015). *Keras, GitHub repository*. https://github.com/keras-team/keras

Cui, Z., Wen, J., Lan, Y., Zhang, Z., & Cai, J. (2022). Communication-efficient federated recommendation model based on many-objective evolutionary algorithm. *Expert Systems with Applications*, *201*, 116963.

Duke, E. (2021, July 17). *OULAD Random Forest*. https://www.kaggle.com/code/zincbottom/oulad-random-forest/notebook#Discussion

Eleks, M., Rebstadt, J., Kortum, H., & Thomas, O. (2023). Privacy Aware Processing. In M. K. Klein Daniel; Winter Cornelia; Wohlgemuth Volker (Ed.), *INFORMATIK 2023 - Designing Futures: Zukünfte gestalten*. Gesellschaft für Informatik e.V. https://doi.org/10.18420/inf2023_67, https://dl.gi.de/handle/20.500.12116/43189

European Commission. (2020). *Digital Education Action Plan (2021-2027)*. Https://Education.Ec.Europa.Eu/de/Focus-Topics/Digital-Education/Action-Plan.

Fang, H., & Qian, Q. (2021). Privacy Preserving Machine Learning with Homomorphic Encryption and Federated Learning. *Future Internet*, *13*(4), 94. https://doi.org/10.3390/fi13040094

GEW. (2020). *Digitalpakt Schule und Digitalisierung an Schulen*.

Harasic, M., Keese, F.-S., Mattern, D., & Paschke, A. (2023). Recent advances and future challenges in federated recommender systems. *International Journal of Data Science and Analytics*, 1–21.

Hevner, A. R., March, S. T., Park, J., & Ram, S. (2004). Design science in information systems research. *MIS Quarterly*, 75–105.

Hutt, S., Das, S., & Baker, R. S. (2023). The Right to Be Forgotten and Educational Data Mining: Challenges and Paths Forward. *International Educational Data Mining Society*.

Imran, M., Yin, H., Chen, T., Nguyen, Q. V. H., Zhou, A., & Zheng, K. (2023). ReFRS: Resource-efficient federated recommender system for dynamic and diversified user preferences. *ACM Transactions on Information Systems*, *41*(3), 1–30.

Jalalirad, A., Scavuzzo, M., Capota, C., & Sprague, M. (2019). A simple and efficient federated recommender system. *Proceedings of the 6th IEEE/ACM International Conference on Big Data Computing, Applications and Technologies*, 53–58.

Jiang, H., Liu, M., Yang, B., Liu, Q., Li, J., & Guo, X. (2020). Customized federated learning for accelerated edge computing with heterogeneous task targets. *Computer Networks*, *183*, 107569.

Kairouz, P., McMahan, H. B., Avent, B., Bellet, A., Bennis, M., Bhagoji, A. N., Bonawitz, K., Charles, Z., Cormode, G., & Cummings, R. (2021). Advances and open problems in federated learning. *Foundations and Trends® in Machine Learning*, *14*(1–2), 1–210.

Kalloori, S., & Klingler, S. (2021). Horizontal cross-silo federated recommender systems. *Proceedings of the 15th ACM Conference on Recommender Systems*, 680–684.

KMK. (2016). *Bildung in der digitalen Welt: Strategie der Kultusministerkonferenz*.

KMK. (2021). *Lehren und Lernen in der digitalen Welt: Die ergänzende Empfehlung zur Strategie "Bildung in der digitalen Welt."*

Kuechler, W., & Vaishnavi, V. (2008). On theory development in design science research: anatomy of a research project. *EJIS*, *17*, 489–504.

Kuzilek, J., Hlosta, M., & Zdráhal, Z. (2017). Open University Learning Analytics dataset. *Scientific Data*, *4*, 170171. https://doi.org/10.1038/sdata.2017.171

Liu, Z., Yang, L., Fan, Z., Peng, H., & Yu, P. S. (2022). Federated social recommendation with graph neural network. *ACM Transactions on Intelligent Systems and Technology (TIST)*, *13*(4), 1–24.

Luo, J., Yang, X., Yi, X., & Han, F. (2023). Privacy-preserving recommendation system based on user classification. *Journal of Information Security and Applications*, *79*, 103630.

Mammen, P. M. (2021). Federated Learning: Opportunities and Challenges. *ArchXiv*. https://doi.org/10.48550/arxiv.2101.05428

McMahan, B., Moore, E., Ramage, D., Hampson, S., & y Arcas, B. A. (2017). Communication-efficient learning of deep networks from decentralized data. *Artificial Intelligence and Statistics*, 1273–1282.

Medio, C. De, Limongelli, C., Sciarrone, F., & Temperini, M. (2020). MoodleREC: A recommendation system for creating courses using the moodle e-learning platform. *Computers in Human Behavior*, *104*, 106168. https://doi.org/10.1016/J.CHB.2019.106168

mmb Institut. (2021). *KI@Bildung: Lehren und Lernen in der Schule mit Werkzeugen Künstlicher Intelligenz*.

Muhammad, K., Wang, Q., O'Reilly-Morgan, D., Tragos, E., Smyth, B., Hurley, N., Geraci, J., & Lawlor, A. (2020). Fedfast: Going beyond average for faster training of federated recommender systems. *Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, 1234–1242.

Neumann, D., Lutz, A., Müller, K., & Samek, W. (2023). A Privacy Preserving System for Movie Recommendations using Federated Learning. *ACM Transactions on Recommender Systems*.

Obeid, C., Lahoud, I., El Khoury, H., & Champin, P.-A. (2018). Ontology-based Recommender System in Higher Education. *Companion Proceedings of the The Web Conference 2018*, 1031–1034. https://doi.org/10.1145/3184558.3191533

OECD. (2015). *Students, Computers and Learning*. OECD. https://doi.org/10.1787/9789264239555-en

Ouyang, F., Zheng, L., & Jiao, P. (2022). Artificial intelligence in online higher education: A systematic review of empirical research from 2011 to 2020. *Education and Information Technologies*, *27*(6), 7893–7925. https://doi.org/10.1007/s10639-022-10925-9

Perifanis, V., & Efraimidis, P. S. (2022). Federated neural collaborative filtering. *Knowledge-Based Systems*, *242*, 108441.

Qin, J., Zhang, X., Liu, B., & Qian, J. (2023). A split-federated learning and edge-cloud based efficient and privacy-preserving large-scale item recommendation model. *Journal of Cloud Computing*, *12*(1), 57.

Ribero, M., Henderson, J., Williamson, S., & Vikalo, H. (2022). Federating recommendations using differentially private prototypes. *Pattern Recognition*, *129*, 108746.

Rohde, D., & Wrase, M. (2022). *Die Umsetzung des DigitalPakts Schule: Perspektiven der schulischen Praxis auf zentrale Steuerungsfragen und -herausforderungen*.

Think Tank iRights.Lab. (2021). *Datenschutz und digitale Schule*.

Truex, S., Baracaldo, N., Anwar, A., Steinke, T., Ludwig, H., Zhang, R., & Zhou, Y. (2019). A Hybrid Approach to Privacy-Preserving Federated Learning. *Proceedings of the 12th ACM Workshop on Artificial Intelligence and Security*, 1–11. https://doi.org/10.1145/3338501.3357370

Wang, Q., Yin, H., Chen, T., Yu, J., Zhou, A., & Zhang, X. (2022). Fast-adapting and privacy-preserving federated recommender system. *The VLDB Journal*, *31*(5), 877–896.

Webster, J., & Watson, R. T. (2002). Analyzing the past to prepare for the future: Writing a literature review. *MIS Quarterly*, xiii–xxiii.

Wen, J., Zhang, Z., Lan, Y., Cui, Z., Cai, J., & Zhang, W. (2023). A survey on federated learning: challenges and applications. *International Journal of Machine Learning and Cybernetics*, *14*(2), 513–535.

Yang, L., Tan, B., Zheng, V. W., Chen, K., & Yang, Q. (2020). Federated recommendation systems. *Federated Learning: Privacy and Incentive*, 225–239.

Yang, Q., Liu, Y., Chen, T., & Tong, Y. (2019). Federated machine learning: Concept and applications. *ACM Transactions on Intelligent Systems and Technology (TIST)*, *10*(2), 1–19.

Zhang, H., Luo, F., Wu, J., He, X., & Li, Y. (2023). LightFR: Lightweight federated recommendation with privacy-preserving matrix factorization. *ACM Transactions on Information Systems*, *41*(4), 1–28.