# Learner Models:

## Requirements and Legal Issues for the Development and Application of Learner Models

Felix Böck[1] [0000-0001-7382-8333], Hendrik Link[2] [0009-0006-0596-2120] and Dieter Landes[1] [0000-0002-0741-3540]

[1] Center for Responsible Artificial Intelligence (CRAI),
Coburg University of Applied Sciences and Arts, 96450 Coburg, Germany
{felix.boeck, dieter.landes}@hs-coburg.de
[2] Public Law, IT Law, and Environmental Law,
University of Kassel, 34109 Kassel, Germany
hendrik.link@uni-kassel.de

***Abstract* — Learners differ vastly in various aspects of what they need for successful learning. Artificial Intelligence (AI) establishes a basis for digital learning environments which adapt themselves automatically to the learners' needs. To be able to do so, these systems presuppose knowledge on the individual learner. Learner models are digital representations of learner characteristics that aim to enable personalised and adaptive learning experiences, touching upon issues in key areas, such as transparency, fairness, data protection, modularity, and sustainability. Such learner models form the core of AI-based adaptive learning environments, as they store data about individual learners. This paper collects and discusses requirements, legal issues, and challenges associated with developing and using learner models, particularly in the context of European regulations. By reviewing existing standards, scientific publications, and practical use cases, we identify gaps in standardisation and propose foundational requirements for the design of interoperable and legally compliant learner models. Our findings lay the groundwork for developing a reference architecture, facilitating scalable and ethical integration of learner models in digital learning environments.**

***Keywords — learner model; learner modelling; requirements engineering; legal issues; compliance; ethical principles; higher education; learning analytics.***

## I. INTRODUCTION

Learners tend to be increasingly heterogeneous as a group since they differ in terms of individual levels of knowledge and competencies, learning styles, individual preferences for (digital) media, and various other factors [1][2]. A potential solution to accommodate this growing heterogeneity might be digital learning environments, which supports users in a given situation. This is possible on various levels. In this work, we focus exclusively on learning environments that support learners on the micro level by adapting to their specific needs in self-directed learning. Yet, adaptation presupposes some knowledge of the individual learner, which is usually stored in a learner model (aka user model). The latter constitutes a collection of user characteristics that are relevant for individual learning support [3][4]. Although a vast amount of literature exists on learner models, there seems to be no consensus concerning the content and purpose of a learner model [4] and the legal constraints that restrict the use of the information embedded in the learner model. This contribution identifies requirements that a learner model should fulfil to provide individual learning support. Requirements are derived from a comprehensive analysis of scientific publications [4] and standards on learner models [3]. In addition, we address legal issues related to the development and operation of digital learning environments that rely on personal data of learners. To do so, we take a European perspective. We aim to contribute to identifying common requirements for learner models, which could be implemented in a further step through a reference architecture, considering legal constraints. The remainder of this paper first clarifies the terminology in Section 2, before Section 3 outlines typical usage scenarios of learner models and Section 4 discusses related work. Section 5 presents requirements that are mandatory, desirable, or otherwise relevant for learner models before Section 6 contrasts this with legal considerations based on European laws and regulations. Section 7 summarises the paper and provides an outlook on future research.

## II. DEFINITION OF TERMS

This paper views a learner model solely as a digitally processed representation of learners' characteristics [3]. Inference mechanisms and reasoning may refer to a learner model, but are separate components rather than part of the model. Furthermore, learner models need to be distinguished from learning analytics: while the latter involves the analysis of group behaviour to predict individual outcomes, a learner model serves as a basis for tailoring the learning process to individual learners. Both approaches examine behavioural data and derive new insights which, once interpreted, enable the implementation of new measures. Open learner models constitute a notable development since they make model contents accessible to the learner or other parties involved in the learning process, such as teachers or parents. Open learner models are visual representations of machine-readable formats of components of learner models [5]. By offering tools for self-reflection in different formats, learners should be supported in various ways [6]. Learning Analytics Dashboards (*LAD*) share a similar objective [7]. Unlike learner models, LADs rely on a static representation of behavioural metrics derived from interaction data, while learner models focus on modelling knowledge and other individual characteristics of the learner [7][8]. Learner models and learning analytics both require personalised data from learners, which is subject to the same legal principles. Although goals and ways of working differ, the requirements and legal aspects can be pretty similar. Therefore, we also consider learning analytics and examine if specific aspects also apply to learner models, possibly with adaptations. We summarise both approaches and refer to them as learning analyses.

Learning environments are (digital) platforms where learners may use different content via various learning elements to educate themselves independently. Learning environments are not defined further here, as they are sufficient for the context at the meta level. It does not matter whether the learning environment is, e.g., a mobile application, an institutional continuing education platform, or even innovative learning settings with XR.

## III. USAGE SCENARIOS

Learner models and adaptation do matter at different levels: from support in planning a degree program [macro level] (which modules are interesting for me?), to learning patterns of student cohorts during their studies [meso level] (which topics prove difficult within certain student cohorts?), to which concepts within a module [micro level]. We will focus mainly on the micro level.

Individual adaption of the learning process is only viable if information and data about the respective learner can be used to respond to the needs of the respective learners and provide them with individualised support. The concept just mentioned will not scale to, e.g., entire study cohorts of degree programmes without further aid. The abundance of learner data, its pre-processing and intelligent aggregation make manual evaluation by the instructor for each individual learner almost impossible, which is why computer-aided methods are usually used. Digital recommendation systems automate the provision of individual support to users for making better decisions [9], often by building upon learner models. Figure 1 visualises the reference workflow of the learning process within a digital learning platform. The starting point is the completion of learning and teaching activities within the digital learning environment. During the learning process, many different types of behavioural data are collected about learners and their activities. This data is pre-processed and then made persistent to enable reproducible analyses subsequently. In addition to behavioural data, other structured knowledge about a learner (such as previous education, learning type, etc.) is also stored and analysed, resulting in individual recommendations for the next learning activities and feedback on previously visited learning elements.

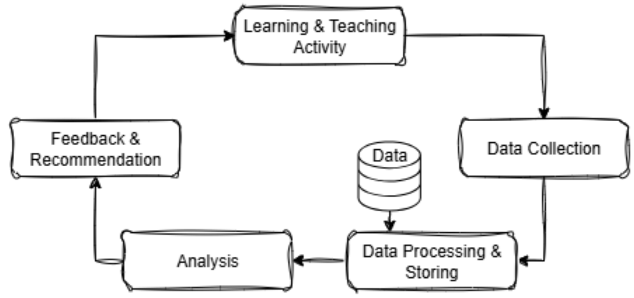


Figure 1. Workflow Reference Architecture (*based on* [10])

The basis for adaptive teaching and learning settings is the learner model, which stores all relevant data persistently and makes it available as required. Yet, the learner model should not be considered in isolation, but in the overall context of the digital learning environment and the associated relationships, as the added value of learner models only comes into play when they are used in the adaptive learning platform. The following section presents examples of typical scenarios from digital teaching to illustrate the specific application purpose. The necessary measures and requirements can then be derived subsequently. These three actors also form the three levels of adaptive learning described above. For the sake of completeness and better understanding, all three are briefly outlined here, although the focus remains on the micro level and thus on the learner.

### *A. Purpose of Use*

Several publications, e.g., ISO/IEC JTC1/SC36 [10] or [11][12], already present typical scenarios of digital learning at different levels. Teaching and learning activities in digital learning environments are the starting point for digital learning analyses. Learning analyses are used to personalise and thus improve the learning environment adaptively. Based on the results of the learning analysis, the learning environment can recommend individualised learning paths in combination with distinct learning content. Three main actors are involved in typical learning scenarios: the learner, the instructor, and the educational institution.

**Learner** (*micro level*)**.** In the past, log entries from the learning environment were difficult to understand for non-technical users, if they could be viewed at all. Nowadays, learners can obtain visual displays of their data in dashboards, e.g., to monitor their learning progress in relation to the average performance of the entire cohort. The early recognition of learners' personal needs and preferences (predictive analytics), including possible performance deficits, and the resulting initiation of preventive remedial measures (timely interventions) increase the effectiveness of the learning process. The use of such learning analyses also contributes to supporting students with disabilities and to identifying accessibility deficits in learning opportunities.

**Instructor** (*meso level*)**.** Instructors can track activities of their entire cohorts and detect progress and potential problem patterns, e.g., lack of learner engagement, early in the course. Consequently, instructors may initiate appropriate support through adaptive teacher response to observed learner needs and behaviour.

**Educational Institution** (*macro level*)**.** The educational institution can contribute to an improved holistic individual education strategy through the possibility of analyses, which can also reduce drop-out rates in the long term. In addition, administration may benefit from information on the entire student population, e.g., for future course planning. If comparisons of the data set across learning modules are permitted, similar patterns can be identified, and indications can be derived, such as a potential accessibility problem.

### B. Legal Principles and Classification

The implementation of learner models in educational institutions has various legal implications. On the one hand, for example, intellectual property is protected from the developer's perspective. On the other hand, learners' rights, especially their fundamental rights, must be considered. In the EU, the fundamental right to data protection, as stated in Article 8 of the EU Charter of Fundamental Rights (*CFR*) [13], is affected. The specific legal implementation of this fundamental right, the General Data Protection Regulation (*GDPR*) [14], is the primary law to be considered when using learner models. These provisions are considered when determining legal requirements.

## IV. Related Work

Learning offers must be modularised so that individual learning paths can be created based on a wide range of criteria. Many publications, particularly in the last decade [4], deal with adaptive teaching and learning settings. For this work, scientific publications on user models, particularly learner models, are the most relevant. Koch [15][16] summarised the seven objectives of a user model: (1) supporting users in learning a specific topic; (2) providing users with customised information; (3) adapting the user interface to the user; (4) supporting users in searching for information; (5) providing users with feedback on their knowledge; (6) supporting collaborative work; (7) supporting the system's use.

Numerous publications deal with learner analyses, although mostly only as a means to the end of adaptive learning systems [17][18] or very superficially, without going into relevant details. Several publications also only consider individual information rather than their aggregation into a learner model [19][20]. A couple of surveys attempt to create an overview of the field [21]-[24], but mostly only compare the different models with each other, without providing direct insight into which learner characteristics can be used, how they are modelled, or where required data comes from. Therefore, we conducted two parallel systematic literature searches on learner models, on the one hand, to get a comprehensive picture of the current state of science research [4] and, on the other hand, to clarify the state of practice based on standards and norms [3]. Ideally, this should pave the way to a standardised approach to how learner models are designed and created, as well as a set of standardised subcomponents – regardless of their purpose and use. 868 standards were reviewed, 16 of which were classified as relevant to the structure and components of learner models. Three standards deal intensively with learner models. Among those, ISO/IEC 29140:2021 [10] describes a mobile learner model that considers specific attributes, such as the device used, connectivity and the learner's location to describe the learning environment better. The 1EdTech consortium [25] focuses on the interoperability of internet-based information systems that support learners in their interaction with other systems. It uses a data model that captures the essential characteristics of learners to monitor and manage their progress, goals, performance, and learning experiences. IEEE P1484.2 [26] attempted to specify the syntax and semantics of a learner model by centralising public and private learner information, which became known as PAPI Learner. This endeavour started in the 1990s, but was not pursued further. Extensive research into norms and standards in the field of learner models has shown that approaches in this area are rare and either specialised [10], complex and more than just learner models [25], or have been abandoned [26]. No standards have been found that describe a generalised realisable learner model that can be extended according to specific use cases or that deal with the creation of such learner models [3].

In parallel, scientific publications of relevant publishers (*IEEE*, *ACM*, *Elsevier*, *pedocs*) from 2014 to 2023 were examined systematically, leading to 197 papers, which were relevant enough to be analysed in detail [4]. Scientific publications on the design and development of learner models reveal a variety of different approaches. Many models integrate characteristics such as learning style, but do not specify why or how these characteristics influence learning and whether or why it makes sense to take them into account. There is also a lack of clear recommendations as to which combinations of characteristics should be included in a learner model. Standardised characteristics, such as demographic data (e.g., name, gender, age), are often used together with behavioural and learning data. The modelling of this data varies greatly in the literature. Knowledge characteristics can be modelled using, e.g., overlay models or fuzzy logic. This diversity makes it hard for developers to make decisions and implement the models. In addition, publications often lack details on data sources, data processing, and practical implementation of model components. Many of these topics are only touched upon in passing or omitted altogether, which makes it difficult to replicate learner models precisely [4].

The two systematic literature searches indicate that the descriptions of learner models are often rather superficial and lack detail. Many explanations refer to frequently used standards and then expand these to include individual aspects [27]-[30][20], but do not describe in detail how these are expanded or what the implementation or modelling of the learner model looks like. Legal implications of learner models, specifically, have not yet been discussed. However, there is research regarding the requirements of the European data protection law in learning analytics, which can also be used for learner models [31]-[36]. So far, proposals are available for the standardisation of leaner models, such as educational metadata or course data. Our main objective is to contribute to this standardisation process with an original proposal for a reference model as one of the first steps towards a reference architecture. This contribution is based on the definition of open software interfaces for each subsystem in the architecture, avoiding any dependency on specific information models. We have already discussed elsewhere a possible reference architecture for an adaptive learning platform which integrates a learner model is [37].

## V. Requirements

Based on the previously analysed publications and the resulting findings on possible components of learner models and their use in digital learning environments, we derive a list of requirements for a learner model and its application. The development of the requirements specification is based on a qualitative analysis. In parallel publications, we already dealt with the functional requirements of learner models in greater detail [37] and also focused on the overall context of digital learning environments and the integration of learner models [38]. Based on literature research, the following section will focus on non-functional requirements and the legal basis for the use of learner models in Europe. The following section is by no means comprehensive, but rather collects and summarises the most relevant requirements in terms of their occurrence and description in the literature and links them to current legal restrictions in Europe.

In addition to the analysed publications, the feedback and experiences of students at our university play an essential role. It is important to emphasise again that learner models are structured models without any interference mechanisms [39]. However, it is also necessary to consider and evaluate the effects, i.e., the integration of a learner model into the overall eLearning environment, when considering the requirements for a learner model. The following requirements are classified according to the Kano model [40] into so-called *must-be requirements*, one-dimensional requirements (*should characteristics*) and attractive requirements (*can characteristics*).

### A. Must-be Requirements

Must-be requirements are essential for the successful use of a learner model in future scenarios and form the basis for its design and subsequent development.

**Collection and Management of Learning Data.** The collection and aggregation of static and dynamic information about the learner (e.g., knowledge, skills, interests, preferences, behavioural data) is the initial step. The utilisation of data from various sources plays a significant role. In the second step, the collected data must be digitally processed, modelled, and persisted.

**Legal Data Protection Requirements.** The development and use of learner models raise a range of legal considerations. In particular, rights relating to data ownership under the EU Data Act may become relevant. Moreover, if learner models fall under the definition of artificial intelligence systems according to the AI Act, additional restrictions may apply. This is especially the case when such systems are classified as high-risk (Annex III No. 3(b)/(c)), which would trigger extensive compliance obligations. However, the following section focuses on the most immediately relevant legal framework: The General Data Protection Regulation (GDPR). All data to be collected must at all times be subject to the applicable legal requirements of the *GDPR* [14][41]. This requirement is closely aligned with the legal constraints of designing and creating learner models. What are the applicable legal requirements for data collection, processing and storage? These questions can neither be generalised, nor answered in general terms, but must be legally considered and examined individually, depending on the purpose of use and the data to be collected. The concept of a data trust model might be applicable here: A neutral, trustworthy entity ensures that data is processed and used by the defined data protection guidelines and only accessible by authorised parties.

### B. Should-be Requirements

Should-be requirements are important prerequisites for making optimal use of the learner model in future scenarios. They represent desirable, yet not mandatory, features that increase the efficiency and flexibility of learner models. These requirements serve as an orientation for further development and improvement of the model.

**Transparency / Traceability.** The transparency of how the system has reached a result [42] creates the trust and acceptance of certain subordinate recommendations, which form the basis of good support [9]. Therefore, it is desirable for learner models to be transparent in every single step of the process if possible, i.e., from the origin of the data to the individual processing steps the data went through and what consequences this has for the result and its explanation. This is important so that backgrounds and issues, such as discrimination potential [43]-[45] in categorisations are illuminated. Transparency thereby draws on individual decisions so that they can be correctly traced. Traceability is achieved by involving the learner in the process right from the beginning and also by taking a learner-centred approach to the design and development of learner models.

**Responsibility.** Traceability is closely connected to responsibility. The data that may be collected must be left to the users' choice, considering applicable legal standards. That is, students can decide which data may be logged and persisted. This topic also includes compliance with ethical principles. For the digital domain, this means that if the learning management system displays ethical behaviour patterns, learners expect the system’s compliance to any ethical guidelines and principles. Briefly summarised, ethics is the view of moral values and their conception (based on Aristotle's *Nicomachean Ethics* and Immanuel Kant's *Categorical Imperative*).

**Fairness & Ethics.** Fairness affects many different steps within the learning process. For example, algorithms for decision-making and data processing must be checked for possible biases to ensure equal opportunities for all learners [46][47].

**Tamper-proof.** The data managed and persisted by learner models must be protected from unauthorised intervention and changes. Any changes need to be logged in a traceable manner. This implies two interlinked requirements, namely that unauthorised data access must be prohibited, and, if access is permitted, changed data shall be checked for plausibility and changes shall be logged for tracking purposes.

### *C. Can Requirements*

Could-be requirements represent possible extensions that might optimise the learner model in certain scenarios through, e.g., additional benefits or improved user experience. If necessary, these requirements can be included in future development phases to increase the flexibility and adaptability of the model.

**Openness & Visualisation.** Learner models should be open [6][48] to promote metacognitive behaviours, such as self-awareness and self-regulation. This means that students may display and analyse their own instance of a learner model for a better understanding of their learning progress and, e.g., misconceptions [5]. In this way, learners can see their current learning status and progress in any area at any time, compare it with their learning goals, and derive follow-up activities. Various visualisation options [49]-[51] are essential for presenting complex learning data in a clear and concise format to the learner. Different representations of the same data can bring learners closer to the various aspects and clarify them [52]. Leaner models designed in this way are called open learner models [53][54].

**Negotiation Options.** In addition to the pure visualisation of personal data, some learner models also offer the option of interacting with this data and, for example, negotiating with the model if the data shows inadequacies from the learner's perspective [5][55]. In this way, (open) learner models give learners not only responsibility for their individual data and progress but also human control over their personal data.

**Data Minimisation & Sustainability.** Only data that is absolutely necessary for adaptation should be collected. Minimising the amount of collected data while maximising the value of the information avoids unnecessary strain on the infrastructure.

**Modularity & Flexibility.** A largely self-contained modular structure of the component of earner models enables flexibility, for example, to swap the technical infrastructure (learning management system) or to export the learner model and integrate and use it in other environments, for example, if the learner changes university after graduation.

**Maintainability & Expandability.** The learner model should have a modular structure (see previous requirement) to ease future adjustments or extensions. This offers the learner model a certain degree of secured prospects.

**Standardised / Interoperability & Integration.** Ideally, the description of learner information is standardised so that it can be easily exported and exchanged between different learning platforms (according to the interoperability defined in [56][57]) through standardised interfaces for data exchange and integration. Standard conformity is essential here, i.e., open standards to ensure compatibility with other systems should be supported. Standardisation also enables cross-platform integration, i.e., the model can be seamlessly integrated into existing learning management systems. As with data minimisation (Art. 5 lit. c GDPR), data portability is also a legal requirement (Art. 20 GDPR).

## VI. LEGAL CONSTRAINTS

Legal considerations are crucial in the design and development of learner models, as lawful use is only possible if these regulations are adhered to. In our case, we are specifically concerned with the legal requirements in Europe regarding the collection, use, and evaluation of personal data. Personal data plays a crucial role in enabling adaptive customisation of learning processes. Therefore, all data collected must always comply with the relevant legal requirements (data protection). However, data protection also encompasses other aspects. On the one hand, only data that is strictly necessary for meaningful adaptation should be collected (data minimisation & data economy). On the other hand, data should only be collected if there is a legal basis for it (lawfulness). All the aforementioned legal principles are enshrined in the General Data Protection Regulation (*GDPR* [41]). The collection and aggregation of static and dynamic information about the learner (e.g., knowledge, skills, interests, preferences) must be compliant with the GDPR. The utilisation of data from various sources plays a significant role. In the second step, this collected data must be digitally processed, modelled and persisted. Unfortunately, the GDPR is technology-neutral (recital 15 GDPR), requiring the broad terms it employs to be defined explicitly in the context of learner models. The GDPR's requirements are diverse, and due to the limited availability of case law and literature regarding the GDPR in relation to learner models, these legal obligations cannot yet be determined with a high degree of certainty. It is highly recommended that the local data protection officer be integrated as soon as possible into the process of implementing learner models in educational institutions. Many obligations of the GDPR, for example, the data protection impact assessment in Art. 35 GDPR will be very hard to meet without professional support. This chapter aims to identify potential issues arising from the GDPR and highlight key aspects to consider when designing and implementing a learner model. In providing an overview, we focus on the principles of the GDPR, which are concretised in the GDPR, and the arising issues regarding learner models.

**Lawfulness - Legal Bases.** Every processing of personal data needs a legal base, as stated in Art. 5 para. 1, Art. 6 para. 1 GDPR. Art. 6 para. 1 lit. a GDPR establishes consent as a legal basis, which must be given voluntarily [58, p. 330]. In hierarchical contexts, such as teaching environments, the GDPR requires a strict interpretation. Learners often depend on lecturers for grading, which may pressure them to agree to the processing of their personal data to align with lecturers' expectations. This exemption is not ubiquitous, though. If the use of a learner model is a voluntary additional offer of the educational institution and is not linked to a specific course, voluntary consent seems possible. For state educational institutions, instead, Art. 6 para. 1 lit. e GDPR serves as the legal basis for data processing when processing is necessary for the performance of a task carried out in the



 ISBN: 978-1-68558-303-3 

public interest, such as education. However, under Art. 6 para. 3 GDPR, the legal basis of Art. 6 para. 1 lit. e GDPR must be complemented by specific legal provisions by the member states that establish obligations and define tasks. Consequently, instructors must ensure that the applicable legal bases in their national laws include the processing of personal data for educational purposes. General requirements that the GDPR imposes on national law are discussed in [59].

Furthermore, the type and extent of data being processed should be carefully evaluated. On the one hand, this enables an assessment of the risks associated with potential data breaches. On the other hand, it highlights whether special categories of personal data are being processed in the learner model. Processing personal data revealing racial or ethnic origin, political opinions, religious or philosophical beliefs, trade union membership, genetic data, biometric data for uniquely identifying individuals, health data, or information about a person's sex life or sexual orientation is subject to stringent requirements. Under Art. 9 para. 2 lit. g GDPR, alongside the requirements of Art. 6 para. 1 lit. e GDPR, a legal basis in Union or Member State law is required, allowing such processing only if it is necessary for reasons of substantial public interest. Meeting this requirement is particularly challenging for learner models.

**Purpose limitation.** Developers of learner models must carefully evaluate the potential sources of personal data used for their development and operation. Often, the data is collected for a different purpose — for example, when student exams, initially collected for grading purposes, are fed into learner models. In addition, the subsequent use of data generated by the learner model should undergo careful evaluation. Art. 5 para. 1 lit. b GDPR requires that personal data be collected for specific, explicit, and legitimate purposes and not further processed in ways incompatible with those purposes. However, Art. 6 para. 4 GDPR provides an exception: if the new purpose for processing personal data is not based on the data subject's consent or authorised by Union or Member State law, its compatibility with the original purpose must be assessed using the criteria outlined in Art. 6 para. 4 lit. a – e GDPR. Scientific research, which is what learner models could be part of, is privileged. Art. 5 para. 1 lit. b assumes scientific research is "not be considered to be incompatible with the initial purposes".

**Transparency.** Transparency is not merely a technical requirement but is also enshrined in Art. 5 para. 1 lit. a GDPR. Personal data shall be processed "in a transparent manner in relation to the data subject […]." This principle is guaranteed by primary law in Art. 8 para. 2 S. 2 CFR by granting every person "the right of access to data which has been collected concerning him or her […]." The requirement of transparency is primarily detailed in Art. 12 ff. GDPR. It stipulates that if personal data is collected, the controller must, pursuant to Art. 13 GDPR, inform the data subject about the details outlined in Art. 13 paras. 1 and 2 GDPR at the time of collection. This information must be provided in a privacy statement. It is recommended to explain the system in clear and accessible language in the privacy statement to help learners understand how the learner model functions.

**Data Minimisation.** Within learner models, various sensitive data will be processed, which is in contrast to the GDPR principle of data minimisation, which allows the processing of personal data only when strictly necessary. The principle does not require minimising the data itself but seeks to limit the connection of data to natural persons, thereby reducing infringements of fundamental rights [60] mn. 96. This takes several technical and organisational measures. First, personal data should be pseudonymised. According to Art. 4 No. 5 GDPR, pseudonymisation is the processing of personal data in such a manner that the personal data can no longer be attributed to a specific data subject without the use of additional information. EDSA provides a detailed guideline for pseudonymisation [61]. As an organisational matter, a role concept should be implemented, which specifies who can access the data collected in the learner model and who can edit this data. For complete guidance on technical and organisational measures according to Art. 25 see [62]. Especially, the scope of people who can undo the pseudonymisation needs to be kept small. Also, the use of a data trustee can be discussed [38]. A data trustee is a neutral third party that acts as a steward for sensitive data, ensuring its secure handling, responsible use, and protection of individuals' privacy while facilitating data-driven innovation.

**Storage Limitation.** The principle of storage limitation, as outlined in Art. 5 para. 1 lit. e GDPR requires that personal data be retained in a form permitting the identification of data subjects only for as long as necessary to achieve the purposes for which it is processed. Temporal storage limitation is a subset of the overarching principle of necessity [60] mn. 122. Educational institutions and lecturers must determine the appropriate retention period for the data. Developers of learning models must ensure that the complete deletion of students' personal data is technically feasible. Typically, learners' personal data should be deleted once they leave the educational institution. If the data is still supposed to be used for the training of learner models, the link between the data and the individual learners can be erased entirely and irreversibly (anonymisation) [63].

**Integrity & Confidentiality.** Art. 5 para. 1 lit. f GDPR states that personal data must be "processed in a manner that ensures appropriate security of the personal data, including protection against unauthorised or unlawful processing and against accidental loss, destruction, or damage, using appropriate technical or organisational measures." This principle underscores the importance of systemic data protection. It is primarily implemented through Art. 25 GDPR, "Data protection by design," and Art. 32 GDPR, "Security of processing". Whenever possible, data should be stored in an encrypted format on a trustworthy server, ideally on a server operated by the educational institution. The use of servers in an EU member state is unobjectionable, as the GDPR establishes a uniform standard for data protection across member states. The use of servers outside the EU is possible; however, the transfer of data to such servers must comply with Art. 45 et seq. GDPR, which aims to ensure continuity of the level of data protection [64] mn. 6.

**Use of Processors.** If an external service provider is engaged in processing personal data on behalf of an institution, the requirements of Art. 28 GDPR must be fulfilled. Processors are required to provide sufficient guarantees that they will implement appropriate technical and organisational measures to ensure compliance with the GDPR. These requirements must be formalised in a contract between the educational institution and the processor. Whenever possible, the processor should store data on servers located within the EU.

**No automatic Decision-Making with a Legal Impact.** Once the learning model is capable of analysing learners' input, it is likely to be used for grading purposes. However, Art. 22 para. 1 GDPR states that decisions based solely on automated processing, including profiling, which produces legal effects, are prohibited. Art. 22 para. 2 GDPR provides exemptions. Still, these are unlikely to apply to the use of learner models unless a member state establishes a legal basis explicitly permitting automated decision-making and defines suitable measures to safeguard the data subject's rights, freedoms, and legitimate interests.

## VII. CONCLUSION AND FUTURE WORK

Learning models have great potential to make education more individualised, efficient and equitable, but they are still a long way from being implemented in a standardised and legally compliant manner. The key challenge is to reconcile technical innovations with strict legal requirements, particularly those of the GDPR.

The GDPR sets out only very general and technology-neutral requirements, offering few concrete implementation guidelines for learner models. Due to the lack of case law and specific regulatory guidance, legal obligations for learner models remain vague and difficult to apply in practice. This underscores the need for stronger support from data protection authorities in clarifying how GDPR principles can be operationalised in educational technology contexts.

However, as a first step, any processing of personal data must be based on a valid legal basis — typically either informed consent or a legal authorisation, such as the public task basis for educational institutions. Once the legal basis is clarified, data minimisation and storage limitation should become a primary focus. However, this does not mean that the amount of data per se must be reduced, but rather that the identifiability of individuals must be minimised. Accordingly, techniques such as pseudonymisation and, where feasible, anonymisation should be considered to reduce the risk of privacy breaches.

In addition to the legal regulations, technical and non-legal requirements must also be considered in order to design learner models that are practical and effective. The variety of possible requirements analysed is vast. It has not yet been aggregated in a form that makes it easier for designers and developers of learner models to get started. Open learner models offer users transparency about their data, thereby promoting self-reflection and independent learning. Visualisations of complex data help users understand their learning status and continue working in a targeted manner.

Modular and interoperable structures ensure that learner models can be flexibly integrated and transferred between different learning environments. Maintainability and extensibility allow for long-term adaptation to new learning contexts. Overall, these technical requirements illustrate that a well-designed reference model forms the basis for the practical, transparent and fair design of adaptive learning environments.

Another key priority is transparency and traceability in all processing steps: learners must be able to understand how their personal data is processed, who has access to it, and how the system derives its outputs or recommendations. Only then can trust in adaptive learning technologies be established. In addition, fairness and the avoidance of algorithmic bias ensure that all learners have equal opportunities. These questions must be addressed in the next steps, including the design, implementation, and evaluation of a prototype learner model that complies with the outlined legal requirements.

The next technical steps are to conceptualise the legal constraints described in this publication with the help of the local data protection authority and the functional requirements [37] into a interoperable and legally compliant learning model, then to implement this model and integrate it into the learning platform [38]. This initial prototype must be evaluated and further developed in compliance with the legal conditions so that the needs of learners are met and the learning process is improved in a sustainable and prosperous manner. Even though the legal requirements within the EU on this topic are not easy to understand, further steps could be taken to develop the prototype into a reference model for learner models, which would enable other educational institutions to get started with adaptive learning environments more quickly, as learner models have been proven to make a decisive contribution to accommodate the heterogeneity of learners and support their learning processes individually and sustainably.

## ACKNOWLEDGMENT

This work is part of the VoLL-KI project and Komp-HI project and funded by the German Ministry of Education and Research (*Bundesministerium für Bildung und Forschung*) under grants 16DHBKI090 and 16DHBKI073.

## REFERENCES

[1] S. Zhang and S. Wang, ‘Modeling Learner Heterogeneity: A Mixture Learning Model With Responses and Response Times’, *Front. Psychol.*, vol. 9, p. 2339, 2018.

[2] T. Kärner, J. Warwas, and S. Schumann, ‘A Learning Analytics Approach to Address Heterogeneity in the Classroom: The Teachers’ Diagnostic Support System’, *Tech Know Learn*, vol. 26, no. 1, pp. 31–52, 2021.

[3] F. Böck, D. Landes, and Y. Sedelmaier, ‘Learner Models: A Systematic Literature Research in Norms and Standards’, in *Proceedings of the 16th International Conference on Computer Supported Education*, Angers, France: SCITEPRESS - Science and Technology Publications, 2024, pp. 187–196.

[4] F. Böck, M. Ochs, A. Henrich, D. Landes, J. Leidner, and Y. Sedelmaier, ‘Learner Models: Design, Components, Structure, and Modeling - A systematic Literature Review’, *User Modeling and User-Adapted Interaction - The Journal of Personalization Research*, Springer, 2025.

[5] S. Bull and J. Kay, ‘Open Learner Models’, in *Advances in Intelligent Tutoring Systems*, R. Nkambou, J. Bourdeau, and R. Mizoguchi, Eds., in Studies in Computational Intelligence. Springer, 2010, pp. 301–322.

[6] S. Bull and J. Kay, ‘Student Models that Invite the Learner In: The SMILI:() Open Learner Modelling Framework’, *International Journal of Artificial Intelligence in Education*, vol. 17, no. 2, pp. 89–120, 2007.

[7] R. Bodily *et al.*, ‘Open learner models and learning analytics dashboards: a systematic review’, in *Proceedings of the 8th International Conference on Learning Analytics and Knowledge*, Sydney New South Wales Australia: ACM, 2018, pp. 41–50.

[8] J. Kay and S. Bull, ‘New Opportunities with Open Learner Models and Visual Learning Analytics’, in *Artificial Intelligence in Education*, vol. 9112, C. Conati, N. Heffernan, A. Mitrovic, and M. F. Verdejo, Eds., in Lecture Notes in Computer Science, vol. 9112. Springer, 2015, pp. 666–669.

[9] F. Ricci, L. Rokach, and B. Shapira, Eds., *Recommender systems handbook*, Third edition. Springer, 2022.

[10] *[ISO/IEC TR 20748-1:2016] Information technology for learning, education and training — Learning analytics interoperability — Part 1: Reference model*, ISO/IEC TR 20748-1:2016, 2016.

[11] G. Siemens *et al.*, ‘Open Learning Analytics: an integrated & modularized platform’, SOLAR, Jul. 2011. [Online]. Available: www.solaresearch.org/publications/position-papers/

[12] M. Cooper, R. Ferguson, and A. Wolff, ‘What can analytics contribute to accessibility in e-learning systems and to disabled students’ learning?’, in *Proceedings of the Sixth International Conference on Learning Analytics & Knowledge - LAK ’16*, Edinburgh, UK: ACM, 2016, pp. 99–103.

[13] *Charter of Fundamental Rights of the European Union*, vol. C 326/391. 2016, p. 17. [Online]. Available: http://data.europa.eu/eli/treaty/char_2012/oj

[14] *Regulation (EU) 2016/679 of the European Parliament and of the Council*, vol. 02016R0679-20160504. 2016, p. 78. [Online]. Available: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A02016R0679-20160504

[15] N. P. de Koch, ‘Software Engineering for Adaptive Hypermedia Systems’, Ludwig-Maximilians-Universität München, München, 2000.

[16] O. Abdennour, H. Kemouss, M. Erradi, and M. Khaldi, ‘From Learner Profile Management to Learner Model Modeling’, in *Advances in Educational Technologies and Instructional Design*, M. Khaldi, Ed., IGI Global, 2023, pp. 71–93.

[17] L. Lin and F. Wang, ‘Adaptive Learning System Based on Knowledge Graph’, in *International Conference on Education and Training Technologies*, in ICETT ’23. New York, NY, USA: Association for Computing Machinery, 2023.

[18] L. Yang, Y. Yu, and Y. Wei, ‘Data-Driven Artificial Intelligence Recommendation Mechanism in Online Learning Resources’, *International Journal of Crowd Science*, vol. 6, no. 3, pp. 150–157, 2022.

[19] X. Li and S. He, ‘Research and Analysis of Student Portrait Based on Campus Big Data’, in *International Conference on Big Data Analytics (ICBDA)*, 2021, pp. 23–27.

[20] Y. Jun-min, X. Song, L. Da-Xiong, W. Zhi-Feng, H. Peng-Wei, and X. Chen, ‘Research on the Construction and Application of Individual Learner Model’, in *International Conference on Computing and Information Technology*, Procedia Computer Science, 2018, pp. 88–92.

[21] D. Hooshyar, M. Pedaste, K. Saks, Ä. Leijen, E. Bardone, and M. Wang, ‘Open learner models in supporting self-regulated learning in higher education: A systematic literature review’, *Computers & Education*, vol. 154, p. 103878, 2020.

[22] S. Ouf, M. A. Ellatif, S. E. Salama, and Y. Helmy, ‘A proposed paradigm for smart learning environment based on semantic web’, *Computers in Human Behavior*, vol. 72, pp. 796–818, 2017.

[23] X. Wang, Y. Maeda, and H.-H. Chang, ‘Development and techniques in learner model in adaptive e-learning system: A systematic review’, *Computers & Education*, vol. 225, p. 105184, 2025.

[24] A. Zanellati, D. D. Mitri, M. Gabbrielli, and O. Levrini, ‘Hybrid Models for Knowledge Tracing: A Systematic Literature Review’, *IEEE Trans. Learning Technol.*, vol. 17, pp. 1021–1036, 2024.

[25] *1EdTech Learner Information Package Specification*, Standard, Mar. 09, 2001. [Online]. Available: https://site.imsglobal.org/standards/sldm

[26] *[IEEE P1484.2] - Standard for Information Technology — Learning Systems — Learner Model*, PAR IEEE P1484.2, 1997.

[27] W. QIU, J. DU, and F. LI, ‘Knowledge Description Frame of Learning Resources for Recommendation System: From the Perspectives of Learning Psychology’, in *2020 15th International Conference on Computer Science & Education (ICCSE)*, 2020, pp. 438–440.

[28] X. Zhao, ‘Research on learner model of adaptive learning system’, in *International Conference on Computer Science & Education (ICCSE)*, 2021, pp. 906–909.

[29] S. Wang, L. Yuan, and H. Yang, ‘User Model Construction of Chinese Learners Based on Learning Style’, in *International Conference on Computer Science and Educational Informatization (CSEI)*, 2021, pp. 70–75.

[30] T. Fei Zhou, Y. Qing Pan, and L. R. Huang, ‘Research on Personalized E-Learning Based on Decision Tree and RETE Algorithm’, in *International Conference on Computer Systems, Electronics and Control (ICCSEC)*, 2017, pp. 1392–1396.

[31] A. N. Cormack, ‘A Data Protection Framework for Learning Analytics’, *Learning Analytics*, vol. 3, no. 1, 2016.

[32] Q. Liu and M. Khalil, ‘Understanding privacy and data protection issues in learning analytics using a systematic review’, *Brit J Educational Tech*, vol. 54, no. 6, pp. 1715–1747, 2023.

[33] T. Hoel and W. Chen, ‘Privacy and data protection in learning analytics should be motivated by an educational maxim—towards a proposal’, *Research and Practice in Technology Enhanced Learning*, vol. 13, no. 1, p. 20, 2018.

[34] M. Paludi, ‘The Right to Privacy and Data Protection for High School Students in the Context of Digital Learning Models and Learning Analytics’, in *Proceedings of the Doctoral Consortium of the Learning Analytics Summer Institute Europe 2024 (LASI Europe 2024 DC)*, 2018, pp. 1–10.

[35] H. Drachsler and W. Greller, ‘Privacy and analytics: it’s a DELICATE issue a checklist for trusted learning analytics’, in *Proceedings of the Sixth International Conference on Learning Analytics & Knowledge - LAK ’16*, Edinburgh, UK: ACM, 2016, pp. 89–98.

[36] J. Doveren, B. Heinemann, and U. Schroeder, ‘Towards Guidelines for Data Protection and Privacy in Learning Analytics Implementation’, in *Online-Labs in Education*, Nomos Verlagsgesellschaft mbH & Co. KG, 2022, pp. 45–52.

[37] F. Böck, ‘Learner Models: Requirements analysis for Application in adaptive Learning Environments and

Recommendation Systems', in *2025 IEEE Digital Education and MOOCS Conference (DEMOcon)*, 2025.

[38] F. Böck, A. Deuerling, and D. Landes, 'Adaptive Learning Environment Reference Architecture for an Optimised Learning Process', in *2025 IEEE Global Engineering Education Conference (EDUCON)*, London, UK: IEEE, Apr. 2025, pp. 1–10.

[39] F. Böck, 'A Research Agenda for Learner Models for Adaptive Educational Digital Learning Environments', in *Futureproofing Engineering Education for Global Responsibility*, Springer Nature, 2025.

[40] N. Kano, N. Seraku, F. Takahashhi, and S. ichi Tsuji, 'Attractive Quality and Must-Be Quality', *Journal of the Japanese Society for Quality Control*, vol. 14, no. 2, pp. 147–156, 1984.

[41] Bundesministerium der Justiz, *Bundesdatenschutzgesetz*. 2017, p. 45. [Online]. Available: https://www.gesetze-im-internet.de/bdsg_2018/BDSG.pdf

[42] J. Kay, 'Learner know thyself: Student models to give learner control and responsibility', in *Proceedings of International Conference on Computers in Education*, 1997, pp. 17–24.

[43] V. Scholes, 'The ethics of using learning analytics to categorize students on risk', *Education Tech Research Dev*, vol. 64, no. 5, pp. 939–955, 2016.

[44] J. Dressel and H. Farid, 'The accuracy, fairness, and limits of predicting recidivism', *Sci. Adv.*, vol. 4, no. 1, p. eaao5580, 2018.

[45] N. Goltz and T. Dowdeswell, *Real world AI ethics for data scientists: practical case studies*. Boca Raton: CRC Press, Taylor & Francis Group, 2023.

[46] K. Kitto and S. Knight, 'Practical ethics for building learning analytics', *Br J Educ Technol*, vol. 50, no. 6, pp. 2855–2870, 2019.

[47] Y.-S. Tsai, C. Perrotta, and D. Gašević, 'Empowering learners with personalised learning approaches? Agency, equity and transparency in the context of learning analytics', *Assessment & Evaluation in Higher Education*, vol. 45, no. 4, pp. 554–567, 2020.

[48] S. Bull and J. Kay, 'SMILI☺: a Framework for Interfaces to Learning Data in Open Learner Models, Learning Analytics and Related Fields', *Int J Artif Intell Educ*, vol. 26, no. 1, pp. 293–331, 2016.

[49] S. Bull, B. Ginon, C. Boscolo, and M. Johnson, 'Introduction of Learning Visualisations and Metacognitive Support in a Persuadable Open Learner Model', in *International Conference on Learning Analytics and Knowledge*, in LAK '16. NY, USA: ACM, 2016, pp. 30–39.

[50] H. Ferreira, G. Oliveira, R. Araújo, F. Dorça, and R. Cattelan, 'An Open Model for Student Assessment Visualization', in *2019 IEEE 19th International Conference on Advanced Learning Technologies (ICALT)*, 2019, pp. 375–379.

[51] J. A. Ruipérez-Valiente, P. J. Muñoz-Merino, D. Leony, and C. D. Kloos, 'ALAS-KA: A learning analytics extension for better understanding the learning process in the Khan Academy platform', *Computers in Human Behavior*, vol. 47, pp. 139–148, 2015.

[52] C.-Y. Law, J. Grundy, R. Vasa, and A. Cain, 'An empirical study of user perceived usefulness and preference of open learner model visualisations', in *2016 IEEE Symposium on Visual Languages and Human-Centric Computing (VL/HCC)*, 2016, pp. 49–53.

[53] S. Bull and J. Kay, 'Open Learner Models', in *Advances in Intelligent Tutoring Systems*, vol. 308, R. Nkambou, J. Bourdeau, and R. Mizoguchi, Eds., in Studies in Computational Intelligence, vol. 308. Springer, 2010, pp. 301–322.

[54] J. Kay, K. Bartimote, K. Kitto, B. Kummerfeld, D. Liu, and P. Reimann, 'Enhancing learning by Open Learner Model (OLM) driven data design', *Computers and Education: Artificial Intelligence*, vol. 3, p. 100069, 2022.

[55] S. Bull, B. Ginon, C. Boscolo, and M. Johnson, 'Introduction of Learning Visualisations and Metacognitive Support in a Persuadable Open Learner Model', in *International Conference on Learning Analytics and Knowledge*, in LAK '16. NY, USA: ACM, 2016, pp. 30–39.

[56] *ISO/IEC 25010:2011 Systems and software engineering — Systems and software Quality Requirements and Evaluation (SQuaRE) — System and software quality models*, 35.080 Software 1, 2011.

[57] *ISO/IEC 25002:2024 Systems and software engineering — Systems and software Quality Requirements and Evaluation (SQuaRE) — Quality model overview and usage*, 35.080 Software 1, 2024.

[58] W. Kotschy, 'Art. 6 GDPR', in *The EU General Data Protection Regulation (GDPR): a commentary*, C. Kuner, C. Docksey, L. A. Bygrave, and L. Drechsler, Eds., Oxford, United Kingdom: Oxford University Press, 2019.

[59] J. T. Helmke and H. Link, 'Das Verhältnis von Art. 6 Abs. 2 und 3 DSGVO', *Datenschutz Datensich*, vol. 47, no. 11, pp. 708–714, 2023.

[60] A. Roßnagel and P. Richer, 'Art. 5', in *General data protection regulation: article-by-article commentary*, First edition., I. Spiecker Döhmann, E. Papakōnstantinu, G. Hornung, and P. de Hert, Eds., in Beck-online Bücher. , Baden-Baden, Germany: Nomos, 2023.

[61] The European Data Protection Board, 'Guidelines 01/2025 on Pseudonymisation'. Jan. 16, 2025. [Online]. Available: https://www.edpb.europa.eu/system/files/2025-01/edpb_guidelines_202501_pseudonymisation_en.pdf

[62] The European Data Protection Board, 'Guidelines 4/2019 on Article 25 Data Protection by Design and by Default, Version 2.0'. Oct. 20, 2020. [Online]. Available: https://www.edpb.europa.eu/sites/default/files/files/file1/edpb_guidelines_201904_dataprotection_by_design_and_by_default_v2.0_en.pdf

[63] Article 29 Data Protection Working Party, 'Opinion 4/2007 on the concept of personal data'. Jun. 20, 2007. [Online]. Available: https://ec.europa.eu/justice/article-29/documentation/opinion-recommendation/files/2007/wp136_en.pdf

[64] P. Schantz, 'Art. 44', in *General data protection regulation: article-by-article commentary*, First edition., I. Spiecker Döhmann, E. Papakōnstantinu, G. Hornung, and P. de Hert, Eds., in Beck-online Bücher. Baden-Baden, Germany: Nomos, 2023.