Learner Models:

The Gap between Educational Opportunities vs. Data Protection Challenges

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) provides a basis for digital learning environments that automatically adapt to learners' needs. To do so, these systems presuppose knowledge of the individual learner. Learner models are digital representations of learner characteristics that aim to enable personalised and adaptive learning experiences, addressing 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 designing interoperable, legally compliant learner models. Our findings lay the groundwork for developing a reference architecture that facilitates the scalable and ethical integration of learner models in digital learning environments.**

Keywords — learner model; learner modelling; requirements; legal issues; compliance; ethical principles; higher education; learning analytics; general data protection regulation; GDPR; lawfulness of processing; data minimisation; learning analytics regulation.

I. Introduction

Learning is an essential process through which people acquire and continuously develop their knowledge and skills. This process takes place in both formal and informal contexts and is therefore part of our daily lives. With rapidly advancing technology, interactive content enables us to promote learning through individualised learning paths, regardless of time or place. This opens up additional opportunities for flexible and self-directed learning.

This work builds on our preliminary published conference publication [1]. It expands on this by providing a comprehensive view of digital learning environments, not only from the learner's perspective (micro level) but also from the perspectives of lecturers (*meso level*) and educational institution administrators (*macro level*). In addition, illustrative examples make the requirements more tangible and concrete, facilitating a quick understanding of this highly complex topic. Recommended strategies for dealing with the topic are then derived from this, and the appendix contains detailed working aids to simplify the introduction to collecting and processing personal data in the context of digital learning environments.

Numerous studies have analysed the factors that contribute to successful learning. All of these studies have one thing in common: the so-called flow experience [2], [3]. The flow experience describes the channel between underchallenging and overchallenging the learner. To make the learning process as successful and sustainable as possible, care should be taken to ensure that learning occurs within this channel. This channel is individual to each learner and must be considered separately.

However, it is not only in terms of the flow experience that learners differ; as a group, they are also becoming increasingly heterogeneous, as they differ in terms of individual levels of knowledge and competencies, learning styles, individual preferences for (digital) media, and various other factors [4], [5]. A potential solution to accommodate this growing heterogeneity might be digital learning environments that support users in specific situations. In digital learning environments, there are two main approaches to adjusting preferences or learning content: personalisation and adaptation. With personalisation, learners can manually adjust the system to their preferences [6], while with adaptation, the system automatically adapts to individual learners based on recognised individual needs and changing conditions [7]. The latter is exciting from the perspective of digital adaptive learning environments. This automatic support is possible on various levels. In our previous work [1], we focused exclusively on learning environments that support learners at the micro level by adapting to their specific needs in self-directed learning. This paper expands the perspective to include the meso and macro levels, so that the requirements for data collection and its digital (automatic) processing are considered not only from a small-scale student perspective but also more generally. 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 relevant to individual learning support [8],

[9]. 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 [9] 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 [9] and standards on learner models [8]. In addition, we address legal issues related to the development and operation of digital learning environments that rely on learners' personal data. To do so, we take a European perspective. With this contribution, we not only want to help identify common requirements for learner models, which could be implemented in a subsequent step through a reference architecture that accounts for legal constraints, but also to lay the foundation for the development and use of digital adaptive learning environments for educational institutions.

In the remainder of this article, Section II first explains the terms used, Section III describes typical application scenarios for learning models from different perspectives (learner, lecturer, and educational institution), and Section IV discusses related work. Section V presents non-functional requirements that are mandatory, desirable or otherwise relevant for learning models and their use in digital adaptive learning environments, before Section VI compares these with legal considerations based on European laws and regulations. Section VII establishes the relationship between the technical and legal requirements and links them, on which Section VIII builds, setting out recommended guidelines for proceeding with this topic, with reference to the working aids in the appendix. Section IX summarises the article and provides an outlook for future research.

II. Definition of Terms

The literature on the science of digital learning presents different views on the same terms or their synonyms. We will therefore briefly review the most important terms to establish a common starting point. As mentioned in the introduction, there are various ways to adjust settings when learning in digital learning environments. Adaptation, often referred to as individualisation in the literature, represents the highest level of differentiation. An earlier stage is personalisation, which allows users to manually adjust the system to their preferences. In a learning context, adaptation refers to the provision of automatic, targeted support throughout the learning process, i.e., tailoring instruction to an individual's performance level and personal interests. For this automatic adaptation to succeed, the learning environment requires a digital representation of the learner, known as the learner model (*LM*).

This paper views a learner model solely as "an explicit representation of the individual characteristics and interaction data of a single learner in the form of a machine internal representation, and comprises many different aspects that ultimately depend on the purpose and goal of the application, [sort of] the digitally processed representation of the learner's characteristics" [10]. 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. The term' learning analytics' (LA) was coined in an academic context by Siemens and Long in their call for papers for the 1st International Conference on Learning Analytics and Knowledge (LAK11) as "the measurement, collection, analysis and reporting of data about learners and their contexts, for purposes of understanding and optimising learning and the environments in which it occurs" [11]. Learning Analytics focuses exclusively on the learning process (including the analysis of relationships among learners, content, institutions, and teachers) to clarify the added value that big data and analytics can bring to higher education. The intended level of analysis is primarily at the course level (discourse analysis, "intelligent curriculum") and departmental level (predictive modelling, patterns of success/failure), to report – e.g., via dashboards through aggregation and descriptive statistics of the collected data – and the subsequent opportunity for optimisation [12], [13].

While the latter involves analysing 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. In short, learning analytics analyses learning activities, whilst learner models digitally represent learners' current status.

Open learner models constitute a notable development, as they make model content accessible to learners and 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 [14]. By offering tools for self-reflection in different formats, learners should be supported in various ways [15]. Learning Analytics Dashboards (*LAD*) share a similar objective [16]. Unlike learner models, LADs rely on a static representation of behavioural metrics derived from interaction data, whereas learner models focus on modelling the learner's knowledge and other individual characteristics [16], [17]. Learner models and learning analytics both require personalised learner data, 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.

Learning environments are (digital) platforms where learners may use different content via various learning elements to educate themselves independently. Put simply, typical digital learning environments usually consist of various components – even when subsequently combined into a monolithic whole. Standard components include, for example, the learner model, which enables individualisation in the first place, something like a domain and competence module, where the content to be learned is structured thematically and competence-oriented (comparable to a module handbook for degree programmes), and the platform with the correspondingly tagged learning elements and the possibility of adaptive learning paths. However, the components can also be divided into smaller parts and cut according to function [18]. Learning environments are not defined further here, as they are sufficient for the meta-level context. 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. Of course, digital learning environments can also be characterised and specified in greater detail, and, in some cases, even classified. For this publication, however, it is only important to note that the term "learning environments" refers to learning platforms that collect implicit and explicit data about learners to support adaptive learning recommendations and adjustments to learning elements.

III. Usage Scenarios

The learning process is a highly multifaceted and complex phenomenon that educators and educational theorists have studied for many centuries. It serves as the basis for effective digital implementation in a digital learning environment.

A. *Learning Process*

Individual adaptation of the learning process is only viable if information and data about the learner can be used to respond to their needs 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 individualised support to users, helping them make better decisions [19], often by building on learner models.

The typical sequence of a possible learning process in digital adaptive learning environments, in which the learner receives support through recommendations for learning elements and adjustments to their individual learning path, is defined by the *LEARN* model (**L**earning **E**nvironment for **A**daptive **R**ecommendations and **N**eeds) [18]. The learner selects a subject area in the learning environment to explore or learn about. For this subject area, the system recommends a learning element that best matches the learner's preferences, current knowledge level, and prior knowledge, based on data stored in their personal learning model. Further alternatives are presented in a ranking list from which the learner can choose if they do not like the recommendation. Once the learner has decided on a learning element, they work through it and complete it with a knowledge check – for example, a quiz or an assignment. Throughout their time in the learning environment, new interaction data is collected and fed back into the learner model, along with the learner's feedback and knowledge-check data. Based on the new information obtained, the next appropriate learning element is then recommended, leaving it up to the learner to continue or take a break for the time being [18].

Figure 1 visualises this 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.

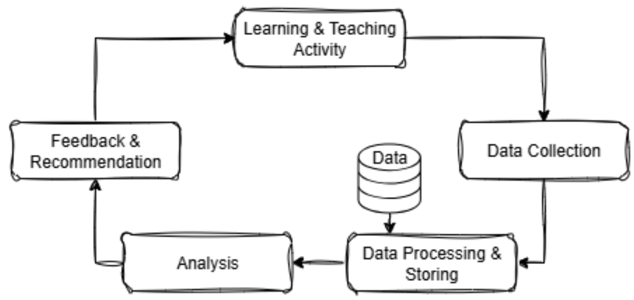


Figure 1. Workflow Reference Learning Process [1]

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 within the overall context of the digital learning environment and its associated relationships, as the added value of learner models only comes into play when they are used in the adaptive learning platform.

B. *Possible Reference Architecture*

There are many approaches aimed at providing learners with personalised learning materials and pathways based on various learner characteristics, such as knowledge level or learning preferences. These approaches often differ in their underlying architectures, the techniques they use, for example, in learner modelling, but also in other design decisions and, finally, in the technologies used for implementation. However, existing research often lacks a holistic view of the learning environment, particularly regarding the system's use in real-world scenarios. The systems are evaluated and tested in a defined environment, disregarding the fact that they must meet further requirements in daily use. Simulations in protected test environments often overlook data use and the need for proper student consent for data collection. However, data protection in general and the protection of personal data in particular have become significant legal issues, particularly in Europe. Furthermore, integration into existing software systems and infrastructure – such as organisational software in a university context – is not considered. A few publications mention the components used in their systems or frameworks, with the focus usually on the recommendation component and the algorithms and techniques employed [18].

Böck et al. have already applied the potential learning process described above to a possible reference architecture in another publication [18]. The infrastructure of the reference architecture for digital learning environments can be broadly divided into five sub-components: adaptive learning elements (managed by the learning element registry), the core, the recommender engine, and the two models: the learner model and the domain and competence model. For example, the management of learning courses or learner authentication is handled by the core component. The domain and competence model contains a machine-readable representation of the domains that can be learned using learning elements on the platform, along with their difficulty levels and associated competence levels. These cannot be viewed in isolation but are always linked to a specific domain. Together with the learner model, the domain and competence model form the basis for the recommendation engine to generate the most suitable recommendations for the next learning elements. The architecture consists of loosely coupled, highly coherent modules that enable flexible communication, ensuring efficient processing and high user acceptance. This structure facilitates the easy integration of new systems and the addition of interfaces. As the system is scalable, affected components can handle increased user volumes as required, and updates to individual modules can be carried out independently without affecting the overall system. The architecture is generalised, so that a wide variety of domain and competence models are supported in compatible formats.

C. *Affected Levels*

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. However, we will also consider the meso and macro levels to a large extent if they have similar or even identical requirements and differ only in their objectives.

However, Figure 1 can be applied not only to the learner perspective (micro level) already mentioned, but also to the meso and macro levels. At the meso level, individual learners can be grouped into a cohort, providing the lecturer with information about a specific learning element or topic. After many learners have completed learning elements, aggregation is added in the 'Data Processing & Storing' step to summarise data from multiple learners in the same learning area. In the analysis step, cohort analyses are carried out, for example, to identify interaction patterns and challenges or learning difficulties for the entire group of learners in specific sub-areas or learning elements, and to counteract these with measures. This information is reflected to the lecturer in the 'Feedback & Recommendation' step, including tips and possible measures to address these challenges, for example, by adapting the content or providing additional learning elements. It is not a single learner that is of interest here, but the cohort as a whole in pseudonymised or even partially anonymised data for the lecturer.

At the macro level, it is viewed from a more distant meta-level. This involves activities across courses and degree programmes, so platform-wide use is relevant here. In the 'data collection' phase, aggregated course data is collected, including completion and dropout rates, as well as accessibility and usage metrics. In the next phase, this collected data is stored long-term and linked to institutional data, such as administrative data (e.g., applications, enrolments, or other data). The analyses aim to simplify strategic decisions. Therefore, comparative analyses of different courses or study programmes, as well as longitudinal trend analyses, are carried out to identify systemic patterns and derive measures. This then happens in the 'Feedback & Recommendation' phase, in which strategic institutional recommendations for the management of the educational institution or institute are derived from the analysis results, for example, to optimise curricula and degree programmes, as well as to make improvements at the platform level.

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. It is normal for several actors to be involved in a learning process. In most cases, however, the selection in adult education can be reduced to three relevant actors: the student (referred to as a learner from now on), the teacher, and the educational institution that provides the formal setting for the education. 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. However, the resulting synergies for the other two actors and related levels in the learning process are also highlighted and considered.

D. *Purpose of Use*

Several publications, e.g., ISO/IEC JTC1/SC36 [19], [21], [22], already present typical scenarios of digital learning at different levels. Teaching and learning activities in digital learning environments are the starting point for analyses of digital learning. Learning analytics are used to personalise and thus improve the learning environment adaptively, and to summarise methods such as learner models and learning analytics. Based on the results of the learning analysis, the learning environment can recommend individualised learning paths alongside 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 for non-technical users to understand, if they could be viewed at all. Nowadays, learners can view visual displays of their data in dashboards, e.g., to monitor their learning progress relative to the cohort average. 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 supports learners with disabilities and helps identify accessibility deficits in learning opportunities.

Instructor (*meso level*)**.** Instructors can track the activities of their entire cohorts and detect progress and potential problem patterns, e.g., a 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 by enabling analyses, which can also reduce dropout rates in the long term. In addition, administration may benefit from information on the entire learner 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 derived, such as a potential accessibility problem.

Table I below summarises the various levels, stakeholders, data types, and their granularity to provide a quick overview. These three levels, along with the three stakeholders, run throughout the publication.

TABLE I. OVERVIEW OF LEVELS IN RELATION TO THE STAKEHOLDERS

Levels	Stakeholders	Type of data (granularity)
Micro	Learner	Personal data (in as much detail as possible)
Meso	Instructor	Pseudonymised data (standard)
Macro	Educational Institution	Anonymised data (abstracted)

E. Legal Principles and Classification

The implementation of learner models in educational institutions has various legal implications. For example, intellectual property is protected from the developer's perspective. On the other hand, learners' rights, especially their fundamental rights, must be taken into account. In the EU, the fundamental right to data protection, as stated in Article 8 of the EU Charter of Fundamental Rights (*CFR*) [20], is affected. The specific legal implementation of this fundamental right, the General Data Protection Regulation (*GDPR*) [21], 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 to enable the creation of individual learning paths based on a wide range of criteria. Many publications, particularly in the last decade [9], deal with adaptive teaching and learning settings. For this work, scientific publications on user models, particularly learner models, are the most relevant. Koch [22], [23] 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 address learner analyses, but mostly only as a means to the end of adaptive learning systems [27], [28], or only very superficially, without going into relevant details. Several publications also consider only individual information rather than aggregating it into a learner model [24], [25]. A couple of surveys attempt to create an overview of the field [26], [27], [28], [29], 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 the required data comes from.

Therefore, we conducted two parallel systematic literature searches on learner models: on the one hand, to obtain a comprehensive picture of the current state of scientific research [9] and, on the other hand, to clarify the state of practice based on standards and norms [8]. We decided to conduct a systematic literature review (SLR) following the methodology described by Kitchenham and Charters [30], [31]. Ideally, this should pave the way for a standardised approach to designing and creating learner models, as well as a set of standardised subcomponents – regardless of their purpose or use. 868 standards were reviewed, 16 of which were classified as relevant to the structure and components of learner models. Three standards focus extensively on learner models. Among those, ISO/IEC 29140:2021 [19] describes a mobile learner model that considers specific attributes, such as the device used, connectivity, and the learner's location, to better describe the learning environment. The 1EdTech consortium [32] focuses on interoperability among internet-based information systems that support learners' 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 [33] 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 are either specialised [20], complex and more than just learner models [37], or abandoned [33]. No standards have been found that describe a generalised, realisable learner model that can be extended for specific use cases, or that address the creation of such learner models [8].

In parallel, scientific publications from relevant publishers (IEEE, ACM, Elsevier, pedocs) from 2014 to 2023 were systematically examined, yielding 197 papers deemed sufficiently relevant for detailed analysis [9]. 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, or whether 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 [9].

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 [34], [35], [36], [25], [37], but do not describe in detail how these are expanded or what the implementation or modelling of the learner model looks like. In summary, there are no ready-made, predefined learner models that can be downloaded and customised. Furthermore, the individual learner models described in the two literature reviews provide only very limited information regarding their implementation details.

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 [38], [39], [40], [41], [42], [43]. Our main objective is to contribute to this standardisation process by proposing an original reference model as a first step towards a reference architecture. This contribution is based on defining open software interfaces for each subsystem in the architecture, thereby avoiding any dependence on specific information models. We have already discussed elsewhere a possible reference architecture for an adaptive learning platform that integrates a learner model [44].

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 separate publications, we have already examined the functional requirements for learner models in greater detail [44] and provided initial insights into the legal classification of these requirements [1]. We have also focused on the broader context of digital learning environments and the integration of learner models [18]. Based on the literature review, the following section will focus on the fundamental functional requirements, the non-functional requirements, and the legal basis for the use of learner models in Europe, especially in Germany. The following section is by no means comprehensive; rather, it collects and summarises the most relevant requirements, based on their occurrence and description in the literature, and links them to current legal restrictions in Europe, especially in Germany.

In addition to the analysed publications, the feedback and experiences of learners at our university play an essential role. It is important to emphasise again that learner models are structured models without any interference mechanisms [10]. However, it is also necessary to evaluate the effects of integrating a learner model into the overall eLearning environment, given the requirements for a learner model.

The following requirements are classified according to the *Kano model* [45] into so-called *must-be requirements*, one-dimensional requirements (*should characteristics*) and attractive requirements (*can characteristics*). The requirements are supplemented with classic examples to facilitate understanding and, where possible, to address any challenges or difficulties that may arise during implementation. The requirements were classified, on the one hand, based on the frequency of mentions in the publications analysed during the systematic literature reviews and, on the other hand, through the authors' assessment of which requirements are essential for operation, which requirements are desirable enhancements to the basic functionality, and which requirements would certainly be useful but are not strictly necessary for the implementation of a functional learner model. These two metrics were consistent with one another. This classification was validated by the authors' experience during the iterative development of their own learner model, as well as by feedback from reviewers of previously published papers and from discussions at conferences attended by experts in these fields.

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. While data about the learner is collected and stored at the micro level, such as (successfully) completed learning elements, behavioural data (in the learning environment, such as interaction data) and knowledge data (e.g., last degree, modules already taken, ...), this information and data are relevant to lecturers at the meso level in aggregated and pseudonymised or anonymised form. The same applies at the macro level, where the data in its entirety serves as the basis for all analyses. The challenge in collecting and persisting the data lies in its granularity and in selecting the data relevant to learning.

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 within the definition of artificial intelligence systems under 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) AI Act [46]), 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 comply with the applicable legal requirements of the *GDPR* [21], [47]. 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 cannot be generalised or answered in general terms; they must be legally considered and examined individually, depending on the purpose of use and the data to be collected. These questions also apply to data aggregation, i.e., the application of the learner model at the meso and macro levels. The challenge here is to check, or have it checked manually, the legal aspects of each application and the required data, to determine whether the project can be implemented legally as planned or what conditions apply. The concept of a data trust model might be applicable here: a neutral, trustworthy entity ensures that data is processed and used in accordance with defined data protection guidelines and is accessible only to 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 but 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 in how the system reaches a result [53] creates trust and acceptance of certain subordinate recommendations, which form the basis of good support [48]. Therefore, learner models should be transparent at every step of the process, 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 to illuminate backgrounds and issues, such as potential discrimination [54], [55], [56], in categorisations. Transparency thereby relies on individual decisions so that they can be traced correctly. Traceability is achieved by involving the learner from the outset and by adopting a learner-centred approach to the design and development of learner models. This helps learners better understand how a decision is made by showing the individual steps involved, the data used, and the information that led to each decision. For example, in a subject area, repeated failures on knowledge tests – such as quizzes, exercises, or similar – could indicate that the learner has not yet internalised the material and therefore needs to repeat it. This information should then be made visible to the learner to improve self-reflection. The difficulty here, however, is that with many machine learning methods, the decisions made do not always allow conclusions to be drawn about the reasons for those decisions, as would be the case with a decision tree, whereas a neural network does. Technically speaking, there are already some approaches: Explainable Artificial Intelligence (XAI) [49], [50] encompasses methods and techniques that make the decisions of machine learning and artificial intelligence systems comprehensible and interpretable to humans [51], [52]. This transparency is also important at the meso and macro levels to deepen understanding among lecturers and educational institution management, and to initiate meaningful measures, above all.

Responsibility. Traceability is closely connected to responsibility. The data that may be collected must be left to users' choice, subject to applicable legal standards. That is, learners can decide which data to log and persist. This topic also includes compliance with ethical principles. In the digital domain, this means that if the learning management system demonstrates ethical behaviour, learners expect it to comply with 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*). Specifically, from the learner's perspective, this means at the micro level that they alone want to decide what data they want to make available to the learning environment and that they can view and change this at any time, for example, whether, in addition to their learning history, the grades of modules and examinations already taken during their studies should and may be used. The system should then comply with this decision and not 'secretly' collect and use the data, as this would demonstrably weaken trust in the learning environment. Considering the other two levels, the responsibility lies in ensuring that this highly personal data is used only by the learner and cannot be viewed, let alone used, by third parties. This means that at the meso and macro levels, the data must not allow any conclusions to be drawn about individual learners, but should provide information about aggregated states, such as: ~80% of course participants have difficulties with the topic' complexity theory' in the learning element 'P-NP problems and NP completeness' (#0815), or the module' theoretical computer science' is very rarely taken voluntarily as an elective module, with only ~3% of learners choosing it.

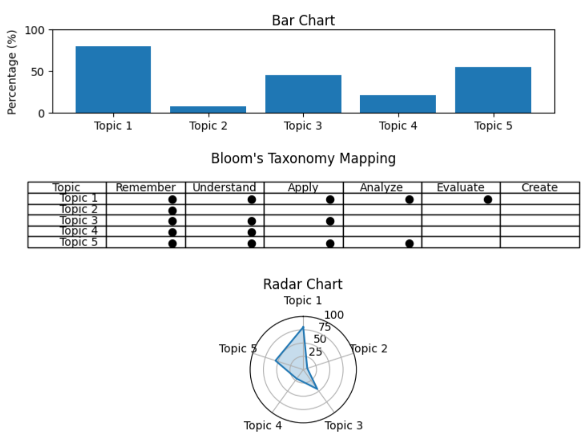
Fairness & Ethics. Fairness affects many different steps within the learning process. For example, algorithms for decision-making and data processing must be checked for possible biases to ensure equal opportunities for all learners [53], [54].

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: that unauthorised data access must be prohibited, and that, if access is permitted, any changes to data shall be checked for plausibility and logged for tracking purposes. This requirement is closely linked to traceability. Learners must be able to be one hundred per cent certain that their data will only be used for the purposes for which they have given their consent, i.e., to assist them in their learning and in reaching their learning goals, and that this personal data cannot be viewed or altered by any unauthorised third party. At the meso level, for example, the lecturer should not be able to identify individual learners based on their learning difficulties, but should only be able to obtain an overall picture of the cohort. The same applies at the macro level, of course. Any inference about individual learners must be prevented by all possible means. The only exception is the learner's prior consent, if they require external assistance, to temporarily disclose a selection of their data to specific third parties.

C. Can Requirements

Could-be requirements represent possible extensions that might optimise the learner model in certain scenarios, e.g., by providing additional benefits or improving user experience. If necessary, these requirements can be incorporated into future development phases to increase the model's flexibility and adaptability.

Openness & Visualisation. Learner models should be open [15], [55] to promote metacognitive behaviours, such as self-awareness and self-regulation. This means that learners may display and analyse their own instances of a learner model to better understand their learning progress and, e.g., misconceptions [14]. 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 [56], [57], [58] are essential for presenting complex learning data to the learner in a clear and concise format. Different representations of the same data can bring learners closer to the various aspects and clarify them [59]. Leaner models designed in this way are called open learner models [60], [61]. The difficulty in opening the learner model is choosing the 'right' visualisation method, which depends on the available data. This is because the visualisation should not overwhelm the learner, but should provide further insights into their learning behaviour, ideally in a self-explanatory way. There are numerous visualisation options for the same data. For example, the knowledge learned can be represented using simple bar or column charts, tables, or radar plots (see Figure 2 for conceptual examples), to name just a few visualisation options [62], [63].



Topic	Remember	Understand	Apply	Analyze	Evaluate	Create
Topic 1	●	●	●	●	●	
Topic 2	●					
Topic 3	●	●	●			
Topic 4	●	●				
Topic 5	●	●	●	●		

Figure 2. A few random examples of visualisation options

The same applies to the meso and macro levels. Visualisation options should provide both teachers and educational institutions with simpler, more easily understandable insights into the data and its interrelationships. For example, the lecturer could see, in pseudonymised form, where learners might be having problems (see Figure 3 as a conceptual example), or the management of the educational institution could see which subjects are popular based on enrolment figures over the last few years (see Figure 4 as a conceptual example).

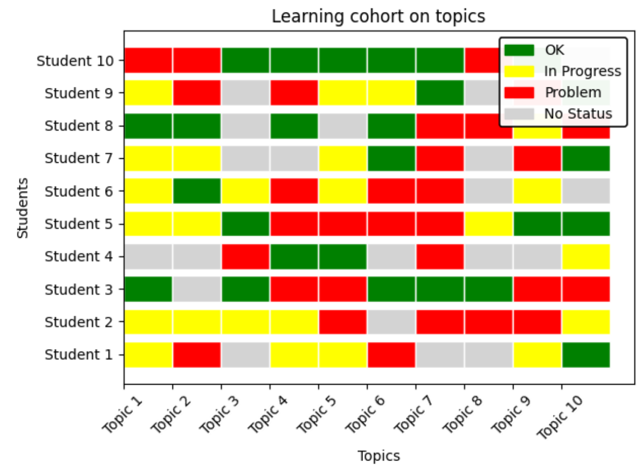


Figure 3. Possible view of a lecturer

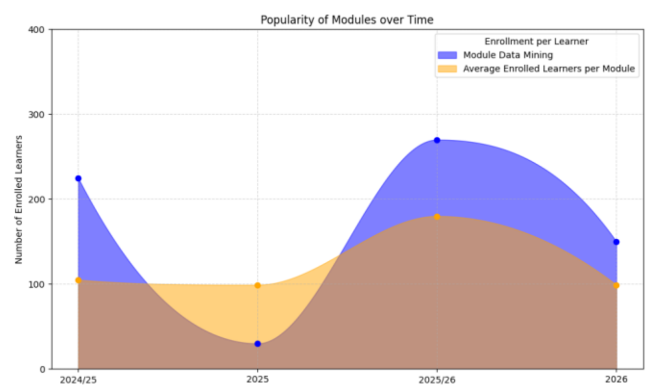


Figure 4. Possible view for the administration of the educational institution

Negotiation Options. In addition to the pure visualisation of personal data, some learner models also offer the option to interact with this data and, for example, negotiate with the model if the data shows inadequacies from the learner's perspective [14], [63]. 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. This requirement is particularly relevant at the micro level, as learners are allowed to correct data they believe is incorrect. For example, if the learner's current level of knowledge does not align with their self-reflection and they believe they are better at a particular topic, they may be able to update this information. There are various approaches to adapting information, which can often be divided into four main types (*according to* [64]):

- inspectable, i.e., the learner can view some or all of their stored data using various visualisation options.
- co-operative, i.e., depending on how easily everyone can fulfil the respective modelling tasks, these are divided up.
- editable, i.e., the learner can change the content as desired.
- negotiated, i.e., the learner and the system discuss the stored contents of the model and agree on a joint solution.

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. In addition to the purely legal considerations of data economy and minimisation, unnecessary data places a strain on the existing infrastructure through the collection, transmission and storage of information. Of course, the initial challenge is to identify which data are relevant for learning. It can support causal inference, but once these questions have been largely clarified, the collection and retention of data that is not relevant for learning should be discontinued.

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. This requirement is independent of levels, as it concerns the overall system rather than the learner or the teacher; in the broadest sense, it also concerns the educational institution, as it involves a technically directional architectural decision and thus becomes less dependent on individual components.

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 secure prospects. Like the previous requirement, this one is also level-independent, as it concerns the overall system.

Standardised / Interoperability & Integration. Ideally, learner information is standardised so it can be easily exported and exchanged across different learning platforms (in line with the interoperability defined in [65], [66]) 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). The challenge clearly lies in providing a possible standard for so-called learner models. Education providers will have to continue addressing this issue over the next few years to arrive at a commonly recognised and accepted standard. However, aside from the challenge, the advantages of standardisation are enormous. For example, when changing educational institutions, learner data can be easily exported and reimported with minimal effort (at the macro level). From the learner's perspective (micro level), learning can continue directly where it left off, without having to start again from the beginning, and the current level of knowledge in various subject areas can be determined through pre-tests or learning preferences assessed via a questionnaire.

One option worth considering is to transfer the learner model topic to a so-called data trust model and treat it as an external entity that can be integrated and used with the learner's permission.

VI. Legal Constraints

Legal considerations are crucial in the design and development of learner models, as lawful use is only possible if these models comply with applicable regulations. It is important to emphasise that the following legal analysis and the resulting practical recommendations are strictly oriented towards the European Union's regulatory landscape. Therefore, while the technical requirements may be universal, the legal compliance strategies discussed here have limited transferability outside the EU jurisdiction.

. Most data processed in learner models falls within the material scope of the GDPR pursuant to Art. 2(1) GDPR, as it constitutes personal data within the meaning of Art. 4 No. 1 GDPR. Personal data encompasses any information relating to an identified or identifiable natural person, including not only basic information such as names or email addresses, but also learning behaviour, performance metrics, and other characteristics that allow conclusions to be drawn about individual learners. The feedback from learner models also contains information about the learners, making it 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 be collected only if there is a legal basis (lawfulness). All the aforementioned legal principles are enshrined in the General Data Protection Regulation (*GDPR* [47]). 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. Notably, the GDPR is technology-neutral (recital 15 GDPR), requiring the broad terms it employs to be defined explicitly in the context of learner models.

To ensure analytical precision throughout the subsequent legal evaluation, a fundamental distinction must be drawn between learner models and learning analytics, as their technical characteristics dictate varying levels of data protection risks. In their standard implementations, both applications inherently process personal data within the meaning of Art. 4 No. 1 GDPR. However, a significant divergence exists regarding the feasibility of anonymisation: because it focuses on cohort-level patterns and aggregated behaviours (meso and macro levels), learning analytics often enables complete anonymisation. Once data is irreversibly anonymised, the personal reference is severed, rendering the GDPR entirely inapplicable (Recital 26 GDPR). Conversely, learner models are structurally tied to the individual (micro-level) to facilitate tailored pedagogical adjustments, making pseudonymization the primary safeguard while the data remain personal. Whenever anonymisation cannot be achieved, the full spectrum of GDPR obligations applies equally to both applications. Consequently, the requirements developed in this framework are fundamentally relevant to both domains. Ultimately, the precise operationalisation of these legal obligations remains a case-by-case determination, as the specific compliance risks heavily depend on the concrete system architecture and data categories involved. In general, however, learner models present a higher intensity of data processing, as they continuously track and model the nuanced learning status, competencies, and developmental trajectories of individual natural persons.

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, at the time of writing this publication, 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 into the process of designing and implementing learner models in educational institutions as soon as possible. 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 under the GDPR and highlight key considerations 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 challenges regarding learner models.

Lawfulness - Legal Bases. Any processing of personal data requires a legal base, as stated in Art. 5(1), Art. 6(1) GDPR. Art. 6(1)(a) GDPR establishes consent as a legal basis, which must be given voluntarily [67, 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. However, this power asymmetry is not present in all educational scenarios. 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 (1)(e) GDPR serves as the legal basis for data processing when processing is necessary for the performance of a task carried out in the public interest, such as education. However, under Art. 6(3) GDPR, the legal basis of Art. 6 (1)(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 imposed by the GDPR on national law are discussed in [68].

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(2)(g) GDPR, alongside the requirements of Art. 6 (1)(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.

For learners, this means that, at the micro level, the learning environment is an additional learning opportunity that is often not mandatory but that supports the learner individually in their learning process. To this end, learners may revoke their consent at any time to the use of their personal data for individual learning support without incurring any disadvantages to their grades or final assessment by the lecturer or educational institution. This additional service is purely voluntary for learners who wish to receive it. Looking at the data to be collected in order to enable the most successful individual adaptation possible, this personal data is divided into two categories: personal data and sensitive personal data, Art. 9 GDPR. The latter includes data relating to health, religious, or philosophical beliefs. However, this data may be relevant to learning. For example, information about the learner's health can help tailor the programme to their needs, such as allowing more breaks if they have been diagnosed with attention deficit disorder, or omitting spelling from the assessment if they have been diagnosed with dyslexia. The same applies to religious data, for example. If this data influences learning behaviour, for instance, because prayer times interrupt active learning, this should be considered by the system, which means that the necessary data is also required. This data should be collected and processed with absolute caution only after explicit notification and, ideally, with the learner's explicit consent (Art. 9(2)(a) GDPR), if necessary to achieve the goal of individually tailored learning support. From a meso- or macro-level perspective, this issue is somewhat simplified when it concerns anonymised, aggregated data at most, which should not allow any conclusions to be drawn about individuals.

Purpose limitation. Developers of learner models must carefully evaluate the potential sources of personal data used for their development and operation. Often, data is collected for a different purpose — for example, when learners' exams, initially collected for grading, are fed into learner models. In addition, the subsequent use of data generated by the learner model should be carefully evaluated. Art. 5(1)(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(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(4)(a)–(e) GDPR. Scientific research, which is what learner models could be part of, is privileged. Art. 5(1)(b) assumes scientific research is "not be considered to be incompatible with the initial purposes".

This requirement becomes especially relevant when repurposing existing learner data that was initially collected for other purposes, such as administrative processes, examination grading, or assessment of learner work. As this data is often ideal for populating and utilising learner models, an evaluation is required to determine whether such a change of purpose falls within the privilege afforded to scientific research.

Transparency. Transparency is not merely a technical requirement but is also enshrined in Art. 5(1)(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(2)(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 et seq. 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(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.

When learners exercise their right of access to their personal data, it must be ensured that all processing of personal data is captured. This encompasses registration data (such as name, learner ID, and email address) used for login, metadata related to system usage, and, most importantly, all input data entered by learners into the system. Additionally, this includes the inferences the system draws from such data, as well as the feedback it provides.

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 [69] mn. 96. To comply with this principle and the requirements of Data Protection by Design (Art. 25 GDPR), institutions are legally obligated to implement appropriate technical and organisational measures. While the GDPR does not dictate one specific technical solution, pseudonymisation, as mentioned in Art. 25(1) GDPR is strongly recommended as a primary safeguard to minimise risks. According to Art. 4(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. *EDPD*, the European Data Protection Board, provides a detailed guideline for pseudonymisation [70]. As an organisational matter, a role concept should be implemented that specifies who can access the data collected in the learner model and who can edit it. For complete guidance on technical and organisational measures according to Art. 25, see [71]. In particular, the scope of people who can undo pseudonymisation needs to be kept small.

Also, the use of a data trustee can be discussed [18]. 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. Data to be collected and processed must be carefully selected, i.e. only for the purpose of optimising the learning process – in other words, adaptive learning. Data that is not required in this context may not be collected or processed further. Pseudonymisation is typically feasible at the meso level. At the macro level, the identity of individual learners becomes irrelevant, allowing personal information to be minimised or even completely removed through anonymisation.

Storage Limitation. The principle of storage limitation, as outlined in Art. 5(1)(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 [69] 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 learners' personal data is technically feasible. Typically, learners' personal data should be deleted once they leave the educational institution. If the data is intended to be retained for further optimisation of the learner models, the link between the data and the individual learners can be erased entirely and irreversibly (anonymisation) [72]. At the micro level, the data collected and processed should be periodically reviewed to determine whether it is still needed or can be reduced to the bare minimum. Nothing further needs to be done at the meso and macro levels, if only pseudonymised or even anonymised data is available at any given time, making it impossible to identify individual learners.

Integrity & Confidentiality. Art. 5(1)(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 one 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, data transfers to such servers must comply with Art. 45 et seq. GDPR, which aims to ensure continuity of the level of data protection [73] 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, so-called Data Processing Agreement (*DPA*). 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 can analyse learners' input, it could be used for automated assessment, for example. However, it is important to emphasise that the learner model itself does not make any evaluations or decisions; it functions strictly as a structural data representation. Consequently, any legal restrictions or prohibitions under Art. 22(1) GDPR apply solely to the downstream system or processing logic that operates upon this data. Art. 22(1) GDPR states that decisions based solely on automated processing, including profiling, which produce legal effects, are prohibited. Art. 22(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. At least for Germany, no such legal basis is known at the time of writing this publication.

VII. The Relationship between Technical and Legal Requirements

Many requirements from technical and legal perspectives are quite similar, although these requirements should be viewed from both angles. Transparency serves as a key example. From a technical perspective, it is essential to design the learner model as transparent and traceable as possible to build trust in the system. Therefore, it might make sense to limit the displayed data, such as learning progress, to prevent information overload. The transparency required under Art. 5(1)(a) GDPR and Art. 12 et seq. GDPR, on the other hand, aim to give the data subject an extensive overview of all their processed data. Thus, a progress bar showing learners' learning progress may fulfil design transparency requirements, but it does not satisfy GDPR transparency obligations. Similar analytical tensions emerge across other core requirements. Regarding openness and negotiability, pedagogical design often allows learners to manually alter their skill levels to foster self-reflection and motivation. However, the legal principle of accuracy (Art. 5(1)(d) GDPR) and the right to rectification (Art. 16 GDPR) demand that institutional data objectively reflect reality, especially when tracking prerequisites. Consequently, to maintain data integrity, a system must store both values separately if a learner overrides a calculated skill state. Furthermore, interoperability reveals a conflict between de jure rights and de facto technical fragmentation. While Art. 20 GDPR grants learners the right to data portability in a machine-readable format, the lack of widely adopted, semantically unified learner-modelling standards means that mere technical transferability does not guarantee pedagogical utility in a new learning management system. Finally, ensuring fairness and ethics highlights an inherent 'accessibility paradox'. To guarantee technical fairness and automatically provide accessible learning materials (e.g., tailored formatting for learners with dyslexia or ADHD), the system must inherently process highly sensitive health data. This immediately triggers the stringent legal restrictions of Art. 9 GDPR, demonstrating that pedagogical adaptivity is frequently constrained by necessary legal privacy safeguards. Overall, this demonstrates that while technological opportunities and data protection regulations can mutually reinforce each other, they often rely on divergent interpretations or result in inherent conflicts of interest.

The learner model must also be evaluated from an ethical perspective, which may not be fully represented in design and legal considerations. It might be legal to collect various personal information about learners over several semesters, but it remains questionable from an ethical perspective. Many higher education institutions have their own ethics commissions, which often need to agree to the use of new technologies for learners. Therefore, both the internal ethics committee and the legal department, together with the data protection officer and, of course, the learners, should be involved in the process from the outset.

To make it easier to understand, we have summarised the technical requirements—both functional and non-functional—in Table II below. We have categorised these requirements by classification and affected level and, where there are legal implications, included them as well; this serves as a synthesis of the two sections described in detail above.

TABLE II. Overview of levels in relation to the stakeholders

Levels	Requirement	Primary Legal Principles	Technical Implications
Micro, Meso, Macro	Collection and Management of Learning Data	Art. 5(1)(c), Art. 5(1)(b) GDPR	To work with data, it must first be collected and stored. In the case of personal data, this always requires the data subject's consent.
Micro	Transparency / Traceability	Art. 5(1)(a) (Transparency), Arts. 12–14 (Information obligations), Art. 15 (Right of access) GDPR	Technical transparency optimises UX by reducing cognitive overload (e.g., simple progress bars). Legal transparency requires a comprehensive, granular disclosure of all processed data, including backend

Levels	Requirement	Primary Legal Principles	Technical Implications
			algorithmic weights and profiling inferences.
Micro	Respon-sibility	---	The responsibility is to abide by the learner's decision and not to do anything without prior agreement that could undermine trust. Even though there is no legal basis for this, it is important to meet this requirement.
Micro	Fairness & Ethics	Art. 5(1)(a) (Fairness), Art. 9 (Processing of special categories of data) GDPR	The Accessibility Paradox: To ensure technical fairness and automatically provide accessible material (e.g., for dyslexia or ADHD), the system must know about the condition. This requires processing highly sensitive health data, which is subject to stringent legal restrictions.
Micro	Tamper-proof	Arts. 25, 32 GDPR	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.
Micro, Meso, Macro	Openness & Visualisation	Art. 5(1)(a),(d) (Accuracy), Art. 16 (Right to rectification) GDPR	Pedagogical negotiability allows learners to alter their skill levels for self-reflection or motivation. Legal accuracy requires that institutional data objectively reflect reality (especially when used to track prerequisites). If a learner manually overrides a system-calculated skill state, the database should store both values separately.
Micro	Negotiation Options	---	The various negotiation options involve purely technical interaction and are not based on any legal framework.
Micro	Data Minimisation & Sustainability	Art. 5(1)(c) GDPR	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.
Micro, Meso, Macro	Modularity & Flexibility	---	This involves a wide range of technical aspects and is not subject to any legal requirements.
Micro, Meso, Macro	Maintain-ability & Expandability	---	This involves a wide range of technical aspects and is not subject to any legal requirements.
Micro, Meso, Macro	Standardised / Interoper-ability & Integration	Art. 20 (Right to data portability) GDPR	The GDPR grants a legal right to transfer data in a machine-readable format; however, due to the lack of widely adopted, semantically unified learner-modelling standards, technical transferability does not guarantee pedagogical utility in a new LMS.

VIII. Recommended Action and Guidelines

Now you are a reader and possibly a future designer or developer of learner models, and you want to take the first step towards adaptive learning environments with learner models, or to integrate a learner model into an existing learning environment. However, you do not know how to get started. With many vague requirements, the GDPR can be hard to comply with, especially without legal expertise. The following section aims to provide a short roadmap for tackling GDPR compliance in a real-world scenario when learner models are to be integrated into higher education. Since a list for full compliance would be too extensive for this work, we focus on the essentials.

A. Common Process

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 primary focus areas. However, this does not mean that the amount of data per se must be reduced; rather, 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, the amount of data to be collected should still be critically examined to determine whether it is necessary and can make a meaningful contribution, and thus be used.

In addition to legal regulations, technical and non-legal requirements must be considered to design learner models that are practical and effective. The variety of possible requirements analysed is wide. 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. In addition, these can help increase transparency and build trust in the learning environment, as the data collected (in excerpts) is visualised, providing insight into what data is being collected and how it contributes to learning support.

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.

B. Recommended Strategy

The following recommendations for action are intended to simplify the introduction for non-lawyers and provide an overview of the most important topics and issues. To this end, we have summarised the top five most important legal points.

1) Assessment of Personal Data

First, developers of learner models must consider the data required to develop and use the model. For instance, one could need data from learners to fill the model. Once the needed data is determined, developers must ask themselves whether this data will give any information about the persons. If so, these data fall within the material scope of the GDPR (Art. 2(1) GDPR). Developers should reconsider whether this specific personal data is really needed. Could the learner model be developed and used without this information linked to an identifiable person? The use of learner models primarily involves processing personal data. All this data should be listed, as it will be needed later, e.g., for the privacy policy.

2) Legal Basis for Processing the Personal Data

Afterwards, it must be ensured that there is a legal ground for processing the personal data, as described under VI. The legal grounds can be discussed with the entity's data protection officer later. If the learning model is to be used at a later date to grade learners, it falls within the scope of Art. 22 GDPR and is generally prohibited.

3) Safeguards

In the following step, developers must consider safeguards, referred to as appropriate technical and organisational measures in Art. 25 GDPR. In most cases, pseudonymisation of personal data is recommended. Regarding data access, a role-based access model should be implemented, distinguishing between user groups such as teachers and researchers. Access rights must be granted in accordance with this role concept, ensuring that each user can only access data necessary for their specific purpose. Furthermore, encryption should be applied both in transit and at rest to protect the data from unauthorised access. Finally, a deletion concept must be established to ensure that personal data is erased when it is no longer required for the purposes for which it was collected, or, at least, anonymised if it remains relevant to the system.

4) Feedback by the Data Protection Officer / Documentation

Data protection officers usually provide a form to document the aforementioned aspects of data protection. The elaborated concept can be discussed with the data protection officer at the higher education institution to obtain final approval. Higher Education institutions must prepare comprehensive documentation for learners and other data subjects, including a privacy policy that explains the AI system in an accessible language and informs users about their rights under the Arts. 15-22 GDPR, especially the right to object under Art. 21 GDPR for learning analytics.

5) Next Steps

In addition, they should be in close contact with their data protection officer to prevent data breaches. At the same time, they should always obtain ethics committee approval before beginning implementation, even if they already have legal approval.

C. Overview and Checklist

In the appendix, we have compiled and summarised some key information to help non-lawyers get started. This is intended as an initial introduction and is in no way a substitute for binding legal advice; rather, it is designed to raise awareness of this highly complex subject.

The checklist in Appendix A provides a quick overview of the most important laws and their classifications. It serves as an initial introduction and working basis. It must then be adapted and expanded to suit your own needs in consultation with specialists (data protection officers and ethics committees).

⇨ *See Appendix A*

We have compiled and briefly summarised key legal principles in Appendix B to ensure that no important principles are missed. This list is not exhaustive and does not cover all cases; rather, it serves as a simplified introduction and quick overview.

⇨ *See Appendix B*

As many legal terms in data protection are not self-explanatory to non-lawyers, we have provided Appendix C, a glossary of the most important legal terms in data protection, as a reference guide. These are grouped into categories and briefly explained.

⇨ *See Appendix C*

IX. Conclusion and Future Work

A. Conclusion

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, in order to subsequently be able to use it in Europe, specifically in Germany.

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. The guidelines and recommendations for action in Section VIII are to simplify the introduction to this topic.

B. Challenges and Invitations

Even though the legal requirements within the EU on this topic are not easy to understand, further steps could be taken to develop 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. However, an open standard for the use of learning models would be important and useful not only from this perspective, for example, to enable educational institutions to differentiate themselves from other providers by offering targeted, individualised support to learners, thereby minimising dropout rates in the long term, but also, of course, from the perspective of learners, who would be more motivated to engage with the content as they are neither under- nor overchallenged, but are individually supported according to their needs [74], [75]. This well-known and scientifically proven flow experience has a demonstrably positive effect on learners and the content [76], [77]. Even though people with disabilities make up only a small part of our society, they should not be excluded but included. The first step is to create an inclusive learning environment that automatically recognises learners with learning disabilities and fully supports them in their personal learning journey through accessible offerings. Unfortunately, it is often overlooked or forgotten that, according to the Charter of Fundamental Rights of the European Union, "everyone [...] has the right to education and to have access to vocational and continuing training" [78] Art. 14, mn. 3. Another major legal advantage of an open standard would be that learner data is standardised. This makes it easier to export and import data to and from another learning environment. From a purely legal point of view, learners have already had this right since the GDPR came into force with Art. 20 (1) GDPR' Right to data portability', which enables them to "receive their personal data [...] in a structured, commonly used and machine-readable format". Apart from the legal framework, this enables easier, closer exchange between (state) educational institutions and provides a remedy when learners switch between courses at different universities, as they could transfer their digital data as a learner model through machine learning. Learners can also benefit from this, as they do not have to answer the same questions every time they register for a new learning environment. Instead, if their personal data were available externally, for example, through a data trust model, and they gave their permission to use individual services, their learner model would be maintained centrally, without redundant information across different educational institutions or learning environments. After completing their studies, this could include further professional training in a work context.

C. Future Work

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 [44] into an interoperable and legally compliant learning model, then to implement this model and integrate it into the learning platform [18]. We have already developed an initial prototype through an iterative process, refining it in successive cycles whilst implementing the various requirements in accordance with their classification (see Section V) [79]. The aim of this initial Minimum Viable Product (MVP) was to demonstrate correct functionality within a legal framework. To this end, this prototype was integrated into a proof-of-concept implementation of the digital learning environment and subjected to initial validation tests [80]. This initial prototype of a learner model 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. An (empirical) assessment can be divided into two aspects: (1) legal validation and (2) an evaluation of the learner model and the associated requirements that have already been implemented. Regarding (1), it can be said that a classic, schematic validation based on a catalogue of criteria is not possible, as the interpretation of legal standards is carried out using legal methodology and, in the event of a dispute, is determined in a binding manner by the courts. The situation is somewhat more difficult when it comes to assessing learner models (2). There is not really much published on this topic in the literature. The learning environment is often validated as an overall system, but not its individual components; if so, most likely the inference component, as shown by a further systematic literature review of possible methods for evaluating learner models [81]. We are therefore currently working on the design and implementation of possible assessments for learner models.

Acknowledgment

This work is a collaboration between the VoLL-KI project and the 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] F. Böck, H. Link, and D. Landes, 'Learner Models: Requirements and Legal Issues for the Development and Application of Learner Models', presented at the AISyS 2025, The Second International Conference on AI-based Systems and Services, Sep. 2025, pp. 27–35. Accessed: Oct. 24, 2025. [Online]. Available: https://www.thinkmind.org/library/AISyS/AISyS_2025/aisys_2025_1_40_80019.html

[2] N. Tagalidou, M. Bui, and M. Vukelić, 'Finding the Sweet Spot: Assessing Skill–Challenge Balance and Flow Using a Novel Game-based Task Paradigm', in *Proceedings of the Mensch und Computer 2025*, Chemnitz Germany: ACM, Aug. 2025, pp. 594–598. doi: 10.1145/3743049.3748556.

[3] S. Deterding and J. Cutting, 'Objective Difficulty-Skill Balance Impacts Perceived Balance but Not Behaviour: A Test of Flow and Self-Determination Theory Predictions', *Proc. ACM Hum.-Comput. Interact.*, vol. 7, no. CHI PLAY, pp. 1179–1205, Sep. 2023, doi: 10.1145/3611065.

[4] S. Zhang and S. Wang, 'Modeling Learner Heterogeneity: A Mixture Learning Model With Responses and Response Times', *Front. Psychol.*, vol. 9, p. 2339, Dec. 2018, doi: 10.3389/fpsyg.2018.02339.

[5] 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, Mar. 2021, doi: 10.1007/s10758-020-09448-4.

[6] E. A. Edmonds, 'Adaptive man-computer interfaces', *Computing skills and the user interface*, vol. 122, pp. 389–426, 1981.

[7] A. Johnson and N. Taatgen, 'User Modeling', in *Handbook of human factors in Web design*, R. W. Proctor and K.-P. L. Vu, Eds, Mahwah, N.J: Lawrence Erlbaum Associates, 2005, pp. 424–439.

[8] 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. doi: 10.5220/0012556100003693.

[9] F. Böck, M. Ochs, A. Henrich, D. Landes, J. L. Leidner, and Y. Sedelmaier, 'Learner models: design, components, structure, and modelling: A systematic literature review', *User Model User-Adap Inter*, vol. 35, no. 15, p. 81, Nov. 2025, doi: 10.1007/s11257-025-09434-4.

[10] F. Böck, 'A Research Agenda for Learner Models: For adaptive Educational Digital Learning Environments', in *Innovation via Collaborative Learning in Engineering Education*, vol. 1847, M. E. Auer and P. Toth, Eds, in Lecture Notes in Networks and Systems, vol. 1847. , Cham: Springer Nature Switzerland, 2026, pp. 3–13. doi: 10.1007/978-3-032-18885-4_1.

[11] Society for Learning Analytics Research, '1st International Conference on Learning Analytics and Knowledge (LAK11) – Call for Papers'. [Online]. Available: https://tekri.athabascau.ca/analytics/

[12] P. Long and G. Siemens, 'Penetrating the fog: Analytics in learning and education', *EDUCAUSE Review*, vol. 46, no. 5, Oct. 2011, Accessed: May 26, 2026. [Online]. Available: https://er.educause.edu/articles/2011/9/penetrating-the-fog-analytics-in-learning-and-education

[13] P. Long and G. Siemens, 'Penetrare la nebbia: tecniche di analisi per l'apprendimento', *TD Tecnologie Didattiche*, vol. 22, no. 3, Dec. 2014, doi: 10.17471/2499-4324/195.

[14] 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. , Berlin, Heidelberg: Springer, 2010, pp. 301–322. doi: 10.1007/978-3-642-14363-2_15.

[15] 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.

[16] 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, Mar. 2018, pp. 41–50. doi: 10.1145/3170358.3170409.

[17] 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. , Cham: Springer International Publishing, 2015, pp. 666–669. doi: 10.1007/978-3-319-19773-9_87.

[18] 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, United Kingdom: IEEE, Apr. 2025, pp. 1–10. doi: 10.1109/EDUCON62633.2025.11016496.

[19] *[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, Feb. 2016. [Online]. Available: https://www.iso.org/standard/68976.html

[20] *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

[21] *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

[22] N. P. de Koch, 'Software Engineering for Adaptive Hypermedia Systems', Ludwig-Maximilians-Universität München, München, 2000. [Online]. Available: https://www.pst.ifi.lmu.de/~kochn/PhDThesisNoraKoch.pdf

[23] 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. doi: 10.4018/978-1-6684-7634-5.ch004.

[24] X. Li and S. He, 'Research and Analysis of Student Portrait Based on Campus Big Data', in *International Conference on Big Data Analytics (ICBDA)*, Mar. 2021, pp. 23–27. doi: 10.1109/ICBDA51983.2021.9403070.

[25] 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, Jun. 2018, pp. 88–92. doi: 10.1016/j.procs.2018.04.189.

[26] 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, Sep. 2020, doi: 10.1016/j.compedu.2020.103878.

[27] S. Ouf, M. Abd 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, Jul. 2017, doi: 10.1016/j.chb.2016.08.030.

[28] 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, Feb. 2025, doi: 10.1016/j.compedu.2024.105184.

[29] 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, doi: 10.1109/TLT.2023.3348690.

[30] B. Kitchenham, 'Procedures for Undertaking Systematic Reviews', Keele University (TR/SE0401) and National ICT Australia Ltd. (0400011T.1), Joint Technical Report, Jul. 2004.

[31] B. A. Kitchenham and S. Charters, 'Guidelines for performing Systematic Literature Reviews in Software Engineering', Keele University, EBSE 2007-001, Jul. 2007. [Online]. Available: https://www.elsevier.com/__data/promis_misc/525444systematicreviewsguide.pdf

[32] *1EdTech Learner Information Package Specification*, Standard, Mar. 09, 2001. [Online]. Available: https://site.imsglobal.org/standards/sldm

[33] *[IEEE P1484.2] - Standard for Information Technology — Learning Systems — Learner Model*, PAR IEEE P1484.2, Oct. 09, 1997. Accessed: Aug. 02, 2003. [Online]. Available: http:/ltsc.ieee.org:80/wg2/

[34] 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)*, Aug. 2020, pp. 438–440. doi: 10.1109/ICCSE49874.2020.9201686.

[35] X. Zhao, 'Research on learner model of adaptive learning system', in *2021 16th International Conference on Computer Science & Education (ICCSE)*, Lancaster, United Kingdom: IEEE, Aug. 2021, pp. 906–909. doi: 10.1109/ICCSE51940.2021.9569673.

[36] 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)*, Jun. 2021, pp. 70–75. doi: 10.1109/CSEI51395.2021.9477714.

[37] T. Fei Zhou, Y. Qing Pan, and L. R. Huang, 'Research on Personalized E-Learning Based on Decision Tree and RETE Algorithm', in *2017 International Conference on Computer Systems, Electronics and Control (ICCSEC)*, Dec. 2017, pp. 1392–1396. doi: 10.1109/ICCSEC.2017.8446741.

[38] A. N. Cormack, 'A Data Protection Framework for Learning Analytics', *Learning Analytics*, vol. 3, no. 1, Apr. 2016, doi: 10.18608/jla.2016.31.6.

[39] 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, Nov. 2023, doi: 10.1111/bjet.13388.

[40] 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, Dec. 2018, doi: 10.1186/s41039-018-0086-8.

[41] 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)*, Jerez de la Frontera: Dimitriadis, Yannis; Cerezo, Rebeca; Balderas, Antonio; Martínez-Monés, Alejandra; Spikol, Daniel, Dec. 2018, pp. 1–10.

[42] 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, United Kingdom: ACM Press, 2016, pp. 89–98. doi: 10.1145/2883851.2883893.

[43] J. Doveren, B. Heinemann, and U. Schroeder, 'Towards Guidelines for Data Protection and Privacy in Learning Analytics Implementation', in *Online-Labs in Education*, D. Uckelmann, G. Romagnoli, J. Baalsrud Hauge, and V. Kammerlohr, Eds, Nomos Verlagsgesellschaft mbH & Co. KG, 2022, pp. 45–52. doi: 10.5771/9783957104106-45.

[44] F. Böck, 'Learner Models: Requirements Analysis for Application in Adaptive Learning Environments and Recommendation Systems', in *2025 IEEE Digital Education and MOOCS Conference (DEMOcon)*, Boca Raton, FL, USA: IEEE, Oct. 2025, pp. 1–7. doi: 10.1109/DEMOcon65705.2025.11282849.

[45] 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, Apr. 1984, doi: 10.20684/quality.14.2_147.

[46] *Regulation (EU) 2024/1689 of the European Parliament and of the Council*, vol. PE/24/2024/REV/1. p. 144. [Online]. Available: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=OJ%3AL_202401689

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

[48] F. Ricci, L. Rokach, and B. Shapira, Eds, *Recommender systems handbook*, Third edition. New York, NY: Springer, 2022.

[49] F. Doshi-Velez and B. Kim, 'Towards A Rigorous Science of Interpretable Machine Learning', *arXiv*, 2017, [Online]. Available: https://arxiv.org/abs/1702.08608

[50] C. Rudin, 'Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead', *Nat Mach Intell*, vol. 1, no. 5, pp. 206–215, May 2019, doi: 10.1038/s42256-019-0048-x.

[51] M. T. Ribeiro, S. Singh, and C. Guestrin, '"Why Should I Trust You?": Explaining the Predictions of Any Classifier', in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, San Francisco California USA: ACM, Aug. 2016, pp. 1135–1144. doi: 10.1145/2939672.2939778.

[52] S. M. Lundberg and S.-I. Lee, 'A unified approach to interpreting model predictions', in *Proceedings of the 31st International Conference on Neural Information Processing Systems*, in NIPS'17. Red Hook, NY, USA: Curran Associates Inc., 2017, pp. 4768–4777.

[53] K. Kitto and S. Knight, 'Practical ethics for building learning analytics', *Br J Educ Technol*, vol. 50, no. 6, pp. 2855–2870, Nov. 2019, doi: 10.1111/bjet.12868.

[54] 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, May 2020, doi: 10.1080/02602938.2019.1676396.

[55] 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, Jan. 2016, doi: 10.1007/s40593-015-0090-8.

[56] 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. New York, NY, USA: Association for Computing Machinery, Apr. 2016, pp. 30–39. doi: 10.1145/2883851.2883853.

[57] 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)*, Jul. 2019, pp. 375–379. doi: 10.1109/ICALT.2019.00117.

[58] 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, Jun. 2015, doi: https://doi.org/10.1016/j.chb.2014.07.002.

[59] 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)*, Sep. 2016, pp. 49–53. doi: 10.1109/VLHCC.2016.7739663.

[60] 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. , Berlin, Heidelberg: Springer Berlin Heidelberg, 2010, pp. 301–322. doi: 10.1007/978-3-642-14363-2_15.

[61] 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, May 2022, doi: 10.1016/j.caeai.2022.100069.

[62] 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)*, Sep. 2016, pp. 49–53. doi: 10.1109/VLHCC.2016.7739663.

[63] 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. New York, NY, USA: Association for Computing Machinery, Apr. 2016, pp. 30–39. doi: 10.1145/2883851.2883853.

[64] S. Bull, 'Supporting learning with open learner models', in *Proceedings of fourth Hellenic Conference on Information and Communication Technologies in Education*, Athen, Griechenland, 2004, pp. 47–61.

[65] *[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, Mar. 2011. [Online]. Available: https://www.iso.org/standard/35733.html

[66] *[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, Mar. 2024. [Online]. Available: https://www.iso.org/standard/78175.html

[67] 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.

[68] 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, Nov. 2023, doi: 10.1007/s11623-023-1848-z.

[69] 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.

[70] 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

[71] 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

[72] 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

[73] 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.

[74] M. Csikszentmihalyi, *Flow: the psychology of optimal experience*, 1st ed. New York: Harper & Row, 1990.

[75] M. Csikszentmihalyi, 'The flow experience and its significance for human psychology', in *Optimal Experience*, 1st edn, M. Csikszentmihalyi and I. S. Csikszentmihalyi, Eds, Cambridge University Press, 1988, pp. 15–35. doi: 10.1017/CBO9780511621956.002.

[76] M. A. Guadagnoli and T. D. Lee, 'Challenge Point: A Framework for Conceptualizing the Effects of Various Practice Conditions in Motor Learning', *Journal of Motor Behavior*, vol. 36, no. 2, pp. 212–224, Jun. 2004, doi: 10.3200/JMBR.36.2.212-224.

[77] C. J. Fong, D. J. Zaleski, and J. K. Leach, 'The challenge–skill balance and antecedents of flow: A meta-analytic investigation', *The Journal of Positive Psychology*, vol. 10, no. 5, pp. 425–446, Sep. 2015, doi: 10.1080/17439760.2014.967799.

[78] C. Calliess and M. Ruffert, Eds, *EUV/AEUV: das Verfassungsrecht der Europäischen Union mit Europäischer Grundrechtecharta: Kommentar*, 6. Auflage. München: C.H. Beck, 2022.

[79] F. Böck and D. Landes, 'PLM - an Open Learner Model for Adaptive Learning Environments – Specification', in *2025 IEEE International Conference on Teaching, Assessment, and Learning for Engineering (TALE)*, Macao: IEEE, Dec. 2025, pp. 1–7. doi: 10.1109/TALE66047.2025.11346719.

[80] F. Böck, A. Deuerling, and D. Landes, 'Towards a Personalised Digital Learning Environment of Tomorrow':, in *Proceedings of the 18th International Conference on Computer Supported Education*, Benidorm, Spain: SCITEPRESS - Science and Technology Publications, 2026, pp. 1479–1490. doi: 10.5220/0014738000004021.

[81] F. Böck, 'Learner Models: Assessment Methods - A Literature Analysis', submitted & under review, 2026.

APPENDIX

APPENDIX A

Overview of the most important GDPR Articles

The following overview table serves merely as a guide to help you get started with the most important provisions of the GDPR and to shed light on the highly complex subject of data protection. This working aid does not claim to be fully compliant, as achieving such a standard would exceed the scope of this paper. It is therefore recommended that you involve your data protection officer or in-house data protection authority right from the start and discuss this complex topic with specialist staff.

☑	Question	Consequenses	
		☑	☒
☐	Am I processing **personal data** (name, ID, performance data, learning behaviour, etc.)? Do the data allow a person to be identified directly or indirectly? Can I fully anonymise data if they are to be reused?	Art. 2(1) GDPR (*filing system*), Art. 4(1) GDPR (*personal data*) → GDPR fully applicable	GDPR not applicable
☐	Do I have a **legal basis** for the processing? Can the processing be based on voluntary consent? Is there a legal basis in the applicable national law?	Art. 6(1) GDPR (l*awfulness of processing*) [consent lit. a or public task lit. e]	Processing unlawful
☐	Am I processing **special categories of personal data** (e.g., health, ethnic origin)?	Art. 9 GDPR (p*rocessing of special categories of personal data*) → only with an explicit statutory basis	No additional requirements
☐	Am I using the data for a **purpose other than originally collected**?	Art. 5(1)(b) GDPR (*purpose limitation*), Art. 6(4) GDPR (*purpose of data collection*) → check purpose compatibility	No additional review required
☐	Are the data used for **automated decision-making with legal effect** (e.g., grading)?	Art. 22 GDPR (a*utomated individual decision-making*) → generally prohibited	No restriction under Art. 22
☐	Is there **profiling** within the meaning of GDPR?	Art. 4(4) GDPR (*profiling*), Art. 22 GDPR (*Automated individual decision-making*) → review required	No specific requirements
☐	Am I collecting **more data than strictly necessary**?	Art. 5(1)(c) GDPR (*data minimisation*) → reduce	No action required
☐	Am I storing data **longer than necessary**?	Art. 5(1)(e) GDPR (*storage limitation*) → deletion concept required	Ok
☐	Is there a **deletion policy** (e.g., upon exmatriculation)?	Art. 5(1)(e) GDPR (*storage limitation*)	Create one
☐	Do I inform data subjects **transparently** about processing?	Art. 5(1)(a) GDPR (*lawfulness, fairness and transparency*), Arts. 12 – 14 GDPR (*information obligations)*	Implement information obligations
☐	Can users exercise their **data subject rights** (access, erasure, objection)?	Arts. 15 – 22 GDPR (d*ata subject rights*)	Establish processes
☐	Is a **Data Protection Impact Assessment (DPIA)** required? (high risk, extensive profiling)	Art. 35 GDPR (*data protection impact assessment*)	No DPIA required
☐	Do I use **processors** (cloud providers, external vendors)?	Art. 28 GDPR (*DPA/AV contract required*)	No action required
☐	Are data processed **outside the EU**?	Arts. 45 – 49 GDPR (*Processor*) [review third-country transfer]	Ok
☐	Are data **pseudonymised**, where possible?	Art. 4(5) GDPR (*pseudonymisation*), Art. 25 GDPR (*privacy by design*)	Check if possible
☐	Do I have a **role and access control concept**?	Art. 25 GDPR (*privacy by design*), Art. 32 GDPR (*technical & organisational measures*)	Implement
☐	Are data **encrypted** (at rest & in transit)?	Art. 32 GDPR (*security of processing*)	Add security measures
☐	Do I enable **data portability** (export of learner model)?	Art. 20 GDPR (r*ight to data portability*)	Add function
☐	Was the **Data Protection Officer** involved early?	Art. 38 GDPR (*data protection officer)*	Urgently catch up
☐	Is there a risk of **algorithmic discrimination/bias**?	Art. 5(1)(a) GDPR (*fairness*)	Conduct a fairness assessment

Appendix B

Key Legal Principles (*Brief Overview*)

The most important data protection laws are summarised briefly below to facilitate entry and serve as a checklist to ensure they are not overlooked. This list does not claim to be complete or to cover all casualties, but merely serves as a simplified introduction and quick overview.

☑	**Law**	**Relevant Content**
☐	Art. 5 GDPR	Basic principles (*lawfulness, transparency, purpose limitation, data minimisation, storage limitation, integrity*)
☐	Art. 6 GDPR	Legal basis
☐	Art. 9 GDPR	Special categories of personal data
☐	Art. 12 – 14 GDPR	Information obligations
☐	Art. 15 – 22 GDPR	Rights of data subjects
☐	Art. 22 GDPR	Prohibition of automated decisions with legal effect
☐	Art. 25 GDPR	Privacy by design and by default
☐	Art. 28 GDPR	Processing by a processor
☐	Art. 32 GDPR	Security of processing
☐	Art. 35 GDPR	Data protection impact assessment
☐	Art. 45 et seq. GDPR	Transfer to third countries.
☐	Art. 20 GDPR	Data portability
☐	Art. 8 EU ChFR	Fundamental right to data protection
☐	Art. 14 EU ChFR	Right to education

In addition to the key data protection rights listed above, there are, however, other laws that play a role in conjunction with the aforementioned laws and which may therefore be worth considering in this context:

- EU Data Act
- AI Act
- Art. 14 EU ChFR (Right to education)

APPENDIX C

Key Legal Terms in Data Protection (Brief Overview)

This glossary serves as a reference work. The most important data protection laws are summarised briefly below to facilitate entry and serve as a checklist to ensure they are not overlooked. This list does not claim to be complete or to cover all casualties, but merely serves as a simplified introduction and quick overview.

Term	Explanation
Core Principles of Data Protection Law	
Lawfulness of Processing (Art. 5 (1)(a) GDPR)	Processing of personal data must be based on a valid legal ground under data protection law.
Purpose Limitation Art. 5 (1)(b) GDPR	Personal data must be collected for specified purposes and not used incompatibly with those purposes.
Data Minimisation Art. 5 (1)(c) GDPR	Only the personal data necessary for the intended purpose should be processed.
Transparency Art. 5 (1)(a) GDPR	Data subjects must be clearly informed about how and why their data is processed.
Storage Limitation Art. 5 (1)(e) GDPR	Personal data should be kept only for as long as necessary for the processing purpose.
Integrity and Confidentiality Art. 5 (1)(f) GDPR	Personal data must be protected against unauthorised access, loss, or damage.
Accountability (Art. 5(2) GDPR	Controllers are responsible for and must be able to demonstrate compliance with data protection principles.
Data Portability (*Art. 20 GDPR*)	Individuals have the right to receive their personal data in a structured and transferable format.
Legal Bases and Justifications for Processing	
Lawfulness Requirement (*Art. 5(1), Art. 6 GDPR*)	Personal data may only be processed when a valid legal basis applies, and principles are respected.
Consent as Legal Basis (*Art. 6(1)(a) GDPR*)	Processing is lawful if the data subject has freely given informed and explicit consent.
Public Interest Processing (*Art. 6(1)(e) GDPR*)	Processing is permitted when necessary to perform a task carried out in the public interest or official authority.
Legal Authorisation under Member State Law	National laws may provide specific legal bases that permit certain data processing activities.
Categories of Personal Data	
Personal Data (Art. 4(1) GDPR	Any information relating to an identified or identifiable natural person.
Sensitive Personal Data / Special Categories of Data (*Art. 9 GDPR*)	– racial and ethnic origin, – political opinions, – religious or philosophical beliefs, – trade union membership, – genetic data, – biometric data, – health data, – data concerning sex life or sexual orientation
Inferred Data and Profiling Results	Data derived from analysis or automated evaluation is used to predict or assess personal aspects.
Data Protection Safeguards and Technical-Legal Measures	
Pseudonymisation (Art. 4(5) GDPR	Processing personal data so it cannot be attributed to a person without additional information.
Anonymisation	Irreversibly removing identifiers so that individuals can no longer be identified.
Encryption of Personal Data	Transforming data into a coded format to prevent unauthorised access.
Role-Based Access Control	Restricting system access based on the user’s role and responsibilities.
Data Protection by Design and by Default (*Art. 25 GDPR*)	Systems must integrate data protection measures from the outset and use privacy-friendly default settings.

Term	Explanation
Security of Processing (*Art. 32 GDPR*)	Appropriate technical and organisational measures must ensure a level of security appropriate to the risk.
Rights of Data Subjects	
Right of Access to Personal Data (Art. 15 GDPR)	Individuals can obtain confirmation and a copy of the personal data processed about them.
Right to Object (*Art. 21 GDPR*)	Individuals may object to processing based on specific legal grounds, especially legitimate interests.
Right to Withdraw Consent (Art. 7(3) GDPR	Individuals can revoke previously given consent at any time with future effect.
Right to Data Portability (Art. 20 GDPR)	Individuals can receive and transfer their personal data between controllers.
Compliance and Governance Mechanisms	
Data Protection Impact Assessment (*Art. 35 GDPR*)	A risk assessment is required for high-risk processing to evaluate impacts on data subjects' rights.
Data Processing Agreement (*Art. 28 GDPR*)	A contract that governs processing activities carried out by a processor on behalf of a controller.
Use of Processors (*external service providers*)	Engaging third parties to process data under the controller's instructions and safeguards.
Role of the Data Protection Officer (Art. 37 GDPR)	An independent advisor who monitors compliance with data protection law within an organisation.
Documentation and Privacy Policies	Written records and notices that demonstrate and communicate compliance with data protection rules.
Emerging Legal-Technical Concepts	
Data Trust Model (*trusted intermediary for data governance*)	A governance framework where an intermediary manages data access and use in the interest of stakeholders.