sustainability

MDPI

Article

Privacy and E-Learning: A Pending Task

Marc Alier [1,*], Maria Jose Casañ Guerrero [1], Daniel Amo [2], Charles Severance [3] and David Fonseca [2]

[1] UPC, BCNseer Research Group, Universitat Politecnica de Catalunya, 08034 Barcelona, Spain; mjcasany@essi.upc.edu

[2] GRETEL Research Group, Universitat La Salle, 08022 Barcelona, Spain; daniel.amo@salle.url.edu (D.A.); david.fonseca@salle.url.edu (D.F.)

[3] School of Information, University of Michigan, Ann Arbor, MI 48109, USA; csev@umich.edu

[*] Correspondence: marc.alier@upc.edu

Abstract: Most educational software programs use and gather personal information and metadata from students. Additionally, most of the educational software programs are no longer operated by the learning institutions but are run by third-party agencies. This means that in the decade since 2020, information about students is stored and handled outside premises and control of learning institutions. The personal information about students and their activity while they interact with learning management systems and online learning tools is increasingly in custody of cloud computing platforms, software-as-a-service providers, and learning tool vendors. There is an increasing will to use all the data and metadata from the activity of the students for research, to develop education management strategies, pedagogy approaches, and develop behavior control tools or learning tools informed by behavior analysis from learning analytics. Many times, these studies lack the ethical and moral perspective. In addition, there is an increasing number of cases in which this information has leaked or has been used in a shady way. Additionally, this information will be around for a long time, tied to the future digital profiles of the students whose data has been leaked. This paper hypothesizes that there has been an ongoing process of technological evolution that leads to a loss of control over personal information, which makes it even more difficult to protect user confidentiality and ensuring privacy, that data surveillance has entered the world of education, and that the current legal frameworks are not enough to really protect the student's personal information. The paper analyzes how this situation came to pass, and why this is wrong. We conclude with some proposals to address it from its different root dimensions: technical, cultural, legal, and organizational.

Keywords: student's privacy; learning analytics; technologies in education; LMS; ethical issues

check for updates

Citation: Alier, M.; Casañ Guerrero, M.J.; Amo, D.; Severance, C.; Fonseca, D. Privacy and E-Learning: A Pending Task. *Sustainability* **2021**, *13*, 9206. https://doi.org/10.3390/su13169206

Academic Editor: Dean Vucinic

Received: 31 May 2021
Accepted: 13 August 2021
Published: 17 August 2021

1. Introduction

For the last 25 years, online software has been gradually but steadily introduced in most educational activities. Nowadays, online software tools are being used in education at all levels. Academic management systems (AMSs) handle the enrollment process and the academic record. Learning management systems (LMSs) provide virtual spaces (courses or classrooms) for the teachers and students to perform the learning activities, both as an extension of the physical classroom or as a complete online learning experience. The online learning activities can extend beyond the software provided (or contracted) by the learning institution, to online learning tools provided on the internet. These online learning tools can be integrated with the institution's LMS [1] or be a completely separate application that the teachers or the students have decided to use. Often, teachers and students use general-purpose websites as information sources or applications for learning support; sometimes, this even includes the use of social networks [2].

This has implications regarding the privacy of the students. Forty years ago, students' records were kept in folders inside filing cabinets in schools. Students did not use computers in the classroom to help them study or learn. However, now technological innovation has made available a large variety of apps and software systems to assist education. This innovation has prompted new educational practices and the rise of data-driven education management, pedagogy, and even policy making [3], even when is not clear that this data-driven approach can lead to better pedagogical results [4].

The low cost of storage and transfer, and high computing power available via cloud computing allow this information to be processed and stored on the cloud. This increases the risk of unauthorized data access, unintentional unauthorized disclosure of information, and generation and storage of student information by third parties.

Ultimately, this situation raises concerns among parents and privacy advocates about secondary uses of information collected from their kids in the school/educational environment [5].

For the purposes of this paper, we define "privacy" as the right and ability of a person or group to control the access and use of information about themselves. Some level of access and handling to personal information is necessary to provide services such as education. To respect privacy rights and comply with privacy regulations, those who are in control of personal information, i.e., information whose referent can be identified, need to ensure "confidentiality." Since adequate confidentiality is necessary for ensuring privacy, in this paper we will use only the term "privacy," though in some cases referring to data, the term "confidentiality" would be more precise.

To provide an analysis of the complex issue of privacy in educational technologies, this paper presents the following hypotheses:

Hypothesis 1 (H1). *The evolution of the last 25 years in educational technologies has led to a situation where privacy for students is difficult to ensure and protect;*

Hypothesis 2 (H2). *There are incentives to gather, mine, and exploit personal data and data about the student's activity in online learning tools. These incentives are not always aligned with the best interests of the students and ethical principles;*

Hypothesis 3 (H3). *While there are laws that aim to protect student's privacy, these laws are not on par with technological advances.*

To prove H1, the authors drew on more than 60 years of combined participative observation within the open source communities developing educational software (such as Sakai.org, Moodle.org, TSugui.org, and laptop.org) and learning technologies standards proponents such as the IMS Global Learning Consortium. Part of the historical perspective that addresses this hypothesis has been presented in the podcast https://podcast.learnerprivacy.org/ (accessed on 1 August 2021); the scripts for the podcast were peer-reviewed before its recording and publication.

To explore H2, the authors performed a literature review to look for (a) possible incentives to gather student's personal information, (b) examples of breaches of privacy of student's information, and (c) the issue of privacy within the research field of learning analytics.

We investigated H3 by performing an analysis of the two main laws that affect privacy in e-learning: FERPA in the USA and the GDPR in the EU. We conducted interviews with lawyers specialized in data privacy and with persons with the responsibility of "data protection officer" in universities to discuss laws, their implications for learning institutions, and current technologies. We present the conclusions of this research.

We conclude with a proposal of actions to take in the future about the issue.

2. Privacy in Educational Technologies: A Historical Perspective

2.1. A Beginning in Good Standing

We find the first developments of the learning management system (LMS) in the first years of existence of the Web. The first systems were custom built in-house, in languages such as Perl or C and plain HTML. As the concept of LMS was refined, the first commercial products hit the market during the com bubble. The usage of LMS started spreading to most universities and schools.

By the year 2000, most educational institutions were using or experimenting with LMS. LMS programs were used as a complement for traditional presence learning, in fully virtual learning environments (VLEs) or virtual campuses for online learning (e-learning) or in mixed contexts (blended learning).

In this initial stage, LMS implemented really good privacy for the student's information. The cause was that all those LMS technologies were usually running on data centers within campuses and operated by technicians from the institution. Additionally, this was true for in-house-developed LMS and for LMS software licensed from companies such as IBM, Blackboard, and WebCT.

All the data about and data generated by the students (including activity logs) were stored in the institution premises, similar to a few years before when all records were kept on paper. Therefore, privacy was not an issue to consider, yet.

2.2. LMS Migrations, 2000–07

Running an LMS is expensive. Back in 2000, it was very expensive. For starters, running applications on ones' own data center is expensive because of the cost of the infrastructure and hardware. Moreover, the skilled personnel also needed to develop it (if needed), customize it, and manage its operation.

Every year, the LMS had to be updated—usually, the IT departments received a stack of CDs with updates for the LMS. Additionally, the whole system had to be updated; all the customizations, functionalities tested, and the performance of the system needed to be checked. Scalability was—and still is—a significant issue in the update processes.

As to the sheer amount of work it took, most installations would not update to the newer versions every year. Needless to say, this was a huge security risk.

Due to the high costs of updating the LMS version, the option of migrating to a new LMS was a realistic one for a lot of institutions. Given the cost of updating, and the rapid evolution of web technologies—languages, frameworks, and platforms—changing the LMS was often considered a better and easier choice than upgrading to the new version of the installed LMS, especially since vendors offered technical support with the migration. In the case of the open source LMS Moodle, a worldwide network of accredited partners offered such technical support and training.

By 2007, most LMS vendors and open source projects started addressing interoperability features to facilitate the integration of other systems and automate processes such as enrollment and sending learning outcomes to AMS.

This might explain why in the early 2000s, LMS such as Moodle (and to a lesser extend Sakai) in the realm of open source applications, and Blackboard (as a private vendor) started gaining market share. Many institutions switched their old LMS products to the point that these LMS vendors gained most of the market of LMS in learning institutions by 2010 [6].

The Open Knowledge Initiative (OKI) made a serious attempt to propose a set of standard interfaces (Open Source Interface Definitions—OSIDs) [7]. Although these standards never fully caught up as such, according to Alier and Casany [8], OKI's OSIDs had a strong influence on the design of the Moodle Webservices API, released in Moodle 2.0. This showcases a movement of a major actor in the LMS market to become more attractive for potential switchers and easy to maintain from version to version.

2.3. Elearning Interoperability Standards, 2007–2010

In the previous period, every popular LMS had developed its own plug-in mechanism. Usually, these were created ad hoc, with few and low-quality documentation, required a high level of experience and inner knowledge of the developer culture of the particular LMS, and had a steep learning curve. A developer who wanted to create a custom feature had to build it and maintain it specifically for each target LMS.

The IMS Global Learning Consortium succeeded in the task of bringing together the main actors in the e-learning sector to draft interoperability specifications that would become established industry standards.

IMS Common Cartridge (IMS CC) provides a specification for creating and sharing educational online interactive multimedia content. The specification describes in detail the packaging format and infrastructure needed to support it and the methods for presenting it to the end user [9]. The SCORM specification by ADL competes with IMS CC.

IMS Learning Tools Interoperability (LTI) is the industry-established standard for software-as-a-service integration of online learning applications. IMS LTI allows the integration of web-based applications for learning within the context of one activity in a course of an LMS. The students can access the learning web application by clicking on a link. Meanwhile, the LMS provides authentication, authorization, and academic context, that is identification of the course, activity, and role of the user to the learning tool provider [10].

LTI is more relevant for the present paper because it sets up the requirements to implement a model of software as a service (SaaS) of learning tool vendors, providing services to learning institutions.

Sakai had a limited market share but exerted an influence on the other LMS providing a reference implementation of IMS LTI, being followed by all the relevant actors. Starting with the leading open source LMS community, Moodle coordinated with the Sakai developers and the standard proponents to develop viable reduced versions of the LTI specification before its official release. Therefore, when the specification was finally released officially, Moodle, Sakai, and several other learning application providers offered tools that already supported it [11,12].

By 2009, there were two major high-quality open source LMS (Moodle and Sakai), solid interoperability standards for e-learning (IMS CC, IMS LTI, SCORRM), and the data generated in the LMS were still managed on the campus data centers; thus, privacy was reasonably respected and protected.

2.4. LMS, Software as a Service and Cloud Computing, 2010–2020

Amazon Web Services (AWS) was launched in 2006 with AWS EC2 (Elastic Cloud Computing) offering Linux Virtual Server (2006) on demand, file storage with the AWS S3 (Simple Storage Service) (2006), and ElastiCache (2011) [13].

AWS and other cloud computing platforms that later emerged solved the problem of hardware scalability and availability, offering the possibility to launch and stop huge servers on demand, instantly at a low hourly cost.

Cloud computing provided hardware availability but did not solve the problem of managing, maintaining, and updating the software. IT departments were stuck with the same LMS software architectures and had to deal with the complexity of legacy platforms such as ORACLE.

Moreover, until 2018, using AWS was complex and IT personnel had to be trained extensively. However, this is no longer true. In 2018, the platform became simplified, many critical tasks were automated and new software development kits (SDKs) and frameworks became totally oriented to cloud deployment. Today, a single developer is able to run servers at production at scale as a side project, a feature that was impossible in 2015.

Then, things started to change. Two trends happened in the late 2000s and early 2010s. Universities started to move servers to the cloud. This started to be considered a best practice, and having the LMS under the SaaS modality was an attractive option.

Canvas was born when cloud computing was booming. Canvas implemented good interoperability and simplicity to outsource everything. Due to IMS LTI and other interoperability standards, it became possible to use a Canvas-based LMS provided as SaaS while still being able to use preferred learning tools from Blackboard, Sakai, or Moodle since, for example, a Moodle learning activity or even a course can be provided as an IMS LTI tool.

Companies such as Blackboard, Desire2learn, Moodle, and Sakai did not react to the arrival of cloud computing nearly as fast as Canvas did. Canvas gained traction and was able to capture about 40% of the market share [14].

The proposal of LMS as a service means "no worries for the rest of your days." It meant that the university IT did not need to have any specific skills or knowledge to run the LMS since it was provided as a service. They only required user-level and admin-level skills where necessary.

However, the outsourcing of in campus online services does not stop here. Google and Microsoft have been offering their SaaS suites to education institutions, with integrated email, calendar, cloud drive storage, office apps, online conferencing, and LMS such as features. They offer the basic package for free. Their business is not the service but the data and metadata.

As is often said, "When the service is free, you (the user) are the product."

The outsourcing of the operations, maintenance, and improvement of the LMS to a SaaS cloud provider led to a loss of IT developer talent. They did not have interesting tasks to perform. They lost the ability to affect the LMS to meet faculty needs.

To summarize, in the early 2000s, the learning institution operated and kept the information about learners in their infrastructure, just as when everything was stored in paper files. The situation did not present any problem regarding privacy. However, this has changed. Over the last 20 years, three technical innovations have transformed the LMS design, architecture, business model, and operations. These innovations are interoperability standards for learning tools and contents, cloud computing, and software as a service.

As a consequence, in 2020, some universities do not have any learner data stored in servers that belong to the institution. This means the learner data and metadata have broken out of the control of the institution. Additionally, this may create potential issues with learner data privacy.

3. Students under Surveillance

3.1. Surveillance Capitalism Enters Education

The aim to obtain and analyze as much data as possible of the user's activity in order to gain the ability to influence their behavior is not exclusive to the usual suspects from big tech (Google, Amazon, Microsoft, Facebook, Apple, Tencent, Alibaba, and Palantir). This same purpose can be found in the big trends in research and innovation in education: learning analytics, gamification, adaptative learning, and proctoring [15].

According to Siemens, learning analytics is the use of data, learner-produced data, and analysis models to discover information and social connections for predicting and advising people's learning [16]. This definition has been debated, criticized, and modified by the research community because often students and are minors and the analysis of data may provide the educational institution the ability to modify their behavior by advising their learning. This definition is not precise enough; teachers are also under observation since managers have intentions of advising the teachers teaching.

The application of game-design elements and game principles in non-game contexts has been named Gamification. It encompasses the design of a set of activities and processes to solve problems to organizational productivity, learning, improve user engagement, knowledge retention, exercise, employee recruitment, evaluation, etc. [17]. Gamification is used in education and is a strong research field in educational innovation [18]. Additionally, we have the elements of environmental design all over again to nudge users, now students, to behave in the way the designer desires.

Adaptive learning or adaptive teaching is a research field of innovation in education that aims to use computer algorithms and machine learning—trained by students' data— to improve the learning experience and deliver customized resources and learning activities to address the unique needs of each student. Computers adapt the presentation of educational material according to students' learning needs, as indicated by their responses to questions, tasks, and experiences. The technology combines aspects derived from various fields of study including pedagogy and psychology—namely, those related to the constructivism school—computer science, machine learning, psychometrics, and even brain science [19].

The authors encountered, in conferences and while reviewing research papers for journals, several presentations of research projects of learning analytics, gamification, adaptive learning, proctoring or analysis of student's emotions from video surveillance, where personal information, video, audio, and activity logs were used, and no ethical considerations were disclosed.

Included in the framework of surveillance capitalism [20], education innovation research about learning analytics, gamification, adaptive learning, and other emerging trends seem to fit in well. The aim of this research is about obtaining as much data about the learner, creating models of behavior, and changing the environment to influence the behavior of the learner with increasing leverage. Many times, this is accomplished without considering the consequences this might have on the learner.

However, in recent years, there is an ongoing academic debate on what actions should be taken to address some of the issues of the application of learning analytics, such as ethics, privacy, and legal aspects. According to Drachsler and Greller, there are fears towards learning analytics concerning (1) privacy and digital identity issues, (2) the asymmetrical power relationship between the data controller and the data subject, (3) the ownership of the data extracted from the student's activity, (4) data integrity concerns, (5) the security of collected data and risk of hacking, and (6) the problem of anonymity [21].

3.2. Examples of Surveillance in Education

As we mentioned before, there are incentives in place to obtain as much data as possible about students and their interactions with learning tools. In this section, we provide several examples of possible misuse of student's data.

The examples are grouped into the following groups: educational apps that collect, extract, and may use student's data; apps used to track students, mostly from higher education institutions; the third group are examples of the use of big data to predict student's enrollment; the fourth group are facial recognition systems; lastly, systems that collect data to perform some kind of e-advertising.

3.2.1. Educational Apps That Collect, Extract, and May Use Student's Data

The inBloom foundation, the Snappet, or Classdojo are some of the most known examples.

The "nonprofit foundation" inBloom, which operated a free software-as-a-service application to manage student data for public school districts across the EUA, created a database with more than 400 different data fields about students that school administrators could fill in. The system was meant to extract student data from school grading and attendance databases, store it in the cloud, and send it to dashboards where teachers could track the progress of individual students effectively. Some school administrators collected sensible data such as family relationships (i.e., foster parent) and reasons for enrollment changes (i.e., leaving school as a victim of a serious violent incident) [22]. The inBloom analytics system was closed down in April 2014 after parents and pressure groups expressed sincere concerns about the misuse of data, the repurposing of data for commercial interests, as well as general safety from cyberattack.

The second example is the Snappet software. Snappet is an organization that rents out tablets to primary schools, with built-in educational software. The tablets are aimed at children from 7 to 9 years of age. Children can read and practice material on the Snappet tablets for subjects such as language, spelling, arithmetic, and reading. At the beginning of July 2014, the tablets were used at more than 400 primary schools in the Netherlands. Snappet collected data from student's apps for processing it afterward [23].

The Dutch Data Protection Authority (CBP) has investigated the processing of personal data by the Snappet Foundation. The CBP has expressed its concern about the very detailed data obtained from seven-to-nine-year-old children because it is sensitive personal data.

In the investigation, the Dutch DPA determined that Snappet processes personal data for different purposes such as advising (or aiming to) advise the schools on possible individuals learning difficulties on the basis of individual learning achievements such as posing an early diagnosis of dyslexia or analyzing the individual student data to classify students into skill levels, by comparing these data with the results per statements of the other children in the class and the results per statements of all other children using the tablets of schools that have agreed to purpose.

The CBP considers that schools are incorrectly informed by Snappet about the overviews of results because of the claim they do not contain personal data. They are also concerned by the lack of information about essential elements of the data processing because the schools cannot determine the purpose and means of data processing, and thus, they do not exercise control over the data [23].

The Dutch DPA found a number of violations in its investigation such as releasing children's personal data to Snappet merely by signing an agreement. This agreement was just signed by the schools. Another aspect was the fact that schools choose to be a part of this experience that creates possibilities to aggregate, combine, and process further personal data without the parents' consent. This partly concerns sensitive personal data, which leads to all kinds of conclusions that can be connected with consequences in (later) social life [23].

Another example of student data collection is ClassDojo. It allows teachers to reward the children with "dojo points" for their performance, and it became an educational success story, used by over 3 million teachers and 35 million children from 180 countries globally.

However, ClassDojo raises dire privacy concerns. In 2016, its Silicon Valley team received over USD 20 million in venture capital investment to extend into a "school-wide" platform, a tool for connecting parents and the school plus leaving the handling of classroom pictures, messages, videos, and digital portfolios of children's in the class dojo system.

The teachers are unknowingly producing and refining huge datasets about children's behavior. This raises some concerns, including the following:

1. Not every school seeks "informed consent" from parents to enter their children's data into the ClassDojo system. ClassDojo's can be used to create a persistent behavioral record of each child across the duration of their schooling, and school managers can use these records to identify children by their behavioral profile. ClassDojo is already in partnership with Stanford University, which is using ClassDojo data to evaluate how well its content promotes children's psychological development [24].
2. The use of ClassDojo in classrooms impacts teacher–pupil contact time; with points awarded by clicking on the mobile app, teachers become responsible for data entry rather than interacting with pupils. Additionally, now is a time when children's mental health has become a subject of serious concern. In this context, ClassDojo might reinforce the idea that it is the behavioral mindset of the child that needs to be corrected. The competition to be the firsts in a ClassDojo ranking (according to their accumulated dojo points) could easily become a further source of stress. In an attempt to monetize the service, ClassDojo is proposing "premium features" for parents and schools, although its vast databank also has potential for monetization. School managers might purchase reports to single out children for specific classes or special behavior programs. Local government departments could buy the data to compare schools' performance [24].

3.2.2. Apps to Track Students

In the context of the COVID-19 pandemic, we found the Aura system, an initiative to control the spread of COVID-19. The Albion College (Michigan) requested all students to install an app that will track student's live locations at all times. Unfortunately, researchers found out that the app had two major vulnerabilities that could expose students' personal and health data [25].

The University of Alabama was using a location-tracking app to detect students who leave football games early. Afterward, these students receive punitive measures for leaving earlier. They used an app to track the location of their students [26]. One of the problems was that a publicly funded university used public funds to purchase an app that tracks students' location. The app can be deleted at any time and only tracks students while they are in the stadium, but when a student leaves the stadium early, the app sends the student to the back of the line for tickets to important games [27]. In this case, the surveillance was unusual and ultra-specific. It was motivated by one of the most powerful football coaches in the nation.

Other schools are, or were, experimenting with more pervasive tracking tied to class attendance [28]. Short-range phone sensors and campus-wide WiFi networks enable colleges across the United States to track hundreds of thousands of students more precisely than ever before. The schools rely on networks of Bluetooth transmitters and wireless access points to piece together students' movements from dorm to desk. One example is Syracuse University that has used the SpotterEDU app to track student's location in order to control class attendance. One professor from this university explained that his lecture had never been so full since they started using the tracking software.

School administrators obtain data that provide comprehensive tracking of student movements. The focus is classroom attendance, which is assisted by hundreds of tiny electronic hall monitors. This tool allows instructors to be notified of missing students so they can send text messages or emails to their phones, hoping they will go back to class.

3.2.3. Big Data to Predict Student's Enrolment

One case of misuse of students' data is the practice of some colleges or universities to track students [29–31] to predict their future enrollment in educational institutions. Many college admissions offices in the USA engage in very sophisticated data-gathering efforts to try to predict the behavior of students in the process of choosing a college. In the USA, admissions offices use permission-based marketing because they have received permission to solicit and communicate with students. Next, college admissions offices collect data with the purpose of trying to determine a good match, which has a mutual benefit for colleges and students. Finally, college admissions offices store and use data during the recruitment process.

However, college admissions offices identify, solicit, collect, maintain, and analyze all sorts of data to recruit, admit and enroll students. College admissions offices spend a large amount of money, use considerable human resources, and rely on big data to help them perform their job.

The fact is that universities are collecting more data about prospective students than ever before, according to school administrators, to help better predict which students are most likely to apply, accept an offer, and enroll [31].

Pressures for admissions officers to meet target enrollment numbers have led some HEIs to consider big data as their solution. According to Rivard's reporting on the use of data analytics in admissions, recruiters are analyzing personal information of potential applicants in order to "target them for certain traits," including income and ethnicity. Furthermore, admissions professionals purchase and analyze datasets sold by the National Research Center for College and University Admissions, the College Board, or by ACT, in order to develop predictive algorithms to score whether or not a specific student is likely to enroll given her or his profile information. In aggregate, administrators manage large datasets that include millions of student names and identifiable information [32].

One example of the prediction model based on the analysis of large datasets is the case of Houston Baptist University. The college administrators discovered that its models based on big data analysis successfully predicted which students would enroll, regardless of whether or not they received viewbooks and mailers. Mailers and expensive viewbooks were the traditional way of marketing to students used years ago. Wichita State University used admissions analytics to avoid hiring admissions consultants. Their analytics predicted "high-yield" (i.e., likely to enroll) students better than their consultants (96% success rate for the algorithm, 82% for the consultants). Augustana College tracks every communication with a potential student, including e-mail, Twitter, and Facebook, and scores these messages to rate the student's "demonstrated interest" in the institution.

3.2.4. Facial Recognition Systems

The Duke Study from Duke University is one example of recorded thousands of student's faces. In 2014, thousands of students were walking around campus, going to and from their classes, minding their own business. What they did not know was that on a particular day, Duke researchers were recording them and putting their likenesses into a data set. This dataset was placed on a public website, and it could be downloaded by academics, security contractors, and military researchers around the globe [33].

In another similar example, a 2014 project by Stanford researchers called "Brainwash," used a camera to take more than 10,000 images from students over 3 days. The data from Brainwash was then shared by the researchers with third parties such as academics in China associated with the National University of Defense Technology and the surveillance technology firm Megvii [34].

In China, a Chinese facial recognition database with information on thousands of children was stored without the necessary security measures on the Internet. A researcher discovered it, raising questions about school surveillance and cybersecurity in China [35].

3.2.5. E-Advertising in Education

Finally, e-advertising has arrived in higher education. Now, advisors can base their guidance on individual students based on insights gleaned from bid data. There are e-advising systems that analyze a student's profile in comparison with her peers in order to evaluate her current and predicted rate of success in particular academic programs. These systems provide instructors with dashboards so they can consult student's "electronic reputation" and academic history [36,37].

Some e-advising tools, such as those employed by Austin Peay State University and Arizona State University, create "personalized degree paths," generating automated course recommendations to students based on their academic and professional goals, courses they need to graduate on time, and courses in which students are predicted to be academically successful. These systems have questionable functionalities such as denying students access to particular courses unless they take specific actions or receive advisor approval when they predicted a low rate of success in a course [38].

To summarize, the migration of student's data to the cloud and the large storage capabilities of cloud-based systems create new concerns about unauthorized access to, or unintentional disclosure of, student information. Many people, including parents or students themselves, fear that the data could be stolen, or that it may compromise the students by providing personal information to possible predators.

Other worries among parents and students include the publication of sensitive information through human or technological errors [39]. Many fear that permanent records limit students' future opportunities based on outdated, inaccurate, or irrelevant information [40].

3.3. Ethics, Privacy, and Learning Analytics

As learning analytics uses student data collection to measure and analyze learning processes, it is necessary to discuss the ethical issues it might raise. Pardo and Siemens define ethics as "the systematization of correct and incorrect behavior in virtual spaces according to all stakeholders" [41].

Pardo and Siemens identified four principles to categorize the numerous issues concerning data privacy and ethics of learning analytics: (1) transparency, (2) student control over the data, (3) security and accountability, and (4) assessment [41].

Transparency means that all stakeholder groups should be informed about when, how, and what type of data is collected, stored, and processed. Student control over the data points out the right of users to access and correct the data obtained about them. Institutions should ensure data security to avoid users' highly sensitive data being exposed. Accountability refers to the identification of responsible entities, and assessment refers to the constant evaluation, revision, and refinement of data collection, security, transparency, and accountability.

Prisloo and Slade use a socio-critical approach based on being critically aware of the way our cultural, political, social, physical, and economic contexts and power relationships shape our responses to the ethical dilemmas and issues in learning analytics [42]. Slade and Prinsloo propose a number of questions from which institutions can develop guidelines and policy frameworks [43]. These questions are as follows: (1) Who benefits and under what conditions? (2) What are the conditions for consent, de-identification, and opting out? (3) What vulnerabilities and possible harms arise from data leaks and bad usages? (4) How are data collected, analyzed, accessed, and stored?

Drechsler proposed the DELICATE checklist, introducing an eight-point checklist named DELICATE that can be applied by researchers, policymakers, and institutional managers to facilitate a trusted implementation of learning analytics [21].

1. Determination: decide on the purpose of learning analytics for your institution;
2. Explain: define the scope of data collection and usage;
3. Legitimate: explain how you operate within the legal frameworks, referring to the essential legislation;
4. Involve: talk to stakeholders and give assurances about the data distribution and use;
5. Consent: seek consent through clear consent questions;
6. Anonymize: de-identify individuals as much as possible;
7. Technical aspects: monitor who has access to data, especially in areas with high staff turnover;
8. External partners: make sure externals provide the highest data security standards.

The DELICATE checklist shows the complexity of the problem. Each point raises a number of difficult questions, and the answers may point out that there is no possibility of agreement between all the actors or stakeholders involved. In the best of cases, when all stakeholders reach an agreement on all the eight points in the DELICATE checklist, there are no technical solutions that will enforce agreements, bringing all parts to the beginning of the process.

According to Willis, the intersection of big data analytics by college administration and ethical reflection is best examined within the obligation of knowing paradigm [44]. The real problem here is that statistical probability within a matrix of academic prediction can have massive consequences for institutions and individual students alike. He points out that there has not been enough ethical reflection on these issues to date. Similar to other forms of technology, the development of software platforms and predictive analytics evolve so quickly that they outpace the ethical issues and the time it takes to consider even the smallest implication. Willis proposes to use the Potter Box, a popular ethical model in business communications to analyze ethical issues in learning analytics.

4. Legal Issues

In this section, we explore the two most known regulations that affect student data: FERPA that regulates specifically student's information in the USA, and the EU's general-purpose data regulation GDPR. The purpose of this analysis is to identify strong points and limitations in these regulations and their implementation. To gather insight into the laws and their implications, the authors conducted interviews with lawyers, data privacy officers in several universities, and participated in discussions in open source communities and standards organizations regarding these laws and the technical and organizational requirements they present. The interviews are documented in one of the authors' PhD Thesis [45].

4.1. FERPA

The Family Educational Rights and Privacy Act (FERPA) was originally enacted by the US Department of Education in 1974, setting a legal framework that granted parents and tutors access to the records kept by schools and districts, and to require their written permission before identifiable information was disclosed to others, addressing privacy concerns.

Specifically, parents and students over 18 years old or enrolled in a post-secondary education institution have the following rights regarding personally identifiable information (PII) maintained in a student's education record: (1) verifying the accuracy of the record; (2) challenging the accuracy of the record at a hearing and provide correction or commentary; (3) preventing information in the student's record from being disclosed to a third party without written consent.

However, there are exceptions that allow educational entities to share information without requiring the parent's or student's consent, for example, exceptions that enable educational institutions to comply with judicial and executive requirements and respond quickly when students' health or security is at risk. Some exceptions also allow educational institutions to share student's information with third parties that perform services, conduct studies, or facilitate evaluations upon request [45].

The school official exception is particularly controversial because it allows schools to disclose information without consent to outside parties who provide learning apps, email support, and learning management systems (LMSs) infrastructure.

Some of the most significant limitations of FERPA mentioned by Zeide [46,47] are the following:

1. Disclosure under the school official exception is informal—FERPA neither specifies how schools decide who is an authorized data recipient nor how to document such authorization or its scope. FERPA does not require a specification of the purposes served by disclosure or a threshold of applicability.
2. Broad discretion over security and approval of data recipients—FERPA requires minimal oversight of data recipients or security requirements. Educational institutions should have "direct control" over third parties that access data. However, while it is suggested that schools control this feature with contracts, it is not a requirement. The standards for "direct control" are loosely defined in non-binding guidance.
3. FERPA lets schools define under their own criteria what constitutes a "legitimate educational interest" required to share information with a school official. In practice, the bounds of what constitutes an appropriate "school official" data recipient and "legitimate educational interest" are not clearly defined.
4. Compliance-oriented enforcement—When a privacy issue is detected, the Family Policy Compliance Office (FPCO) of the Department of Education notifies the institution, which then has "a reasonable period of time" to comply voluntarily with its FERPA obligations. If the entity does not comply, the FPCO can initiate "any legally available enforcement action" to compel compliance. At a practical level, this limits enforcement to the unlikely case of an educational institution intentionally and repeatedly violating FERPA after FPCO attempts to bring it into compliance.
5. Limited regulatory scope—FERPA only applies to educational agencies or institutions that receive federal funds. It does not apply to the data recipients or to entities, such as Massive Open Online Courses (MOOCs) that collect and use information about students and do not receive federal funding. If a data recipient violates FERPA, the disclosing school is responsible for non-compliance with the law. In this case, the DOE can prohibit a publicly funded institution or agency from providing information to an entity found in violation of FERPA for at least five years. No punitive action is taken against private institutions.

4.2. GDPR

The European Union's (EU) General Data Protection Regulation (GDPR) was approved in April 2016 and enforced in all EU member states on 25 May 2018. The main goals of the GDPR are to protect the personal data of EU citizens, to enable full informed user consent, and to raise trust in personal data treatment while using digital services. In terms of user consent, previous to GDPR, the EU enacted the ePrivacy Directive in 2009 requiring third parties to enable ". . . methods of providing information and offering the right to refuse . . . as user-friendly as possible," to avoid websites to track users through unnecessary cookies to run services without user consent.

The GDPR reinforces restrictions on the treatment of personal data, especially the user consent process, in the following:

1. The Seventh Article "Conditions for Consent";
2. The 32nd Recital, specifying that "Consent should be given by a clear affirmative act establishing a freely given, specific, informed and unambiguous indication of the data subject's agreement to the processing of personal data relating to him or her, such as by a written statement, including by electronic means, or an oral statement";
3. The 42nd Recital, specifying that "For consent to be informed, the data subject should be aware at least of the identity of the controller and the purposes of the processing for which the personal data are intended";
4. The 43rd Recital, stating "Consent is presumed not to be freely given if it does not allow separate consent to be given to different personal data processing operations."

Fields related to data collection, storage, and analytics continue to evolve, growing exponentially in their capabilities of data and metadata gathering and processing to infer knowledge about learners and citizens, integrating technologies such as cloud computing, data mining, big data, and machine learning.

The EU regulator's concerns about possible negative consequences of personal data treatment and analysis include Article 25 "Data protection by design and by default" in the GDPR. It introduces two principles—(1) privacy by design and (2) privacy by default—that enforce privacy measures at the technical and organizational levels. This encompasses measures such as pseudonymization at the beginning of any software design, "for ensuring that, by default, only personal data which are necessary for each specific purpose of the processing are processed."

The GDPR grants citizens with a set of rights over their personal information, (GDPR, Recital 156)—right to rectification, right to erasure, right to be forgotten, right to restriction of processing, right to data portability, and right to object to the processing personal data for archiving purposes in the public interest, scientific or historical research purposes or statistical purposes, and automated individual decision making. The recipients of these rights include the students and this sets educational institutions on the spot of having to comply with these clauses.

Nevertheless, the GDPR has significant limitations, listed as follows:

1. Issues with concepts—To exercise one's rights granted by the GDPR, one needs to understand some concepts about data privacy and the implications of its use. Research shows that "students are not aware of the use of their data and metadata by others" and "do not know who can access them (their data), whether they are used for unwanted purposes." Therefore, data privacy needs to be among the contents to learn in the curriculum if the students are to be able to exercise their GDPR rights [48].
2. Threats to individual control—GDPR gives control to users to consent to data sharing and treatment. However, there are some threats to user consent, such as cookie acceptance, that diminishes user criteria to evaluate consent or automatically consent by information overload and information complexity. The GPDR fails to avoid these pitfalls with the right to explanation or the use of icons to simplify complex concepts. Icons can only provide a partial description of data treatment and processing, and explanation falls short to explain actual implications for an individual [49].

3. The GDPR introduces in Article 25 the principle of data protection by design and by default. Most educational software does not comply with this principle. This means that many codes must be redesigned and refactored with data protection in mind by design and by default.
4. The directive is designed to act in a punitive manner when a privacy breach is denounced. However, it does not require standards to be attained nor any measure of technical and organizational certification or quality assurance with regard to privacy.

5. Discussion

Thus far, in this paper, we have seen that the ongoing process of digital transformation of the educational system worldwide is encountering problems in data privacy. Our research suggests that the following three initial hypotheses have some truth to them:

Hypothesis 4 (H4). *The evolution of the last 25 years in educational technologies has led to a situation where privacy for students is difficult to ensure and protect;*

Hypothesis 5 (H5). *There are incentives to gather, mine, and exploit personal data and data about the student's activity in online learning tools. These incentives are not always aligned with the best interests of the students and ethical principles;*

Hypothesis 6 (H6). *While there are laws that aim to protect student's privacy, these laws are not on par with technological advances.*

H4 implies that there is an ongoing process that leads to a loss of control over personal information, which makes it even more difficult to protect user confidentiality and ensure privacy. The technological and organizational evolution has led to a prevalent model of SaaS for learning online tools and hosting on public cloud platforms of LMS and AMS. As a consequence, the student's personal information has moved outside the campus and the learning institution. The practice of automatic and by default gathering huge amounts of information about the student's activity has spread. All of this information can be used for a number of purposes with positive and negative impacts on students' lives during their time as students and afterward.

H5 indicates that data surveillance has entered the world of education. We have seen that institutions, companies, and researchers are eager to use this information and that the ethical concerns about first-, second-, and third-order consequences of these actions are just starting to arise in recent years. Connected with this issue is the growing interest in ethics in the sector of information technologies, where we can see universities, such as the employers of the authors, governments, companies, and professional associations put together ethical committees, and new books and documentaries are being published [20].

H6 suggests that the current legal frameworks are not enough to really protect the student's personal information. In his book *Being Digital*, Nicolas Negroponte already observed that "The combined forces of technology and human nature will ultimately take a stronger hand in plurality than any laws Congress can invent" [50]. We can detect, in the adaptations to GDPR of the LMS, that the students are presented with forms to accept GDPR compliance terms of service, similar to cookies and websites. However, personal information is still gathered, mined, analyzed, and stored in unsecured systems, attracting hackers and information thieves since there is a market for such data.

6. Conclusions

These challenges we currently face need to be addressed at different levels: social and cultural, organizational, and technical.

First of all, we need to solve at a social and cultural level the ethical dilemma presented by data privacy. On the one hand, there are the possible benefits of the usage of personal data and metadata fed to big data, analytics, and machine learning systems. On the other hand, there are the downsides of forfeiting citizens' personal privacy and providing unrestricted massive access to personal data to government and corporations. Different societies will have different views and priorities. It is obvious that western liberal democracies might have a different position on the issue than the People's Republic of China.

It is important to have a discussion at the social and cultural level. It starts with education about privacy and raising awareness of the dilemma of data privacy. In recent years, we have seen a proliferation of books, documentaries, fiction movies, and series, which signal that western culture is raising its awareness on the privacy problem.

We, the people, need to solve this ethical dilemma have an informed social debate about what are the accepted uses of personal information, online activity records, and surveillance data. When this issue is settled in the culture, the proper laws will follow, fueled by popular pressure on the politicians.

Second, on the educational institution (organization) level, we need to develop best practices for privacy and include these practices in the business processes and training. We need to create quality standards to measure how the privacy best practices are followed. These quality standards can be certified to ensure the compliance of privacy best practices when outsourcing to third parties. Then, we will truly comply with the principle of data privacy and by default.

Last but not least, the problem needs solutions at a technical level. We are dealing with a complex problem with many legacy systems, codes, and standards in place. Additionally, we have also to beware of partially effective solutions such as data depersonalization in learning analytics data warehouses or plain bad ideas such as Google's Federated Learning from Cohorts [51]. Security researchers have shown time and again that depersonalized data can be accurately be re-personalized with very few data points [52].

After several years of research, the authors propose the following preliminary list of technical issues and possible solutions in the context of LMS, AMS, and learning tools:

- Encryption of personal data on the server datastores—The personal information of the students is stored plain and unencrypted in many database systems. Any superuser, developer, sysadmin, or hacker who made it into the system has full access to it. This is a complex technical problem because many legacy codes and systems access these datasets, and we have also to address performance and scalability issues. A data storage system such as the "personal data broker" could be used to encrypt sensitive data in the LMS. The authors developed a prototype running on Moodle [53].
- Apply differential privacy techniques to the data logs—The LMS usually logs all the activity in the system. Every action every user (student, teacher, admin) has performed is recorded with a unique identifier for every user, which can be easily traced to the user identity. These logs feed learning analytics systems and are unencrypted of course. These logs should be anonymized, and differential privacy techniques should be applied when recording these logs, inserting noise, which would prevent the depersonalization of the information while allowing for statistical inferences to those researchers who are entrusted with the noise pattern [54].
- Masking the student's identity under an alias—If students wish, for whatever justified reason, to make use of their right to object to this kind of exposure of their data, they would not be able to use the system. The current GDPR compliance implementations of most systems require the acceptance of terms of use to access the LMS. Therefore, this right is violated. Let us point out that the personal information of the student is not only accessed by academic staff. LMS programs are designed for interaction between students and teachers. The students gain access to a lot of personal information of their peers: access to course rosters, fellow students' profiles, forum posts, wiki edits, etc. The authors developed a Moodle plugin prototype that enables the students who want or need to exercise their right to oppose to not lose their right to education, by enabling a system of alias profiles. Students can show themselves under alias identities to their peers [55].
- We need to establish privacy practices for the learning tools that interoperate with LMS. The privacy features present in protocols such as IMS LTI need to be enabled

in the default configurations and strengthened. For untrusted LTI providers, webbots acting as fake students could feed noise to the provider implementing a kind of differential privacy.

Open source implementations of these and many more privacy measures need to be included in the open source LMS and reference implementations of interoperability standards.

Data privacy is a second-order problem that has arisen after decades of moving part of the education system online. Similar to most of the effects of digitized activities, this impact follows an exponential pattern: deceptive at first until it starts to display disruptive effects when it reaches the knee in the hockey-stick-like curve [56]. Currently, we are at a point where the issue has been noticed, but very few and incomplete measures have been taken to address it. The data from students are being massively gathered, used with little or no supervision, and often leaked to unknown actors. These data will be bound to digital profiles of the students for countless years to come.

Author Contributions: Conceptualization, M.A. and C.S.; investigation, M.J.C.G., D.A. and M.A.; writing—original draft preparation, M.A., M.J.C.G., D.A. and C.S.; writing—review and editing, M.J.C.G., M.A. and D.F.; All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

AMS	Academic Management System
AWS	Amazon Web Services
AWS EC2	Elastic Cloud Computing
AWS S3	AWS Simple Storage Service
FERPA	Family Educational Rights and Privacy Act
IMS CC	IMS Common Cartridge
IMS LTI	IMS Learning Tools Interoperability
GDPR	General Data Protection Regulation
LMS	Learning Management System

References

1. García-Peñalvo, F.J.; Alier, M. Learning management system: Evolving from silos to structures. *Interact. Learn. Environ.* **2014**, *22*, 143–145. [CrossRef]
2. Wang, Q.; Woo, H.L.; Quek, C.L.; Yang, Y.; Liu, M. Using the Facebook group as a learning management system: An exploratory study. *Br. J. Educ. Technol.* **2012**, *43*, 428–438. [CrossRef]
3. Zeide, E. The structural consequences of big data-driven education. *Big Data* **2017**, *5*, 164–172. [CrossRef]
4. Guzmán-Valenzuela, C.; Gómez-González, C.; Rojas-Murphy Tagle, A.; Lorca-Vyhmeister, A. Learning analytics in higher education: A preponderance of analytics but very little learning? *Int. J. Educ. Technol. High. Educ.* **2021**, *18*, 23. [CrossRef]
5. Polonetsky, J.; Tene, O. Who is reading whom now: Privacy in education from books to MOOCs. *Vanderbilt J. Entertain. Technol. Law* **2014**, *17*, 927.
6. Hill, P. State of Higher Ed LMS Market for US and Canada: Spring 2016 Edition. In eLiterate. 2016. Available online: https://eliterate.us/state-higher-ed-lms-market-spring-2016/ (accessed on 20 May 2021).
7. Alier, M.; Mayol, E.; Casañ, M.J.; Piguillem, J.; Merriman, J.W.; Conde, M.Á.; García-Peñalvo, F.; Tebbens, G.; Severance, C. Clustering projects for eLearning interoperability. *J. Univers. Comput. Sci.* **2012**, *18*, 106–122.
8. Casany, M.J.; Alier Forment, M.; Mayol, E.; Piguillem, J.; Galanis, N.; García-Peñalvo, F.J.; Conde González, M.Á. Moodbile: A framework to integrate m-learning applications with the LMS. *J. Res. Pract. Inf. Technol.* **2012**, *44*, 129–149.
9. Common Cartridge: How Common Cartridge Benefits K-20 Institutions. Available online: https://www.imsglobal.org/activity/common-cartridge (accessed on 20 May 2021).

10. Learning Tool Interoperability: IMS LTI 1.3 and LTI Advantage. Available online: http://www.imsglobal.org/activity/learning-tools-interoperability (accessed on 20 May 2021).
11. Alier, M.; Casany, M.J.; Conde, M.A.; García-Peñalvo, F.J.; Severance, C. Interoperability for LMS: The missing piece to become the common place for e-learning innovation. *Int. J. Knowl. Learn.* **2010**, *6*, 130–141. [CrossRef]
12. Casany, M.J.; Alier, M.; García-Peñalvo, F.J. SOA initiatives for eLearning. A Moodle Case. In Proceedings of the IEEE 23rd International Conference on Advanced Information Networking and Applications, AINA 2009, Bradford, UK, 26–29 May 2009; pp. 750–755.
13. Timeline Amazon Web Services. In Wikipedia. Available online: https://en.wikipedia.org/wiki/Timeline_of_Amazon_Web_Services (accessed on 20 May 2021).
14. LMS Market Share for US & Canadian Higher Ed. Institutions. *Online Learning Distance Education Resoure*. 2018. Available online: https://tonybates.wpengine.com/wp-content/uploads/LMS-market-trends-2.jpg (accessed on 20 May 2021).
15. De Bruyckere, P. Gartner Hipe Cicle for Education. The Economy of Meaning. 2016. Available online: https://theeconomyofmeaning.com/2016/08/09/gartners-hype-cycle-for-education-2016/ (accessed on 26 August 2020).
16. Siemens, G. Learning analytics: Envisioning a research discipline and a domain of practice. In Proceedings of the 2nd International Conference on Learning Analytics and Knowledge, Leuven, Belgium, 29 April–2 May 2012; pp. 4–8.
17. Huotari, K.; Hamari, J. Defining gamification—A service marketing perspective. In Proceedings of the 16th International Academic MindTrek Conference, Tampere, Finland, 3–5 October 2012.
18. Casteneda, L. Gamificación, Que Podria Haber Más Allá De Micro-Estimulitos. 2014. Available online: https://www.lindacastaneda.com/es/mushware/gamificacionesytic1314/ (accessed on 20 May 2021).
19. Capuano, N.; Caballé, S. Adaptive learning technologies. *AI Mag.* **2020**, *41*, 96–98. [CrossRef]
20. Zuboff, S. Big other: Surveillance capitalism and the prospects of an information civilization. *J. Inf. Technol.* **2015**, *30*, 75–89. [CrossRef]
21. Drachsler, H.; Greller, W. 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, Edinburg, UK, 25–29 April 2016; pp. 89–98.
22. Ravitch, D. Is inBloom engaged in identity theft. *Diane's Ravitch Blog*. 2013. Available online: https://dianeravitch.net/2013/04/07/is-inbloom-engaged-in-identity-theft/ (accessed on 7 April 2021).
23. CBP 2014. College Bescherming Persoonsgegevens Onderzoek. CBP Naar De Verwerking Van Persoonsgegevens Door Snappet Rapport Definitieve Bevindingen Van 14 Juli 2014 Met Corrigendum Van 27 Augustus 2014 Juli 2014. Available online: https://cbpweb.nl/sites/default/files/downloads/mijn_privacy/rap_2013_snappet.pdf (accessed on 20 May 2021).
24. Williamson, B.; Rutherford, A. ClassDojo Poses Data Protection Concerns for Parents. Available online: https://blogs.lse.ac.uk/parenting4digitalfuture/2017/01/04/classdojo-poses-data-protection-concerns-for-parents/ (accessed on 16 August 2021).
25. Cox, K. College contact-tracing app readily leaked personal data, report finds. *ArsTechnica*. 2020. Available online: https://arstechnica.com/tech-policy/2020/08/college-contact-tracing-app-readily-leaked-personal-data-report-finds/ (accessed on 7 April 2021).
26. Cushing, T. Tracking college students everywhere they go on campus is the new normal. *Techdirt*. 2019. Available online: https://www.techdirt.com/articles/20191226/12031843636/tracking-college-students-everywhere-they-go-campus-is-new-normal.shtml (accessed on 7 April 2021).
27. Cushing, T. University of Alabama is using a location-tracking app to punish students for leaving football games early. *Techdirt*. 2019. Available online: https://www.techdirt.com/articles/20190915/13384942992/university-alabama-is-using-location-tracking-app-to-punish-students-leaving-football-games-early.shtml (accessed on 7 April 2021).
28. Geoffrey, A. Colleges are turning students' phones into surveillance machines, tracking the locations of hundreds of thousands. *Wash. Post*. 2019. Available online: https://www.washingtonpost.com/technology/2019/12/24/colleges-are-turning-students-phones-into-surveillance-machines-tracking-locations-hundreds-thousands/ (accessed on 7 April 2021).
29. McMillan, D.; Anderson, N. Student tracking, secret scores: How college admissions offices rank prospects before they apply. *Wash. Post*. 2020. Available online: https://www.washingtonpost.com/business/2019/10/14/colleges-quietly-rank-prospective-students-based-their-personal-data/ (accessed on 7 April 2021).
30. MacMillan, D. Some colleges are tracking students before they even apply. *Wash. Post*. 2019. Available online: https://www.washingtonpost.com/podcasts/post-reports/some-colleges-are-tracking-students-before-they-even-apply/ (accessed on 7 April 2021).
31. Barnds, W.K. Does big data know best? NSA and college admissions. *Huffington Post*. 2013. Available online: http://www.huffingtonpost.com/w-kent-barnds/does-big-data-know-bestn_b_3460096.html (accessed on 7 April 2021).
32. Rubel, A.; Jones, K. Data analytics in higher education: Key concerns and open questions. *Univ. St. Thomas J. Law Public Policy* **2017**, *11*, 25.
33. Satisky, J.A. Duke study recorded thousands of students' faces. Now they're being used all over the world. *Chronicle*. 2019. Available online: https://www.dukechronicle.com/article/2019/06/duke-university-facial-recognition-data-set-study-surveillance-video-students-china-uyghur (accessed on 7 April 2021).
34. Owen, M. Facial recognition bolstered by mass database scraping, but not from apple. *Apple Insider*. 2019. Available online: https://appleinsider.com/articles/19/07/13/facial-recognition-bolstered-by-mass-database-scraping-but-not-from-apple (accessed on 7 April 2021).

35. Lin, L. Thousands of chinese students′ data exposed on internet. *Wall Str. J.* 2020. Available online: https://www.wsj.com/articles/thousands-of-chinese-students-data-exposed-on-internet-11579283410 (accessed on 7 April 2021).
36. Parry, M. Big data on campus. *New York Times*. 2012. Available online: https://www.nytimes.com/2012/07/22/education/edlife/colleges-awakening-to-the-opportunities-of-data-mining.html (accessed on 7 April 2021).
37. New initiatives advance asu′s efforts to enhance student′s success. *Arizona State University News*. 2011. Available online: https://news.asu.edu/content/new-initiatives-advance-asus-efforts-enhance-student-success (accessed on 7 April 2021).
38. Denley, T. Advising by algorithm. *New York Times*. 2012. Available online: https://archive.nytimes.com/www.nytimes.com/interactive/2012/07/18/education/edlife/student-advising-by-algorithm.html?ref=edlife (accessed on 7 April 2021).
39. O'Neil, M. Data Breaches Put a Dent in Colleges′ Finances as well as Reputations. Available online: https://www.chronicle.com/article/data-breaches-put-a-dent-in-colleges-finances-as-well-as-reputations/ (accessed on 7 April 2021).
40. Kamenetz, A. What parents need to know about big data and student privacy. *NPR*. 2014. Available online: https://www.npr.org/sections/alltechconsidered/2014/04/28/305715935/what-parents-need-to-know-about-big-data-and-student-privacy?t=1617701713223 (accessed on 7 April 2021).
41. Pardo, A.; Siemens, G. Ethical and privacy principles for learning analytics. *Br. J. Educ. Technol.* **2014**, *45*, 438–450. [CrossRef]
42. Prinsloo, P.; Slade, S. An evaluation of policy frameworks for addressing ethical considerations in learning analytics. In Proceedings of the LAK '13 Proceedings of the 3rd International Conference on Learning Analytics and Knowledge, New York, NY, USA, 8–13 April 2013.
43. Slade, S.; Prinsloo, P. Learning analytics: Ethical issues and dilemmas. *Am. Behav. Sci.* **2013**, *57*, 1510–1529. [CrossRef]
44. Willis, J.E., III. *Ethics, Big Data, and Analytics: A Model for Application*; Purdue University: West Lafayette, IN, USA, 2013.
45. Amo Filvà, D. Privacidad Y Gestión de la Identidad en Procesos de Analítica de Aprendizaje, phd dissertartion, Programa de Doctorado Formación en la Sociedad del Conocimiento. Universidad de Salamanca. 2020. Available online: https://repositorio.grial.eu/handle/grial/1951 (accessed on 1 July 2021).
46. Zeide, E. Student privacy principles for the age of big data: Moving beyond FERPA and FIPPs. *Drexel Law Rev.* **2015**, *8*, 101–160.
47. Zeide, E. The limits of education purpose limitations. *Univ. Miami Law Rev.* **2016**, *71*, 496–526.
48. Marković, M.G.; Debeljak, S.; Kadoić, N. Preparing students for the era of the General Data Protection Regulation (GDPR). *TEM J.* **2019**, *8*, 150. [CrossRef]
49. van Ooijen, I.; Vrabec, H.U. Does the GDPR enhance consumers′ control over personal data? An analysis from a behavioral perspective. *J. Consum. Policy* **2019**, *42*, 91–107. [CrossRef]
50. Negroponte, N.; Harrington, R.; McKay, S.R.; Christian, W. Being digital. *Comput. Phys.* **1997**, *11*, 261–262. [CrossRef]
51. Cyphers, B. Google′s FLoC is a terrible idea. *Electron. Front. Found.* 2021. Available online: https://www.eff.org/deeplinks/2021/03/googles-floc-terrible-idea (accessed on 16 August 2021).
52. Li, H.; Chen, Q.; Zhu, H.; Ma, D.; Wen, H.; Shen, X.S. Privacy leakage via de-anonymization and aggregation in heterogeneous social networks. *IEEE Trans. Dependable Secur. Comput.* **2017**, *17*, 350–362. [CrossRef]
53. Amo, D.; Fonseca, D.; Alier, M.; García-Peñalvo, F.J.; Casañ, M.J.; Alsina, M. Personal data broker: A solution to assure data privacy in EdTech. In Proceedings of the International Conference on Human-Computer Interaction, Orlando, FL, USA, 26–31 July 2019; Springer International Publishing: Berlin/Heidelberg, Germany, 2019; pp. 3–14.
54. Dwork, C.; Roth, A. The algorithmic foudation of differencial privacy. *Found. Trends Theor. Comput. Sci.* **2014**, *9*, 211–407. [CrossRef]
55. Amo, D.; Alier, M.; García-Peñalvo, F.J.; Fonseca, D.; Casañ, M.J. Protected users: A moodle plugin to improve confidentiality and privacy support through user aliases. *Sustainability* **2020**, *12*, 2548. [CrossRef]
56. Diamandis, P. THE 6 D′S. Available online: https://www.diamandis.com/blog/the-6ds (accessed on 7 April 2021).