The influence of conception paradigms on data protection in e-learning platforms: a case study

Christophe Kiennert, Nathan de Vos, Manon Knockaert, Joaquin Garcia-Alfaro

IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received April 1, 2019, accepted April 24, 2019, date of publication May 7, 2019, date of current version May 29, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2915275

The Influence of Conception Paradigms on Data Protection in E-Learning Platforms: A Case Study

CHRISTOPHE KIENNERT[1], NATHAN DE VOS[2], MANON KNOCKAERT[2], AND JOAQUIN GARCIA-ALFARO[1]

[1]SAMOVAR CNRS UMR 5157 Télécom SudParis, Institut Polytechnique de Paris, 91128 Palaiseau, France
[2]Centre de Recherche Information Droit et Société, University of Namur, 5000 Namur, Belgium

Corresponding author: Joaquin Garcia-Alfaro (jgalfaro@ieee.org)

This work was supported in part by the H2020-ICT-2015/H2020-ICT-2015 TeSLA Project *An Adaptive Trust-based e-assessment System for Learning* under Grant 688520, and in part by the European Commission (H2020 SPARTA Project) under Grant 830892.

ABSTRACT The wide adoption of virtual learning environments such as Moodle in numerous universities illustrate the growing trend of e-learning development and diffusion. These e-learning environments alter the relationship between the students and academic knowledge and learning processes considerably stimulating the students' autonomy by making most of the course material freely available at any time while inducing a progressive reduction of physical student-teacher interactions with virtual ones. Recent advances, as proposed in the TeSLA project, even introduces an e-assessment environment. This entire virtual learning framework raises new concerns in terms of privacy, given that such environments are potentially able to track the students, profile their habits, and retrieve personal data. In this paper, we analyze the influence of conception paradigms of e-learning platforms on personal data protection, based on a classification of these platforms in two antagonistic approaches. We illustrate our analysis with a case study of the TeSLA project and examine how the design choices impact the efficiency and legal compliance of personal data protection means. We finally propose alternative designs that could lead to significant improvements in this matter.

INDEX TERMS E-learning, privacy, data protection.

I. INTRODUCTION

The transformation of educational environments over time is often regarded mostly as a technological evolution, aiming to improve the accessibility of students to course material as well as to encourage interactivity with teachers. Indeed, many universities have already integrated e-learning platforms, such as Moodle, to enhance the educational content of their courses. However, such evolution also entails deeper, more fundamental changes in the relationship between students and academic knowledge.

In particular, some institutions have chosen to fully embrace this change by specializing in online education. These so-called open universities are attempting to deploy bachelor and master degree programs in all subjects taught in traditional universities with certification. The latter raises logistical problems related to the very nature of online activities characterized by spatial and temporal asynchronity (students carry them out at different times and places). These concerns lead to a lack of legitimacy of remote certification, perceived as unreliable to prove the presence and absence of cheating, which these institutions hope to counterbalance by the development of an innovative system.

E-learning systems, and even more so e-assessment systems, are potentially subject to privacy issues. Remote authentication of students requires exchanges of personal data that must be carried out within the legal framework provided for in this respect. The objective of this article is to show with the help of a case study how an e-assessment system can be associated with such concerns. The analysis will be placed at the level of the technical object during its development, and will be comprised of sociological, legal and technical aspects.

Our analysis will be illustrated with an e-assessment system called TeSLA (Adaptative Trust-based e-assessment System for Learning), designed through a H2020 project funded by the European Commission. This system relies on several anti-cheating countermeasures which consist in the real-time gathering of several biometric samples to ensure the authentication of the learner during the whole remote assessment session. This process allows the university instructors to be convinced of the learner's identity when grading the assessment. However, biometric samples are regarded as highly

The associate editor coordinating the review of this manuscript and approving it for publication was Ahmed M. Elmisery.

sensitive data, from both ethical and legal points of view. Therefore, on top of normalizing the use of biometrics in an academic context, the choices made in the conception of this framework lead to an extremely large collection by the TeSLA system of various biometric samples from all enrolled learners, which should necessarily come with strong guarantees regarding the justification and the processing of such sensitive data. Analyzing this process implies regarding these choices as an attempt to manage users (both learners and teachers) in a way that could lead to some unintended outcomes in terms of privacy, even from the point of view of the system designers, as pointed out in a seminal work by Suchman [27].

Indeed, our approach stands in the middle of a naive vision supposing the validity of the development approach without questioning it, and of a deterministic vision that would picture the developers as careless about privacy issues. Our idea is to approach the TeSLA system, our case study, as a contingent result of a set of social variables that nevertheless gives us indications on the normativity specific to these online learning and examination systems. Our analysis actually relies on the premise that technologies are not inert and disembodied devices, and that they embody a set of norms (representations of what a *good student* is, of what cheating is, etc.) that influence the behavior of their users [10]. Technical choices always involve social and even political choices as they empower, or conversely disadvantage, the people getting in touch with them.

The paper is organized as follows. Section II provides a brief state of the art on approaches for the development of e-learning tools, under sociological and technical perspectives. Section III presents the legal framework on data privacy, particularly at the European level. Section IV describes the TeSLA use case and describes the architecture of the project. On the basis of this description, Section V highlights the privacy issues raised by this project due to its fundamental choices. Section VI proposes alternatives to the current TeSLA architecture to highlight the impact of different choices on data protection and privacy. Section VII concludes the paper.

II. APPROACHES ON E-LEARNING PLATFORMS

A. SOCIOLOGICAL ASPECTS

The purpose of this section is to highlight the two major ways of thinking about the technical innovation that has recently accompanied the development of e-learning and e-assessment systems. These are especially reflected in an article by Feenberg and Hamilton [13], which illustrates them with examples from the early days of e-learning in the modern sense of the term (via online information exchanges).

On one hand, Feenberg and Hamilton pinpoint what we could call today the *mainstream* learning management systems that have an initial aim, which is often information spreading. The approach is not sensitive to the different contexts of e-learning, systems are seen as a tool that teachers are going to be familiar with regardless of their study field or pedagogical method. The two authors describe this system as based on behaviorist assumptions that strongly imbued the e-learning of the time (and that still have a clear influence today). The technology aims to pre-define the behaviors to be generated. Here, the best way to accommodate spatial and temporal asynchrony is to make the student's behavior *calculable* via various indicators, thus indirectly encouraging them to self-manage.

One the other hand, Feenberg and Hamilton are giving an insight of a *context sensitive* systems. During the early days of e-learning, the idea of such systems, called *WBSI*, has been developed on the fact that e-learning differs from traditional learning, as the first implies only temporally and geographically asynchronous relations between professors and students. WBSI's principal objective was to widen communication possibilities allowed by these technologies. The system was thought in an inclusive way toward different teaching methods. There was no preconception of the definitive goal of the system, which made its development rich in terms of newness but chaotic in terms of implementation.

These two distinct approaches are similarly found in the dichotomy drawn by the anthropologist Suchman. In a famous fieldwork [27], she explains how engineers first presuppose what the so-called *plans* of future users are. These correspond to the precise way in which users must use a technical object to meet a predetermined objective. It is an *optimistic* projection of the designer, since users often reinterpret this *plan* in their own way, sometimes leading to unexpected results. The aim is to ensure that a predictable standard mode of use of the object is imposed on everyone. In line with this approach, and in the context of the expansion of e-learning technologies, we aim to analyze unintentionally induced uses of a system conceived to legitimate the remote assessment of learners, and the impact of such unplanned uses on data protection and users' privacy.

According to Suchman, engineers generally use this development mode while a second way of doing things exists. This different approach is intended to be more open, an objective is defined upstream but does not prefigure the exact way to achieve it. The final use of the object is not predictable, which makes it more adaptable according to the circumstances encountered. This philosophy of innovation is based more on experimentation than on the achievement of a goal.

B. TECHNICAL ASPECTS

Related work in the literature reports existing models and platforms for electronic examination of learners. These systems focus on the need to objectively authenticate students at a distance in order to overcome the concerns about the legitimacy of e-assessment faced by universities. Two recent approaches can be mentioned:

- Services like Safe Exam Browser and Secure Exam may require the installation of dedicated software into learners' computers to take remote assessments. The software controls the execution of unauthorized actions

during the exam, such as executing multi-task applications, web connections, etc. From a pedagogical point of view, this type of solution has negative effects on the learners, creating stress and affecting the results of the examination [2].

- An alternative approach is the use of Proctor-based assessments. Human proctors are selected by the learners, whose main responsibilities include face-to-face monitoring of learners, combined with some technological solutions, such as web cameras and voice services, during the execution of special examinations. Online services such as Kryterion, ProctorU and Pearson VUE rely on this type of approach. The main drawbacks of these solutions consist in the technical scalability issue, as well as the lack of authorship guarantees [14].

Overall, such approaches are flawed with a few limitations, the most noticeable being the lack of a continuous process relying on solutions addressing authentication and authorship as a whole. These limitations constitute the main technical challenges that the TeSLA project aims to address.

III. LEGAL FRAMEWORK ON PRIVACY

The General Data Protection Regulation [23] (hereafter GDPR) is the new European tool to ensure the protection of personal data. It aims at modernizing the legal framework and strengthening the responsibility of the data controller, defined as the entity that determines the purposes and means of the processing of personal data. According to Article 4.1 of the GDPR, *personal data* refers to "any information relating to an identified or identifiable natural person; an identifiable natural person is one who can be identified, directly or indirectly, in particular by reference to an identifier such as a name, an identification number, location data, an online identifier or to one or more factors specific to the physical, physiological, genetic, mental, economic, cultural or social identity of that natural person".

The GDPR applies to the processing of personal data wholly or partly by automated means, and to the processing other than by automated means of personal data which are part of a filing system or are intended to be part of a filing system (defined as any structured set of personal data accessible according to specific criteria). The GDPR may apply even where neither the controller nor the processor (defined as the natural or legal person which processes personal data on behalf of the data controller) are established on the territory of the European Union. It will be the case when the processing concerns the offering of goods or services irrespective of whether a payment of the data subject is required to such data subjects in the Union or if the monitoring of their behavior as far as their behavior takes place within the European Union.

The GDPR recalls the following principles established in Directive 95/46/EC on the protection of individuals with regard to the processing of personal data and on the free movement of such data: transparency, purpose limitation, data minimization, accuracy, storage limitation, integrity and confidentiality.

This section aims to present the main aspects of data protection regulation that apply for e-learning activities. Indeed, the implementation of a system such as TeSLA involves the processing of a large amount of personal data that receive a particular attention from the GDPR. To follow our analysis, the institutions are considered as data controllers, and TeSLA is the data processor.

A. PRIVACY BY DESIGN

Privacy by design is a newly formalized obligation introduced by the GDPR to reinforce these data protection principles. The concept of privacy by design already existed in the previous Directive 95/46/EC, but now, its implementation is clearer and formally mandatory in specific circumstances [1].

This imperative requires the data controller to ensure that the system put in place is compliant with the fundamental principles of personal data protection. In addition, the GDPR encourages the technology to follow the movement in order to ensure effective protection of the personal data. In other words, it is about thinking the process differently. This is no longer a question of developing a system and then mending rules. The logic is reversed: the architectural design of the system and the different algorithmic operations must integrate in themselves the guarantees of data protection, at all stages of the processing of the personal data from the collection to the deletion or anonymization after a specified retention period. By an *a priori* integration of legal norms, the objective pursued by the European legislator is to reverse a situation in which the development of technology precedes the legal constraints.

In addition to be a binding obligation for the data controller, the GDPR intends to go further by encouraging product manufacturers, service providers and application producers to take into account the data protection right when developing and designing their products or services. The notion of data protection by design receives an echo in the recent modernization of Convention 108 for the Protection of Individuals with regard to Automatic Processing of Personal Data, which is the first and only international instrument concerning data protection.

If privacy by design was at first a legal concept at which legislation and case law paid attention to, it became a genuine obligation. Indeed, Article 25 of the GDPR states that the data controller has to implement appropriate technical and organizational measures, both at the time of the determination of the means for processing and at the time of the processing itself. Each institution that wants to integrate e-learning tools must have a clear policy concerning the rules of access, the situations justifying a processing, the time of storage, the confidentiality as well as the security by avoiding a single point of failure and a proliferation of storage locations that could lead to the loss of control over personal data [12].

B. MINIMIZATION OF PERSONAL DATA COLLECTION

The starting point in the conception of a system compliant with the data protection law is the principle of minimization

provided in Article 5 of the GDPR. This principle imposes to only collect personal data that are adequate, relevant and limited to what is necessary in relation to the purposes for which they are processed. This implies that the data controller considers the data that are strictly necessary to achieve its purpose and promote the use of anonymized or at least pseudonymized data. For example, one of the major objectives for the institutions using e-learning is to be able to have a strong identification of the students. Note that a minimization policy does not apply solely to the amount of data collected, but also to the number of people who can access and process the data and the retention period [12].

This legal principle of minimization can receive two technical implementations. Firstly, the data controller should design his or her database expurged of any personal data not strictly necessary in order to accomplish the goal pursued. Secondly, it is also possible to select — initially — the data strictly necessary for the accomplishment of the purpose announced and — in a second step — to add personal data in order to ensure and maintain the integrity of the developed system. This second approach tries to legitimize the collection and processing of additional personal data for ensuring the security and correctness of a system. If both approaches are potentially valid, they must nevertheless be applied in accordance with all the principles of the GDPR.

The second approach consists of a balance between two legal principles: data minimization and security. While it is true that security requirements are, for the first time, raised as a real principle within the GDPR, it must be emphasized that the law only requires proportional security to the risks through the use of appropriate measures. Consequently, the collection of additional personal data to those strictly necessary to achieve the objective of the system requires a case-by-case analysis depending on the operating of the tool, the nature and the volume of data as well as the risk to rights and freedoms for the data subject. If the data controller can reasonably consider the collection of additional personal data to verify the identity of the person in order to guarantee a safe processing, the Article 11 of the GDPR demonstrates the overall philosophy of the data protection law. This provision describes the situation where the identification of the data subject is not required. If the purpose for which personal data are processed does not or does no longer require the identification of a data subject, the data controller is not obliged to maintain, acquire or process additional informer in order to be able to identify the data subject [29]. In any case, it appears that the collection of additional data must be transparent and the concern for security must be raised to the rank of a specific purpose.

It implies that each institution determines, prior to the collection of personal data of the students, the information really necessary for the fulfillment of the educational activities. Each institution must decide whether it is really necessary to combine several biometric recognition tools or whether this can enhance the security of proper identification.

However, the possession of some personal data may be particularly intrusive in relation to the data subject privacy, and requires a specific regime. For instance, personal data revealing racial or ethnic origin, political opinions, religious or philosophical beliefs, or trade union membership, and the processing of genetic data, biometric data for the purpose of uniquely identifying a natural person, data concerning health or data concerning a natural person's sex life or sexual are considered as sensitive data. The processing of such personal data is, in principle, prohibited, but this prohibition does not apply if the data controller, for example, receives the explicit consent of the data subject (Article 9.2 of the GDPR lists several possibilities to process sensitive data).

C. PRACTICAL STEPS FOR COMPLIANCE IN SYSTEM CONCEPTION

In order to respect the privacy by design principle, the data controller has to define beforehand a data processing purpose. The second step is the determination of the personal data strictly necessary to achieve this purpose. The condition of necessity covers two realities: on the one hand a quantitative necessity and, on the other hand, a qualitative necessity. Although it is important to avoid collecting too many data in relation to the objective pursued, it is just as important to verify that the collected data, even in a very small quantity, by their nature, are not qualitatively excessive. This means that it must be determined whether the processing of data does not unreasonably infringe the rights and freedoms of the data subject in relation to the benefits that the data controller could obtain.

This double check is essential when a data controller considers using sensitive personal data, such as biometric data. Firstly, it is important to consider the relevance of using sensitive data to achieve the purpose. Secondly, if the collection and processing of sensitive data are relevant, the data controller has to realize a qualitative necessity and consider what sensitive data will be collected and processed. For example, to authenticate users in a system, the data controller has to identify if it is relevant and proportionate to process both facial and vocal recognition in order to authenticate the user, when facial recognition alone might be enough. The controller must then operate a delicate and careful balance between efficiency and restriction of the intrusion into the privacy of the individual.

The data controller must also respect the minimization principle and have a strong justification for the processing of biometric data. In other words, the data controller has to demonstrate that there is no other option. However, in that case, the data controller has to put in place all the technical and organizational measures surrounding the processing of sensitive personal data to counteract the interference into data subject privacy while ensuring the confidentiality of these data [9]. In particular, the data controller has to enforce a strong policy for the access to the biometric data and to have procedural rules to check log files.

IV. USE CASE: THE TeSLA PROJECT

As mentioned previously, the TeSLA project (Adaptive Trust e-Assessment System) [28] is a EU-funded project that aims at designing an e-assessment architecture in order to allow learners from various universities to take remote examinations, while deploying the necessary countermeasures to prevent cheating. Thus, e-assessment takes existing e-learning environments one step further, since traditional e-learning systems, such as Moodle, are primarily designed to offer learners a restricted access to course material. In this section, we present an overview of the TeSLA projects and present the designed architecture, before focusing on the security and privacy issues raised in the context of this project.

A. OVERVIEW OF TeSLA

The objective of the TeSLA project is to provide an online environment to take remote assessments with enough anti-cheating countermeasures to guarantee a legitimacy of the e-assessment equivalent to that of traditional examinations. These countermeasures are comprised of various features such as real time biometric authentication and authorship verification.

Compared to the other existing e-assessment projects described in Section II, the TeSLA framework aims at covering several security requirements by combining technologies such as biometrics, digital certificates and trusted time stamping [20]. The biometric modalities on which authentication relies in the context of TeSLA include keystroke detection [4], [24], face recognition [26] and voice recognition [18]. The authorship of learners is addressed with plagiarism detection solutions [5], [11]. By combining these various approaches for identification, authentication, and cheating prevention, TeSLA aims at building a strong trust in the assignments submitted by the learners.

B. THE TeSLA ARCHITECTURE

The architecture that has been designed and implemented during the TeSLA project allows secure interactions between the traditional e-learning environment, and the e-assessment environment. The former is by nature tightly linked to the university, while the latter is independent from the institution. Therefore, the TeSLA architecture can be divided into two domains: the university on one side, and TeSLA on the other side, each domain being comprised of various components, described hereafter. The global TeSLA architecture is represented on Figure 1.

The TeSLA domain is composed of three main elements:

- The TeSLA E-assessment Portal (TEP), whose role is to acts as a service broker, receiving and forwarding requests to other TeSLA components.
- The TeSLA Portal, whose role is to compute statistics related to the e-assessment activities.
- Several instruments, whose role is to provide anti-cheating countermeasures.

It is worth nothing that the aim of the instruments is either to analyze the learner's biometric samples sent during the e-assessment in order to provide a feedback to the instructors, or to perform a post processing of the assignment, e.g. with an anti-plagiarism tool.

Regarding the university domain, the following elements can be listed:

- A Virtual Learning Environment (VLE), which corresponds to the environment offered by a traditional Learning Management System (LMS) such as Moodle.[1]
- A plugin integrated to the VLE, which offers a client side interface with the TeSLA domain.
- A number of tools added to the VLE, that send requests as well as data to the TeSLA domain through the plugin. These tools can be sorted in three sets: one learner tool, one instructor tool, and various external tools. Instructors prepare and submit online assessment using the instructor tool, while the learners take these e-assessments via the learner tool. External tools are meant to sample the learner's biometric modalities and to send them to TeSLA, where they will be redirected to instruments for analysis and evaluation, as a way to ensure real time authentication during the assessment.
- The TeSLA Identity Provider (TIP), which is an entity that converts the learner's real name into a randomly generated identifier, called TeSLA ID , in order to prevent any component of the TeSLA domain to know the learner's identity.

As mentioned above, the TeSLA architecture is illustrated in Fig. 1. All the exchanges between the many components are secured by the TLS protocol [7], which is deployed on the whole architecture with mutual authentication. The purpose of securing the data exchanges with the TLS protocol is on one hand to ensure the confidentiality and integrity of the data, and on the other hand, to make sure that no one is impersonating any of the components in the architecture. The underlying Public Key Infrastructure for the TLS management in the TeSLA architecture is explained in [17].

A classic scenario of a learner taking an e-assignment can be summed up as follows. First, the learner has to log in on the VLE where the client-side plugin has been integrated. Then, the learner requests the e-assignment through the learner tool available on the VLE. Before sending the request to the TeSLA domain through the TEP, the real name of the learner is replaced by the TeSLA ID obtained from the TIP. Then, the TEP fetches the e-assignment in its database and sends it back to the VLE. Finally, the learner can start to take the e-assignment while the external tools regularly gather biometric data, sent to the TeSLA instruments in real time. Once the e-assessment is over, the instructor will be provided with the analysis results of the biometric samples, as well as with the e-assignment to grade.

The biometric data gathered through the external tools can be as varied as keystroke dynamics, voice and face

[1] https://moodle.org/

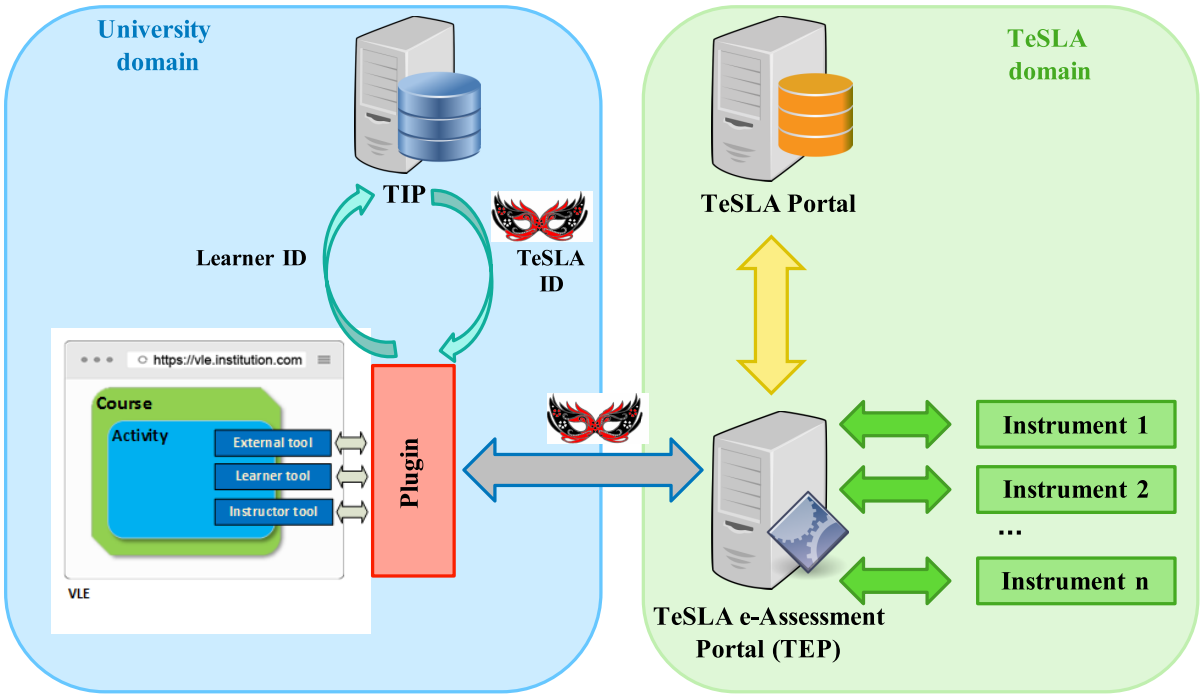

FIGURE 1. Simplified TeSLA architecture representation.

recognition. Keystroke dynamics consists in a real time measurement of the learner's keystroke frequency all through the e-assessment, to be compared with a test sample built during an enrolment phase. Similarly, voice samples and face pictures can be gathered by TeSLA at regular intervals. An algorithm then compares the collected biometric data to the preexisting samples, and for each sample, a feedback is provided on the teacher's page with a score indicating the computed level of trust.

C. PRIVACY MANAGEMENT IN TeSLA

Ensuring privacy consists in minimizing the personal information retrieved by the TeSLA system during its interactions with the learners, while anonymizing the data exchanged and stored in the databases whenever practical [6]. The TeSLA architecture provides the learner with pseudonymized identifiers, which hide the learner's genuine identity when taking e-assessment activities.

Yet, in the context of e-assessment, full anonymity cannot be provided to learners. Indeed, the very nature of the notion of assessment requires to grade identified students, and makes it mandatory to store the association between the TeSLA ID and the real name of the learner. Therefore, in such context, anonymity can only be partial. In other words, only pseudonymity can be provided to learners when taking an e-assessment. Indeed, since the association between the learner's true identity and the TeSLA ID is stored in the TIP database, any claim of anonymization is impossible, even though the learner's identity is never transmitted to the TeSLA domain.

The TeSLA ID is a random number computed by the TIP following version 4 of the UUID standard [19]. The TIP database shall be shared with all the VLEs. Since any interaction between the university domain and the TeSLA domain involves the client-side plugin and the TEP, ensuring pseudonimity only requires that all requests sent to the TEP are first redirected to the TIP in order to obtain the learner's TeSLA ID .

Regarding sensitive personal data, the most crucial aspect of this platform has to do with the management of the biometric samples. Indeed, as highlighted in Section IV-B, several types of biometric data are gathered in real time from the learner during the e-assessment process. These data are transmitted over a secured TLS channel to the TEP in the TeSLA domain, for analysis and comparison with the enrolled samples.

The samples themselves are usually not accessible to the instructors, which imposes a boundary on the availability domain of such sensitive data. Only the results of the samples analysis are returned to the teacher's page on TeSLA, providing only an estimation of the trustworthiness of these samples. However, though not designed as a feature of the original architecture, it was eventually made possible for a teacher facing samples with poor matching rates to gain access to the doubtful biometric samples. This choice was made in order to avoid punishing honest learners because of false positives. Though adding a human factor before coming up with the final decision to punish a learner for cheating undoubtedly helps to reduce the risks of punishing a honest learner, it raises serious issues regarding how easy it becomes for a teacher to

gain access to sensitive data such as biometric samples. Such issues are discussed next.

V. ANALYSIS OF THE TeSLA PLATFORM

A. IMPACT ON LEARNERS AND ASSESSMENT

The purpose of the TeSLA platform is to provide an e-assessment platform with strong anti-cheating countermeasures, while at least partly ensuring users' privacy. Therefore, TeSLA primarily aims at addressing the lack of legitimacy of online assessments due to the multiple cheating possibilities, that include impersonation. By choosing to exclude any human interaction during the assessment, the platform shifts the burden of authentication to the learners themselves, imposing various rigid constraints such as real time biometric sampling.

Rather than focusing on alternative assessment practices that would minimize the requirements on personal data, TeSLA rather grants universities the possibility to propose all kinds of exams to their learners, with various requirements on the types of biometric modalities to be sent to the system, all while taking potential disabilities of the learners into account. By doing so, learners are left with very little margin in their way to use the system. Similarly, instructors have to design exams in accordance with the features of the TeSLA system, in order to make sure that enough types of biometric modalities will be retrieved for the authentication of the learner. Both of them are guided by the technical features implemented in TeSLA, thus corresponding to the *plan* concept developed by Suchman [27], mentioned in Section II. Similarly, TeSLA aims at keeping track of the learners' activity in real time, extending the numerous possibilities for learners' activity surveillance in current e-learning platforms such as Moodle, where authentication allows the teachers to know exactly when, e.g., a learner is logged in or when he accessed course material.

Another aspect of TeSLA that is worth highlighting has to do with the conception of cheating. First, the system relies on the implicit assumption that taking an e-assessment without cheating requires being alone. Though the system is not able to guarantee that the learner is alone in the room (another person may be present behind the camera for example), the focus on face and voice recognition shows that the system is intended to replace classic examination instead of designing new examination practices, e.g. team work based assessment. Second, in a system such as TeSLA, anti-cheating countermeasures are all focused on detecting impersonation. Contrary to classic exams, where cheating mostly consists in accessing unauthorized sources of knowledge, TeSLA cannot check if a learner is browsing the Internet in search for answers. However, the instructors taking part in the project are well aware of this situation and accept it as part of the e-assessment process, which constitutes a significant evolution in the conception of assessments.

It should be emphasized that students have to give their explicit consent before taking an e-assessment that will gather biometric samples from them, hence leaving them the choice to refuse the remote test. However, TeSLA itself provides no guarantee regarding the consequences of a refusal from a learner. Therefore, whether or not other it can be replaced with a less intrusive assessment, or with a face to face exam, entirely depends on the university, which could lead to consent being given by learners less freely than originally designed.

Finally, in a context where the sampling of biometric modalities is still under strong law regulation (and emphasized by the GDPR, as detailed in Section III), a system so heavily based on biometrics such as TeSLA might take part in normalizing the mass collection of biometric samples by institutions over the years, in that learners will tend to regard such practice as more commonplace than it is now, which could be a significant source of evolution regarding people's relationship with their personal data.

B. PRIVACY ISSUES

Although TeSLA provides guarantees regarding privacy, such as relying on a TeSLA ID , several issues can still be raised. As explained in Section IV-C, the TeSLA ID allows learners to hide their identity to the TeSLA components, thus ensuring pseudonymity during e-assessment activities. However, it is not enough to prevent the acquisition and correlation of personal data by the TeSLA system. For example, the TeSLA ID does not ensure multi-session unlinkability, since the e-assessment system is obviously able to know when the same learner is logging in over two different sessions and track his activity even without knowing his identity.

Such flaw in the learner's privacy could lead to function creep, i.e. unintended treatment of sensitive data first collected for benign purposes. An example would be learner profiling that might be exploited to infer which learners are more likely to cheat, hence generating a bias in the system by considering the overall matching score of the learners' samples as objective representation of their likelihood to cheat, similarly to Porter's description of the emergence of quantitative data in social sciences [25]. Moreover, it should be reminded that the quality of the samples will vary from one learner to another depending on the quality of their own equipments. Therefore, it should not be excluded that learners with less costly, but less sophisticated devices are more likely to produce doubtful samples, and might therefore be more prone to be regarded as suspicious by the system. On top of creating a potential bias in the analysis of the samples, this issue may therefore also indirectly leak private information regarding the learner's standard of living.

Though such profiling is not implemented in TeSLA as part of the current project, no technical countermeasures have been added to prevent such future evolution. Such concern about the possibility of function creep is justified by the numerous controversies related to the extension of the way biometric data are used in various systems, as highlighted in [21].

Due to the fact that privacy filters were not clearly defined from the very start of the project, some uncertainty remained regarding the treatment and exploitation of the biometric samples gathered by TeSLA, and automatically analyzed by the instruments in order to provide feedback to teachers. These samples were supposed to remain strictly confined to the TeSLA domain, but as stated in Section IV-C, a sample that has a low matching ratio will lead the teacher to check the sample in person, and judge if it was a false positive or not. Such leak of sensitive data from the TeSLA domain to the university domain creates an important issue in terms of ethics, and makes it technically possible for teachers to exploit these biometric data at their own will, which is highly questionable with respect to the GDPR requirements, and stems from the lack of privacy by design development process as advised by the GDPR. Moreover, since biometric samples are gathered at a relatively high frequency, any learner is extremely likely to produce at least one bad sample during an e-assessment session (for example, when the learner's face is not clearly visible by the camera), such situation is bound to happen frequently.

Therefore, the choice to develop an e-assessment architecture that heavily relies on biometry without defining clear boundaries with respect to privacy is an eloquent illustration of the biases of a technology, introduced by the plan that here consists in collecting a large amount of sensitive and private data to first and foremost ensure learners' authentication. The main purpose of such plan is to provide the e-assessment system with a credibility on par with face to face examination, at the cost of serious privacy issues.

VI. ALTERNATIVE DESIGNS FOR TeSLA

A. PRIVACY-PRESERVING SCHEMES

Two additional enhancements of the TeSLA architecture towards improved privacy features relies on the use of anonymous certification and malleable signatures.

As described in [15], [16], anonymous certification allows us to perform privacy-friendly access control, in order to certify that users are allowed to access a resource because they own some attributes required by the verifier, without revealing their identity. As explained in Section IV-C, given the very principle of assessments, where learners receive personalized grades, it is necessary to identify each learner taking an assessment, whether through their true identity or a pseudonymzed identifier. For that reason, anonymous certification cannot replace the authentication scheme deployed in TeSLA. However, anonymous certification can be applied in a few other ways. First, it can be used as the main access control scheme for the pages hosting course material on the VLE. Indeed, the VLE has theoretically no need of the learners identity to decide whether they should have access to the course material or not. The only information possibly required by the VLE is:

- whether the student is enrolled at the university giving the course
- whether the student has registered for the course

These two items correspond to the two attributes that would be checked in the context of anonymous certification. By doing so, it would be not be technically possible for the VLE to track the learners' activity, and to profile the learners according to their hours of activity, the frequency of their access to the course material, etc., hence significantly improving the learners' privacy.

Another way to integrate anonymous certification into the TeSLA architecture consists in anonymizing the post processing of completed e-assignments. Indeed, as described in Section IV-B, after a learner takes an e-assessment, the assignment is sent to various external instruments, that perform anti-cheating post processings such as checking whether the assignment contains plagiarism. However, if the assignment is sent along with the TeSLA ID of the learner, it becomes technically possible for the instruments to keep track of a learner's data over time, as well as to correlate the data, for example in order to perform a more in-depth analysis for learners who happened to raise suspicions of cheating according to the results of other instruments. In the case where anonymous certification is used instead, relying on the aforementioned attributes, it becomes impossible for instruments to perform such tracking and correlation, hence enhancing the learners' privacy.

In terms of malleable signatures, we can envision the adaptation of the TeSLA architecture towards the combination of anonymization and sanitization. Similar approaches exist in the literature, in order to enforce lifelogging protection schemes [22], i.e., environments expected to record information about everyday lives of users via smart devices. Similarly to TeSLA, lifelogging environments involve the collection of large datasets, including sensitive information about users interacting with smart devices such as personal assistants, mobile cellphones, and Internet-connected watches [30]. This can lead to very serious privacy risks of personal data disclosure, as these data can be exploited in isolation, as well as combining the information generated between several of these devices.

Malleable signatures are mathematical construction that allow a designated party, the *sanitizer*, to modify given parts of a ciphertext, created by a *signer*, in a way that allows the modifications in a controlled way. The signer divides the signed information into several chunks or blocks, and provides a subset of them to the sanitizer. The subset, representing the series of admissible modifications, allows the sanitizer to modify given parts of a previously signed datum, while keeping the resulting signature still valid, under the public key of the signer. The scheme can satisfy privacy properties such as *unlinkability*, i.e., making unfeasible to the involving parties to distinguish between the initial source of data and the sanitized one [3]. The combination of malleable signatures and anonymous certification can also be used in order to handle record linkage attacks, while maintaining the utility of the data being processed [8].

B. TRUST AND TRANSPARENCY

In terms of trust, enhanced features beyond learners' privacy can also be added to future releases of the architecture. A system like TeSLA, where learners have to take e-assessments under strict anti-cheating countermeasures, requires a high degree of trust from learners in order to be widely deployed and accepted as a legitimate assessment tool. TeSLA should provide public guarantees that its claims regarding privacy and security are met, meaning that TeSLA is as transparent as possible with respect to personal data management processes. Though it is not directly related to security and privacy, TeSLA should also ensure transparency regarding the anti-cheating decision processes, and let learners know how these decisions are made while informing them of possible resorts at their disposal in case of false positive detection.

However, the choice of biometric-based authentication for learners who are taking e-assessments entails other issues. Firstly, the biometric samples are collected from the learner's computer, which by definition has no guarantee whatsoever regarding security. Even if the samples are not meant to be stored on the learner's computer, the risk of personal data theft at this point is independent from the TeSLA architecture, but is induced by the choice to rely on biometry. As such, it should be taken into account for further improvement of the TeSLA system. Secondly, even though the biometric samples are anonymized before they are sent to the TeSLA instruments, it may be better not to send such sensitive data to TeSLA at all, and decentralize Trusted Third Parties (TTPs) as much as possible. The role of TeSLA is to offer a specific service, namely the possibility to take e-assessments. It does not, and could not act as a TTP. In the current configuration, what happens to the biometric samples depends on how TeSLA is managed. With a TTP, which would have no specific connection to TeSLA or to the academic institutions, there would be a dedicated entity whose explicit role would be to guarantee the treatment of these sensitive data, independently of the current TeSLA policy. Notice that anonymous certification will benefit of such a TTP-decentralization, as well.

To sum up, we consider that improving trust and privacy in TeSLA requires further de-centralization of its fundamental choices, in order to offer the best guarantees to both learners and instructors. Even if the use of biometry is maintained as it is nowadays in TeSLA, extending current TTP elements, such as the TeSLA Public Key Infrastructure (PKI), and the underlying Certification Authorities (CAs), would be a significant step in this direction.

VII. CONCLUSION

We have addressed the influence of conception paradigms on data protection in virtual learning environments such as Moodle. We have argued that new e-learning environments may alter the relationship between students and academic knowledge, if they are not properly conceived. E-learning environments can potentially be used to track the students, profile their habits, and retrieve personal data. We have used as use case the recent advances of a EU project called TeSLA, which promotes the use of an e-assessment framework to remotely assess students, i.e., by reducing physical student-teacher. We have illustrated our analysis by examine how the design choices impact the efficiency and legal compliance of personal data protection means. We have also proposed some alternative designs that could lead to significant improvements in these matters.

REFERENCES

[1] A. Lee Bygrave, "Data protection by design and by default: Deciphering the EU's legislative requirements," *Oslo Law Rev.*, vol. 4, no. 2, pp. 105–120, Jun. 2017.
[2] K. M. Ala-Mutka, "A survey of automated assessment approaches for programming assignments," *Comput. Sci. Edu.*, vol. 15, no. 2, pp. 83–102, Jun. 2005.
[3] S. Canard and A. Jambert, "On extended sanitizable signature schemes," in *Proc. Int. Conf. Topics Cryptol.*, Mar. 2010, pp. 179–194.
[4] M. Choraś and P. Mroczkowski, "Recognizing individual typing patterns," in *Proc. Iberian Conf. Pattern Recognit. Image Anal.* Berlin, Germany: Springer, Jun. 2007, pp. 323–330.
[5] B. Cook *et al.*, "Academic integrity: Differences between computing assessments and essays," in *Proc. 13th Koli Calling Int. Conf. Comput. Educ. Res.*, Nov. 2013, pp. 14–17.
[6] G. Danezis *et al.* (Jan 2015). "Privacy and data protection by design—from policy to engineering." [Online]. Available: https://arxiv.org/abs/1501.03726
[7] T. Dierks and E. Rescorla, *The Transport Layer Security (TLS) Protocol Version 1.2*, document RFC 5246 (Proposed Standard), Updated by RFCs 5746, 5878, 6176, Aug. 2008.
[8] J. Domingo-Ferrer and D. Rebollo-Monedero, "Measuring risk and utility of anonymized data using information theory," in *Proc. EDBT/ICDT Workshops*, Mar. 2009, pp. 126–130.
[9] Appl. Note 30562/04 ECHR 1581 and 30566/04.S, 2008.
[10] A. Feenberg, *Transforming Technology: A Critical Theory Revisited*. London, U.K.: Oxford Univ. Press, 2002.
[11] O. H. Graven and L. M. MacKinnon, "A consideration of the use of plagiarism tools for automated student assessment," *IEEE Trans. Educ.*, vol. 51, no. 2, pp. 212–219, May 2008.
[12] S. Gürses, C. Troncoso, and C. Diaz, "Engineering privacy by design reloaded," in *Proc. Amsterdam Privacy Conf.*, 2015, pp. 1–21.
[13] E. Hamilton and A. Feenberg, "The technical codes of online education," *E-Learn. Digit. Media*, vol. 2, no. 2, pp. 104–121, Jun. 2005.
[14] P. Ihantola, T. Ahoniemi, V. Karavirta, and O. Seppälä, "Review of recent systems for automatic assessment of programming assignments," in *Proc. 10th Koli Calling Int. Conf. Comput. Educ. Res.*, Oct. 2010, pp. 86–93.
[15] N. Kaaniche, M. Laurent, P.-O. Rocher, C. Kiennert, and J. Garcia-Alfaro, "*PCS* a privacy-preserving certification scheme," in *Proc. Int. Workshop Data Privacy Manage.* (Lecture Notes in Computer Science). Oslo, Norway: Springer, Sep. 2017, pp. 239–256.
[16] C. Kiennert, N. Kaaniche, M. Laurent, P.-O. Rocher, and J. Garcia-Alfaro, "Anonymous certification for an e-assessment framework," in *Proc. Nordic Conf. Secure IT Syst.* (Lecture Notes in Computer Science). Cham, Switzerland: Springer, vol. 10674, Nov. 2017, pp. 70–85.
[17] C. Kiennert, P. O. Rocher, M. Ivanova, A. Rozeva, M. Durcheva, and J. Garcia-Alfaro, "Security challenges in e-assessment and technical solutions," in *Proc. 21st Int. Conf. Inf. Visualisation*, London, U.K., Jul. 2017, pp. 366–371.
[18] T. Kinnunen, E. Karpov, and P. Franti, "Real-time speaker identification and verification," *IEEE Trans. Audio, Speech, Language Process.*, vol. 14, no. 1, pp. 277–288, Jan. 2006.
[19] P. Leach, M. Mealling, and R. Salz, *A Universally Unique IDentifier (UUID) URN Namespace*, document RFC 4122 (Proposed Standard), Jul. 2005.
[20] Y.-C. Lu, Y.-S. Yang, P.-C. Chang, and C.-S. Yang, "The design and implementation of intelligent assessment management system," in *Proc. IEEE Global Eng. Educ. Conf. (EDUCON)*, Mar. 2013, pp. 451–457.
[21] E. Mordini, *Ethics and Policy of Biometrics*. London, U.K.: Springer, 2009, pp. 293–309.

[22] D. Pàmies-Estrems, N. Kaaniche, M. Laurent, J. Castellà-Roca, and J. Garcia-Alfaro, "Lifelogging protection scheme for internet-based personal assistants," in *Proc. Int. Workshop Data Privacy Manage.* Cham, Switzerland: Springer, Sep. 2018, pp. 431–440.
[23] Parliament and Council of European Union, "Regulation (EU) 2016/679 of the European parliament and of the council of 27 Apr. 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing directive 95/46/ec (general data protection regulation)," *Off. J.*, vol. 119, 2016.
[24] A. Peacock, X. Ke, and M. Wilkerson, "Typing patterns: A key to user identification," *IEEE Security Privacy*, vol. 2, no. 5, pp. 40–47, Sep. 2004.
[25] T. M. Porter, *Trust in Numbers—The Pursuit of Objectivity in Science and Public Life*. Princeton, NJ, USA: Princeton Univ. Press, 1995.
[26] P. Sinha, B. Balas, Y. Ostrovsky, and R. Russell, "Face recognition by humans: Nineteen results all computer vision researchers should know about," *Proc. IEEE*, vol. 94, no. 11, pp. 1948–1962, Nov. 2006.
[27] L. A. Suchman, *Plans and Situated Actions: The Problem of Human-Machine Communication* (Learning in Doing: Social, Cognitive and Computational Perspectives). New York, NY, USA: Cambridge Univ. Press, 1987.
[28] TeSLA Consortium. (2016). *Trust Based Authentication & Authorship e-Assessment Analysis*. [Online]. Available: http://tesla-project.eu/
[29] M. Veale, R. Binns, and J. Ausloos, "When data protection by design and data subject rights clash," *Int. Data Privacy Law*, vol. 8, no. 2, pp. 105–123, Apr. 2018.
[30] P. Wang and A. F. Smeaton, "Using visual lifelogs to automatically characterize everyday activities," *Inf. Sci.*, vol. 230, pp. 147–161, May 2013.

CHRISTOPHE KIENNERT received the Ph.D. diploma degree in computer science from Télécom ParisTech. He is currently an Associate Professor with the Networks and Telecommunication Services Department, Télécom SudParis Campus (Institut Mines-Telecom), Institut Polytechnique de Paris. His research activities include various topics in cybersecurity, which focuses on embedded systems, digital identity, intrusion detection, and mathematical models for security optimization in information systems.

NATHAN DE VOS is currently a Researcher with the Namur Digital Institute, University of Namur. His research focuses on sociology, organizational sciences, and technological policy. He also participates in the activities of the Centre of Research Law, Information and Society (CRIDS), University of Namur, in terms of legal and ethical issues raised by digital technologies. His activities include dissemination of sociological research, the publication of articles, and organization of conferences.

MANON KNOCKAERT received the master's degree in law from the Catholic University of Louvain, and the master's degree in law in information and communication technologies law from the University of Namur, where she is currently a Researcher, and more particularly with the Centre of Research Law, Information and Society (CRIDS). Her research interests are the intellectual property law in the information society, privacy, and data protection. She frequently gives conferences on the subject of data protection and intellectual property law. She is also involved in several research projects at national and European level.

JOAQUIN GARCIA-ALFARO received the Ph.D. diploma degree in computer science from the Autonomous University of Barcelona and the University of Rennes, and a research Habilitation from the Université Pierre and Marie Curie (Sorbonne Paris VI). He is currently a Full Professor with the Networks and Telecommunication Services Department, Télécom SudParis Campus (Institut Mines-Telecom), Institut Polytechnique de Paris. His research interests include a wide range of network security problems, with an emphasis on the management of security policies, analysis of vulnerabilities, and enforcement of countermeasures. He is involved in several research projects at the European level, related to ICT security.