Unplugged institutions: towards a localization of the cloud for Learning Analytics privacy enhancement

Daniel Amo-Filvà[1], David Fonseca[1], Marc Alier[2], Francisco José García-Peñalvo[3] and Mª José Casañ[2]

[1] *La Salle, Universitat Ramon Llull, Barcelona, Spain*
[2] *Universitat Politècnica de Catalunya, Barcelona, Spain*
[3] *Universidad de Salamanca, Salamanca, Spain*

Abstract

The debate on privacy issues in Learning Analytics processes has been going on for a long time. In academic terms, various researchers attempted to identify the origin of the problem, provide solutions, and propose alternatives. However, the problem is complex, not yet solved, and increasingly pressing and serious. We reflect on cloud computing technologies as a generator of privacy issues and new derivatives. We assume that the technology used in the cloud is aggravating the problem, not Learning Analytics itself. Considering data capitalism, we argue that it is hopelessly impossible to solve the privacy problem, nor even mitigate it, when educational institutions use data ubiquity services in the cloud. We point to the paradox of Learning Analytics as the in-compatibility factor with third-party cloud computing services, where the latter is the link to all the associated privacy issues. To mitigate privacy issues, we propose the deconstruction of cloud computing for its localization. The localization is the basis of a new concept related to the disconnection of educational institutions from the cloud. New technological perspectives, legal frameworks, and social, cultural, and political changes are required.

Keywords

Learning Analytics, Privacy, Localization, Cloud Computing, Un-plugged Institutions.

1. Introduction

We need to talk. Privacy issues in Learning Analytics processes have long been debated. In 2010, three years before Pardo and Siemens linked ethical and privacy principles to Learning Analytics [1], Bach [2] expressed some concerns about analyzing educational data: "Who should be able to access the data and view results? Which data should be reported anonymously? Which can be tagged to students for educational purposes?". Researchers have made different attempts to identify ethical and privacy problems [3] and afterward provide possible solutions [4, 5]. However, beyond identifying ethical and privacy problems related to automated data processing in education and trying to solve them, these concerns are becoming increasingly serious and pressing. It seems that we are at the same starting point. As stated by Bach in 2013, "…any institution engaging in these […] activities should plan on developing a concurrent, and systematic, conversation that addresses for the institution how it approaches the ethical and legal issues that might arise along…". Bach was right then and now, after nine years from his asseverations: we are still in those initial conversations, now trying to define the role of the Chief Privacy and Chief Ethics Officers in an undeniable "shortage of leadership" [6] (p. 113) regarding data treatment.

In this conceptual article, thus not a research paper, we briefly explore cloud computing as the problem and expose what we call unplugged institutions (from cloud computing) as a proposal for

[1]earning Analytics Summer Institute Spain (LASI Spain) 2022, June 20–21, 2022, Salamanca, Spain

EMAIL: daniel.amo@salle.url.edu (A. 1); fonsi@salle.url.edu (A. 2); granludo@gmail.com (A. 3); fgarcia@usal.es (A. 4); mjcasany@essi.upc.edu (A. 5)

CEUR Workshop Proceedings (CEUR-WS.org)

student privacy protection enhancement, thus advancing ethical and privacy concerns pointed out nine years ago.

The remainder of this paper is structured as follows: Section 2 explains how the cloud diminishes students' privacy. Section 3 presents how the cloud generates entropy in Learning Analytics processes. Section 4 reflects on the proposed concept of unplugged institutions. Finally, conclusions as Section 5 ends the paper.

2. The cloud is students' data hell

Students have a big problem regarding privacy, and academics have too much homework to be done. Alier et al. [7] reveal some challenges that they state as a "pending task" in privacy and e-learning, being the surveillance capitalism one of the most dangerous to consider. Surveillance capitalism means "obtain and analyze as much data as possible of the user's activity in order to gain the ability to influence their behavior". In education, considering definitions of Learning Analytics such as the elaborated by Bach [2] or Siemens [8], influencing the students' behavior is intended as positive; but surveillance capitalism makes it the worst practice.

2.1. Surveillance capitalism in the data capitalism era

The surveillance capitalism, under the data capitalism movement, is a long-standing problem once ago called the "dossier effect". This is defined by Goldberg et al. in 1997 as "dangerous" due "it is so easy to build a comprehensive profile of individuals," and "many will be tempted to take advantage of it" [9] (p. 104). This "dossier effect" is exemplified in the educational context as applications that extract, collect and store student data that can be used by actors without any authorization. This student data, both personal and metadata, is used without permission to, among other, create profiles to predict enrollment, to deny tuitions to those students with lower financial status or to serve targeted advertisements [7]. Data collected without permission has even been sold and used in the name of improving educational products as seen in the inBloom scandal, where "data (was) available to (New York) district-approved third parties to develop tools and dashboards so the data could more easily be used by classroom educators." [10]. inBloom organization was closed due to breaking privacy rules.

2.2. Cloud computing is everywhere

All the processes mentioned above have a common denominator: cloud computing. The concept of "cloud computing" is the abstraction of a set of real computers as interconnected and hyperconnected servers that make available the use of developments under service in high efficiency to treat a massive number of requests and data. These servers and programs are managed by third parties that follow their ideology, ideas, and policies, or worse, follow other third parties' ideologies at different levels of depth. Therefore, in this article, we avoid defining these services as solutions since most of them impose unresolved problems of great magnitude related to privacy, security, and data ethics. Hence, we refer to cloud computing as those "public cloud" services ruled by the market.

Unfortunately, educational institutions migrate their educational infrastructures to cloud computing. The reasoning is different for each one, but mostly with arguments based on economics. Educational institutions shift from active controllers to passive users during this technological transition: they go from controlling infrastructures and data to delegating computing power and data management, treatment, and storage location. All educational institutions that move their infrastructures to third-party cloud computing lose absolute control of enrolled students data (and potentially to any user stored in the database) [11] (pp. 263-265); by becoming passive users, the control of the data is inexistent, and management transferred to entities that grow in centralization and power. In the last three months of 2021, Amazon concentrated 33% of the cloud computing market share worldwide, followed by Microsoft, Google, and Alibaba. This growth in differences is constant; the shares of Microsoft and Google added together do not exceed the share of Amazon. Dindsdale makes it very clear "…the rising tide continues to lift all boats, but some are being lifted more swiftly than others." [12].

All third-party applications (cloud-based or not) mean a risk to educational data. However, on-premises third-party software and hardware, can be controlled in some manner or another by IT staff, even unplugged from the internet to avoid data transfers. Without local infrastructures, the necessary knowledge and experience in educational institutions IT staff to develop, implement, deploy, and manage technologies intended for education, will be transferred to third-party services IT staff. This fact generates an absolute dependency that grows longitudinally and exponentially. The longer the know-how remains in the hands of third parties, the higher the cost of returning the location of the infrastructures, and so power and control, to educational institutions. In a possible dystopian scenario, this outsourcing debt will be impossible to reduce or eliminate. This catastrophic scenario will consider full outsourcing of the connectivity with third-party internet-managed devices such as IoT or even those wi-fi devices that feed educational campuses with the internet. This kind of scenario would complete the transfer of power and control from local institutions to third parties, thus, generating power asymmetries. It is no longer about hardware or software beyond LMSs, but third-party applications that orbit the teaching-learning processes that store and analyze educational data. For each application used in the classroom, the risk of data leakage increases from installation to installation among students, and thus also the loss of power and control in these processes.

2.3. More actors, less privacy

The common denominator in the use of Learning Analytics is also the cloud computing associated with technologies within the Big Data movement such as IoT, Machine Learning, and Artificial Intelligence techniques that arises ethical and privacy issues [11, 13]. These technologies run in remote data centers adding actors all along with the data processing. Hence, as with many (unknown) actors, the user's privacy diminishes.

Unfortunately, the trend among institutions in education is to use third-party cloud computing services to provide educational technology solutions, thus, raising the number of (unknown) actors that may have access to data. In these terms, once again, the new wave of technological solutionism [14] proves to be a failure, now in the form of cloud computing, considering ethical and privacy issues. Hence, considering the above, it is hopelessly impossible to solve or even mitigate the privacy problem in the cloud, considering that the more actors are involved, the less privacy is ensured. We dare to make this statement since cloud computing implies permanent and uncontrolled access to information from anywhere and by unknown actors.

3. Too much entropy: Learning Analytics in the cloud paradox

Bach [2] defines Learning Analytics, in which could be the first public definition, as "…the use of advanced modeling techniques integrated with learning outcomes assessment to better understand student learning and more efficiently and meaningfully target instruction, curricula and support.". The meaning given to Learning Analytics (even considering the definition from SOLAR [15]) implies that the teacher uses (any) specific students' data to provide personalized support of any kind at the agreed time. Thus, the very genesis of the concept of Learning Analytics undermines privacy to the minimum possible.

Enabling full privacy Learning Analytics is not nourished with data. Without fully identified student data, the teacher cannot articulate directed actions but rather general ones with general effects. With no identifying data, there is no personalized Learning Analytics. Therefore, considering the micro-level, Learning Analytics consists of data exchange between teachers (as responsible for success [16]) and students.

Considering the above, using cloud computing in Learning Analytics breaks the rule of data exchange between teachers and students because data is [possibly] accessible by (unknown) third parties. In this sense, a paradox is outlined when Learning Analytics uses cloud computing. Since principals cannot reduce the actors involved, the uncertainty of access to data generates a high entropy (lack of order that let chaos in controlling users' privacy) and pushes to anonymize the data as the final solution. However, this anonymization is not possible because of the meaning given to Learning

Analytics regarding the need for identifying students' data. Hence, Learning Analytics processes require privacy, but this privacy is impossible to provide when using cloud computing due to its definition.

Although some laws, such as the GDPR (General Data Protection Regulation, Regulation (EU) 2016/679), accept the pseudonymization of data, it is technically possible that third parties use multiple attributes to end up identifying students. This is what Rocher et al. [17] conclude in its model of re-identification in incomplete and anonymized datasets with the effectiveness of 99.98%, thus, breaking the standards imposed by the GDPR. The legality is punitive and not preventive, so as a society we must rely on the acceptance and practice of this legal codex. However, countries have different legalities that can be seen as disobediences by each other, thus generating much more entropy, and a system with too much entropy collapses [18].

4. Unplugged institutions

The abovementioned is the foundation of the meaning of unplugged institutions: those totally or partially disconnected from the cloud that minimizes third-party actors not related to the institution itself. Thus, these unplugged institutions will execute Un-plugged Learning Analytics (ULA).

Prinsloo et al. [19] point out that the solution is not technological, but a social change is necessary. Society requires that the laws support its decisions. It is the society that pushes new ways of legislating in response to its social desires. Other authors, such as Alier et al. [7] point out the privacy problem at a social level: "we need to solve at a social and cultural level the ethical dilemma presented by data privacy" (p. 13). In European education, the GDPR regulates the data treatment and the data transfer to third countries. However, this legislation does not reference how to reduce actors. Thus, the current legislation cannot be a regulation for unplugged institutions. Therefore, new social, political, and cultural rules are required for those institutions [partially] free from cloud computing.

The idea of the unplugged institution requires profound social change, a new technological perspective, and new (or old) engineering to make it possible. New developments capable of reducing actors and improving student privacy in this regard will emerge from such foundations.

4.1. Turn off cloud computing

We are now in an infinite loop. As we have commented in the introduction, non-technological actions, and privacy-enhancing technologies (PET) are proposed in a techno-solutionism that leads us to an apparent advance. However, the problem remains unsolved. How can the technology that supports Learning Analytics use cloud computing to preserve the identity of students from third parties? The answer is (it seems to be) a negative due to profound uncertainty. To reduce the risk of the Learn-ing Analytics paradox, we must forbid cloud computing. This dilemma uncovers serious questions that (apparently) fight against the development of cloud computing, and perhaps this is the case in education and Learning Analytics processes (the technology per se is not an opportunity); among them, surely the most important: is it possible to reduce the high levels of entropy generated by the cloud computing in Learning Analytics?

Considering what has been stated in the previous sections, the answer to this question leads us to an irremediable total or partial disconnection of institutions from cloud computing, thus, move to unplugged institutions. If reducing the [privacy] risks [in the cloud] is unacceptable, it only remains to eliminate the risks. The way to do this is to reduce the intermediary actors in cloud technologies. The previous means getting out of the cloud as the more actors involved, the less privacy for students. It is necessary to initiate a relocation of power over technologies, services, and data. This movement implicitly implies new ways of thinking and understanding automated data processing.

5. Conclusions

The concept of unplugged institutions as a solution to (some) social, cultural, and political problems regarding data privacy is an alternative to the technical problematics created by a techno-solutionism of faith in cloud computing. The economic costs of infrastructures are reduced when migrating services to the cloud. Besides, the risks and harmful effects on students are increased. Hence, the economic reduction is subsidized by students' data.

The proposal assumes that, in the educational system, the closer the data is to third-party cloud computing, the more unknown actors will appear, and the more entropy will generate. Besides, the closer the student's data is to the teacher, the less actors will be involved, and the less entropy will arise. Therefore, we establish a connection between privacy and proximity. Reducing data proximity solves or mitigates the problem of entropy in education. We communicated it in our previous work [11] "…data proximity is therefore essential to ensure non-leakage, non-misuse, non-prohibited, or inappropriate storage, and non-processing without permission of any data generated…" (p. 270).

Without excluding cloud computing or considering an absolute localization, the proposal of unplugged institutions promotes a balance of those local and remote policies and techniques that help reduce the actors involved in Learning Analytics processes. Thus, "the unplugged institution" is not a technical proposal, although technology is an important part. As Alier et al. [7] expose, the changes must be accompanied by debate in social, cultural, and political perspective; this argument is being shared by other authors such as Prinsloo et al. [19] when they state that "we should approach data privacy as primarily a social problem" (p. 3). Likewise, we pointed out our commitment to local technologies and 7 regulatory principles [20]. All these proposals are the beginning of alternatives to solve privacy issues that open new ways of looking, knowledge, and actions.

6. References

[1] Pardo, A., Siemens, G.: Ethical and privacy principles for learning analytics. British Journal of Educational Technology. 45, 438-450 (2014). https://doi.org/10.1111/bjet.12152

[2] Bach, C.: Learning Analytics: Targeting Instruction, Curricula and Student Support. En: Proceedings of the The 8th International Conference on Education and Information Systems, Technologies and Applications: EISTA 2010. EISTA, Orlando, Florida, USA (2010)

[3] Slade, S., Prinsloo, P.: Learning Analytics: Ethical Issues and Dilemmas. American Behavioral Scientist. 57, 1510-1529 (2013). https://doi.org/10.1177/0002764213479366

[4] Amo, D.: Privacidad y gestión de la identidad en procesos de analítica de aprendizaje con Blockchain, https://repositorio.grial.eu/bitstream/grial/1951/1/TesisDanielAmoFilva-LearningAnalytics-Blockchain_v43.pdf, (2020)

[5] 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. En: Zaphiris, P. y Loannou,

[6] (eds.) International Conference on Human-Computer Interaction. pp. 3-14. Springer, Orlando, USA (2019)

[7] Timofte, R.S.: Ethics and Privacy in Learning Analytics: The Rise of Chief Privacy and Chief Ethics Officers. En: Mâță, L. (ed.) Ethical Use of Information Technology in Higher Education. pp. 113-126. Springer, Singapore (2022)

[8] Alier, M., Casany, M.J., Severance, C., Amo, D.: Learner Privacy, a pending assignment. En: Proceedings of the Eighth International Conference on Technological Ecosystems for Enhancing Multiculturality. pp. 725-729. Association for Computing Machin-ery, New York, NY, USA (2020)

[9] Siemens, G.: Learning Analytics: The Emergence of a Discipline. American Behavioral Scientist. 57, 1380-1400 (2013). https://doi.org/10.1177/0002764213498851

[10] Goldberg, I., Wagner, D., Brewer, E.: Privacy-enhancing technologies for the Internet. En: Proceedings IEEE COMPCON 97. Digest of Papers. pp. 103-109. IEEE, San Jose, CA, USA (1997)

[11] Herold, B.: InBloom to Shut Down Amid Growing Data-Privacy Concerns, http://blogs.edweek.org/edweek/DigitalEducation/2014/04/inbloom_to_shut_down_amid_growing_data_privacy_concerns.html

[12] Amo, D., Prinsloo, P., Alier, M., Fonseca, D., Kompen, R.T., Canaleta, X., Herrero-Martín, J.: Local Technology to Enhance Data Privacy and Security in Educational Technology. International Journal of Interactive Multimedia and Artificial Intelligence. 7, 262-273 (2021). https://doi.org/10.9781/ijimai.2021.11.006

[13] Richter, F.: Infographic: Amazon Leads $180-Billion Cloud Market, https://www.statista.com/chart/18819/worldwide-market-share-of-leading-cloud-infrastructure-service-providers/

[14] Alshohoumi, F., Sarrab, M., AlHamadani, A., Al-Abri, D.: Systematic Review of Existing IoT Architectures Security and Privacy Issues and Concerns. International Journal of Advanced Computer Science and Applications (IJACSA). 10, 232-251 (2019). https://doi.org/10.14569/IJACSA.2019.0100733

[15] Morozov, E.: To save everything, click here: The folly of technological solutionism. PublicAffairs, New York, NY, USA (2013)

[16] SoLAR: About SoLAR, https://solaresearch.org/about/

[17] Knight, S., Littleton, K.: Discourse-centric learning analytics: mapping the terrain. Journal of Learning Analytics. 2, 185-209 (2015). https://doi.org/10.18608/jla.2015.21.9

[18] Rocher, L., Hendrickx, J.M., de Montjoye, Y.-A.: Estimating the success of re-identifications in incomplete datasets using generative models. Nat Commun. 10, 3069 (2019). https://doi.org/10.1038/s41467-019-10933-3

[19] Franceschi-Bicchierai, L.: Facebook Doesn’t Know What It Does With Your Data, Or Where It Goes: Leaked Document, https://www.vice.com/en/article/akvmke/facebook-doesnt-know-what-it-does-with-your-data-or-where-it-goes, (2022)

[20] Prinsloo, P., Slade, S., Khalil, M.: The answer is (not only) technological: Considering student data privacy in learning analytics. British Journal of Educational Technology. 00, 1-18 (2022). https://doi.org/10.1111/bjet.13216

[21] Amo, D., Torres, R., Canaleta, X., Herrero-Martín, J., Rodríguez-Merino, C., Fonseca, D.: Seven principles to foster privacy and security in educational tools: Local Educational Data Analytics, (2020)