Guest editorial: The datafication of student life in higher education: privacy problems and paths forward

Overview of the special issue's theme

Higher education institutions continue to datafy student life in all its forms: academic, social, personal, health, etc. Some of these actions are intentional. Universities build data infrastructures to strategically capture student behaviors, communications and profiles to better serve their educational interests. Other forms of student datafication arise as a consequence of the ubiquity of information and communication technologies on campuses and how they create revelatory and analyzable data trails, which can directly or indirectly identify students. Learning sciences researchers recognize the potential to inform teaching strategies, improve learning processes and increase outcomes by studying and better understanding the relational dynamics among student data trails, cognition, affect and performance behaviors and artifacts. Higher education administrators and institutional researchers see benefits, too; for them, acting on data can improve higher education's effectiveness and create new efficiencies. But as the guest editor wrote in a Future of Privacy Forum special report:

> [. . .] these opportunities bring significant, undeniable social, political, ethical, and legal problems that education stakeholders should neither discount nor ignore. Chief among these problems is student data privacy, from which one could argue that most of these other social, political, ethical, and legal issues stem. (Jones, 2022, p. 3)

This special issue focuses on student and learner privacy as a *central problem* in the fields of educational data mining, learning analytics (LA) and information science (IS) research and practice, with an emphasis on laying out paths forward to *constructively* and *pragmatically* address privacy. Since around 2010, the early years of LA, researchers and practitioners alike have published a significant body of work:

- identifying privacy as a complicated issue (Rubel and Jones, 2016); and
- cataloging its related ethical facets (e.g. autonomy, trust, intellectual freedom) (Slade and Prinsloo, 2013; Pardo and Siemens, 2014).

This critical, ethics-forward work has been valuable in that it has enhanced sensitivities around privacy and demonstrated the role of privacy in educational practices. However, only more recently has research in the education and learning sciences and IS attempted to address how to build and implement educational technologies and data practices with privacy and consent as a central feature (Cormack, 2016; Jones, 2019; Li *et al.*, 2021; Paris *et al.*, 2022). Other research has just begun to treat faculty and students as key stakeholders whose rights and interests in particular require greater consideration and protection in the design and deployment of LA tools (Klein *et al.*, 2019; Sun *et al.*, 2019; Jones *et al.*, 2020; Mahmoud *et al.*, 2020; West *et al.*, 2020). Furthermore, research has begun to also envision faculty and students as codesigners of the tools who can agentively work alongside programmers and data scientists in designing system affordances as well as implementation, customization and deployment plans (Buckingham Shum *et al.*, 2019;

Alvarez *et al.*, 2020; Michos *et al.*, 2020; Paris *et al.*, 2022; Sarmiento and Wise, 2022). In their own way, literature in this body of research has demonstrated how to address privacy issues in a practical way.

The call for this special issue specifically requested contributions that provided practical pathways forward to further enhance this facet of the student privacy research agenda in the higher education context. We are pleased to present six articles traversing diverse themes focused on algorithmic discrimination, governance, policy development, technical approaches to privacy, student behaviors and trust. Taken as a set, we propose they constitute a contribution to the growing literature in this area, as they develop and advance key themes surrounding the datafication of student life in the context of higher education and educational technology. They importantly also raise new questions and start new pathways forward for other researchers to enter the conversation.

Summaries of included articles

Greenhalgh *et al.* (2023) "Platforms, perceptions, and privacy: ethical implications of student conflation of educational technologies", *Information and Learning Sciences*, doi: 10.1108/ILS-03-2023-0030:

> Published research on student perceptions of and expectations for how their institutions should protect their privacy has focused on student reactions to sociotechnical practices within their current university. In their article, Greenhalgh *et al.* (2023) take a different track. They begin with the premise that a student's privacy preferences are molded by experiences with educational technologies during their secondary education. Their research centers on students' use of the widely used ClassDojo platform with communication and behavior management affordances – or another platform with similar features. Greenhalgh *et al.* (2023) surveyed 528 undergraduate students in the spring and fall of 2020. Their findings suggest that students see educational technologies as tools, not platforms in which their lives are turned into datafied, analyzable objects. They conclude by emphasizing "that the collection and analysis of students' data as well as students' relative unawareness of this phenomenon both begin long before they reach higher education." Consequentially, higher education institutions "may have to contend not just with a blank slate of ignorance about these phenomena but rather with entrenched, practiced attitudes toward educational technology."

Holmes *et al.* (2023) "PIILO: an open-source system for personally identifiable information labeling and obfuscation", *Information and Learning Sciences*, doi: 10.1108/ILS-04-2023-0032:

> Anonymizing student data for research and technology development purposes is the gold standard for protecting student privacy and possible downstream harms. But, that gold standard is often impossible to achieve when data sets are replete with student identifiers from there very outset of a datum's creation. In their article, Holmes *et al.* (2023) note that "public sharing of student data is a consequential scenario because shared data sets present many valuable opportunities for replication research and, in the case of predictive analytics, the ability to benchmark new predictive models." However, while researchers desire to create, share, and study student data sets comprising of unstructured text (e.g. discussion posts) – and the intellectual behaviors and ideas expressed therein – such data sets risk putting students at risk if personally identifiably data is not sufficiently scrubbed or obfuscated; processes that have proven to be impractical if done manually and challenging if done technically. To address this problem, the authors introduce "an open-source automatic deidentification system for student text called the Personally Identifiable Information Labeling and Obfuscation (PIILO) system." PIILO uses identification and hiding in plain sight (HIPS) strategies to recognize and mask student names and other PII. The authors tested PIILO on two data sets, one consisting of student writing samples from a MOOC and the other from discussion posts in an LMS. PIILO corrected identified

96% of full names in the MOOC data set and 91% in the discussion post data set. The application of HIPS reduced – but did not eliminate – the risk of reidentifying a student.

Mann *et al.* (2023) "Tracking transparency: an exploratory review of Florida academic library privacy policies", *Information and Learning Sciences*, doi: 10.1108/ILS-04-2023-0038:

In the USA, privacy policies in academic libraries in colleges and universities are informed by long-standing professional values to protect the intellectual freedom of library users – students. But as Mann *et al.* (2023) note in their article, "there is a deep tension between the professional value of privacy in librarianship and the ubiquitous collection of data and learning analytics frequently required from higher education to show value." Their study focuses on the state of Florida and 70 of its public and private higher institutions (excluding medical, law, or faith-based institutions) to investigate the presence and content of library privacy policies. The found that only 15 of the studied institutions presented a "separate library webpage with policy-oriented text," and institutions with a research focus were more likely to have such pages. Policy pages resided in multiple, differently named locations (i.e. they were not all named something akin to "Privacy Policy"), and there was no transparency concerning student data collection that could be used for learning analytics practices. Their work underscores the need for academic libraries to make more transparent how and to what ends they use data about students, and what practices are in place to protect their privacy.

Prinsloo *et al.* (2023) "'Trust us', they said. Mapping the contours of trustworthiness in learning analytics", *Information and Learning Sciences*, doi: 10.1108/ILS-04-2023-0042:

Educational technologies, generally, and learning analytics, specifically, have and increasingly struggle with the question of whether they are trustworthy artifacts and implemented with trustworthy, justifiable ends in mind. These questions have become more pointed since the COVID-19 pandemic, when students were forced into online learning spaces, some of which exposed their personal lives with the use of webcams and proctoring applications. Khalil *et al.* (2023), in their article, emphasize that "trust has always been central to the social contract between students, communities, educational providers and governments. However, there are increasing concerns that an 'uncritical embrace of technology' subverts trust and goodwill." And while these questions of trust have not gone undiscussed in the literature, especially regarding the collection, analysis and use of student data, they argue that there is a gap regarding the "contours" of trust, or "the elements and the importance of separate and mutually constitutive elements of trust." To fill this gap, they conducted a two-round Delphi study, whereby 31 (of 99 invited) authors who have written on learning analytics and trust were surveyed. Survey findings established a common definition of trust; elements of trust; and trust factors affecting specific practices (e.g. student learning, data use, use of learning analytics). Other findings suggest elements that could improve the trustworthiness of learning analytics practices.

Sanfilippo *et al.* (2023) "Privacy governance not included: analysis of third parties in learning management systems", *Information and Learning Sciences*, doi: 10.1108/ILS-04-2023-0033:

Learning management systems (LMSs), and the third-party plug-ins and learning tools interoperabilities (LTIs) embedded within, have become commonplace educational technologies in higher education. However, Sanfilippo *et al.* (2023) note how these tools and the reliance thereon by instructors and students "was amplified throughout the COVID-19 pandemic, as digital solutionism paved the way for virtual classrooms and digital proctoring via synchronous and asynchronous means of collaboration between students and instructors." The problem for student privacy is that, they write, "higher education institutions have access to, and in many cases locally host, substantial amounts of student data on LMS[*s*] that provide third-party mechanisms to enhance interfaces, add new functionalities, and customize user experiences for specific institutions, departments, or courses. The tight integration of first and third-party tools in this ecosystem raises concerns that student data may be accessed and shared without sufficient

244

transparency or oversight." To address governance issues associated with LMSs and LTIs, Sanfilippo *et al.* (2023) conducted a multimodal study consisting of an online questionnaire collected from information technology professionals at seven universities in the USA and Canada, in-depth interviews with 25 data governance professionals and decision-makers at 14 US research universities, and documentation from 112 universities regarding LMS, plugin and LTI usage and management. Using the Governing Knowledge Commons framework in support of their analysis, Sanfilippo *et al.* (2023) found few legal protections for students concerning third-party data flows, mostly failures of governance at the institutional level (with a few notable exceptions), and opaque decision-making processes. They argue that "that student privacy is being overlooked, ignored, and, in some cases, intentionally sacrificed [....] to prioritize convenience, cost or control over the interests of student privacy." Still, opportunities exist to build on existing governance norms and rules in higher education to specifically address the problems identified with LMSs, plugins and LTIs.

Von Winckelmann (2023) "Predictive algorithms and racial bias: a qualitative descriptive study on the perceptions of algorithm accuracy in higher education", *Information and Learning Sciences*, doi: 10.1108/ILS-05-2023-0045:

Noting that predictive algorithms "have become the most common analytic tool used in higher education [. . .] and open a window into the educational lives of students," Von Winckelmann (2023) catalogs how higher education institutions have used these tools to investigate and improve student success and engagement, support alumni fundraising strategies, identify students likely to default on their student loans and target students whose timeline to graduation is slower than institutionally expected. Like with uses of predictive algorithms in other contexts, they warn that inappropriate uses of predictive algorithms "places students in historically underrepresented groups (HUGs) in a precarious position as there are significant risks of racial biases infiltrating the data." Von Winckelmann (2023) used a questionnaire and interview protocol informed by data justice theory to investigate how higher education data professionals perceive and vet the accuracy of the algorithms their institutions use. The study confirmed that participants were "aware of both systemic and racial bias in their [predictive algorithm] inputs and outputs and acknowledge their responsibility to use [predictive algorithms] recommendations ethically with students in HUGs." Among other findings, resulting practical implications from the study recommend that higher education data professionals would be well served by social justice professional education related to data practices. Furthermore, like has been found in other published studies, institutions should transparently communicate their uses of predictive algorithms to students.

Steps forward
The motivation for this special issue was, as written above, to "to address privacy issues in a practical way." Was that goal achieved? Indeed, we propose it was. Each article uniquely presents pragmatic insights into how practitioners, researchers, administrators and even technologists could affect positive change in their uses of educational technologies and student data to either reduce privacy concerns or enhance privacy protections. The articles suggest frameworks for informing data practices (Von Winckelmann, 2023), the power of existing governance norms to fill governance gaps (Sanfilippo *et al.*, 2023), a clear call to action to create and update privacy policies in response to sociotechnical changes (Mann *et al.*, 2023), advancements in privacy-protecting technologies that support research endeavors (Holmes *et al.*, 2023), new ways of thinking about student behaviors and attitudes toward learning technologies (Greenhalgh *et al.*, 2023) and a concrete definition and actionable conceptualizations of trust that can drive how LA can be implemented in trustworthy ways. There is always room for theorizing and philosophizing about the concept of privacy and its value – especially in relation to sociotechnical change – but

privacy must also be practiced. The included articles will help us all to be more informed privacy practitioners.

Kyle Jones
Luddy School of Informatics, Computing and Engineering,
Indiana University-Indianapolis, Indianapolis, Indiana, USA

References

Alvarez, C.P., Martinez-Maldonado, R. and Buckingham Shum, S. (2020), "LA-DECK: a card-based learning analytics co-design tool", *Proceedings of the Tenth International Conference on Learning Analytics and Knowledge, Association for Computing Machinery (LAK '20), New York, NY*, pp. 63-72, doi: 10.1145/3375462.3375476.

Buckingham Shum, S., Ferguson, R. and Martinez-Maldonado, R. (2019), "Human-centred learning analytics", *Journal of Learning Analytics*, Vol. 6 No. 2, pp. 1-9, available at: www.learning-analytics.info/journals/index.php/JLA/issue/view/463 (accessed 30 June 2022).

Cormack, A.N. (2016), "A data protection framework for learning analytics", *Journal of Learning Analytics*, Vol. 3 No. 1, pp. 91-106, doi: 10.18608/jla.2016.31.6.

Greenhalgh, S.P., DiGiacomo, D.K. and Barriage, S. (2023), "Platforms, perceptions, and privacy: ethical implications of student conflation of educational technologies", *Information and Learning Sciences*, doi: 10.1108/ILS-03-2023-0030.

Holmes, L., *et al.* (2023), "PIILO: an open-source system for personally identifiable information labeling and obfuscation", *Information and Learning Sciences*, doi: 10.1108/ILS-04-2023-0032.

Jones, K.M.L. (2019), "Learning analytics and higher education: a proposed model for establishing informed consent mechanisms to promote student privacy and autonomy", *International Journal of Educational Technology in Higher Education*, Vol. 16 No. 1, p. 24, doi: 10.1186/s41239-019-0155-0.

Jones, K.M.L. (2022), "The datafied student: why students' data privacy matters and the responsibility to protect it", Future of Privacy Forum, available at: www.studentprivacycompass.org/resource/the-datafied-student-why-students-data-privacy-matters-and-the-responsibility-to-protect-it/

Jones, K.M.L., *et al.* (2020), "'We're being tracked at all times': student perspectives of their privacy in relation to learning analytics in higher education", *Journal of the Association for Information Science and Technology*, Vol. 71 No. 9, pp. 1044-1059, doi: 10.1002/asi.24358.

Klein, C., *et al.* (2019), "Technological barriers and incentives to learning analytics adoption in higher education: insights from users", *Journal of Computing in Higher Education*, Vol. 31 No. 3, pp. 604-625, doi: 10.1007/s12528-019-09210-5.

Li, W., *et al.* (2021), "Disparities in students' propensity to consent to learning analytics", *International Journal of Artificial Intelligence in Education*, doi: 10.1007/s40593-021-00254-2.

Mahmoud, M., *et al.* (2020), "Learning analytics stakeholders' expectations in higher education institutions: a literature review", *The International Journal of Information and Learning Technology*, Vol. 38 No. 1, pp. 33-48, doi: 10.1108/IJILT-05-2020-0081.

Mann, E.Z., *et al.* (2023), "Tracking transparency: an exploratory review of Florida academic library privacy policies", *Information and Learning Sciences*, doi: 10.1108/ILS-04-2023-0038.

Michos, K., *et al.* (2020), "Involving teachers in learning analytics design: lessons learned from two case studies", *Proceedings of the Tenth International Conference on Learning Analytics and Knowledge, Association for Computing Machinery (LAK '20), New York, NY*, pp. 94-99, doi: 10.1145/3375462.3375507.

Pardo, A. and Siemens, G. (2014), "Ethical and privacy principles for learning analytics", *British Journal of Educational Technology*, Vol. 45 No. 3, pp. 438-450, doi: 10.1111/bjet.12152.

Paris, B., Reynolds, R. and McGowan, C. (2022), "Sins of omission: critical informatics perspectives on privacy in e-learning systems in higher education", *Journal of the Association for Information Science and Technology*, Vol. 73 No. 5, pp. 708-725, doi: 10.1002/asi.24575.

Prinsloo, P., Slade, S. and Khalil, M. (2023), *'Trust Us', They Said. Mapping the Contours of Trustworthiness in Learning Analytics*, Information and Learning Sciences, doi: 10.1108/ILS-04-2023-0042.

Rubel, A. and Jones, K.M.L. (2016), "Student privacy in learning analytics: an information ethics perspective", *The Information Society*, Vol. 32 No. 2, pp. 143-159, doi: 10.1080/01972243.2016.1130502.

Sanfilippo, M.R., *et al.* (2023), "Privacy governance not included: analysis of third parties in learning management systems", *Information and Learning Sciences*, doi: 10.1108/ILS-04-2023-0033.

Sarmiento, J.P. and Wise, A.F. (2022), "Participatory and co-design of learning analytics: an initial review of the literature", *LAK22: 12th International Learning Analytics and Knowledge Conference*, Association for Computing Machinery (LAK22), New York, NY, pp. 535-541, doi: 10.1145/3506860.3506910.

Slade, S. and Prinsloo, P. (2013), "Learning analytics: ethical issues and dilemmas", *American Behavioral Scientist*, Vol. 57 No. 10, pp. 1510-1529, doi: 10.1177/0002764213479366.

Sun, K., *et al.* (2019), "It's My data! Tensions among stakeholders of a learning analytics dashboard", *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems. CHI '19: CHI Conference on Human Factors in Computing Systems*, ACM, Glasgow Scotland UK, pp. 1-14, doi: 10.1145/3290605.3300824.

Von Winckelmann, S.L. (2023), "Predictive algorithms and racial bias: a qualitative descriptive study on the perceptions of algorithm accuracy in higher education", *Information and Learning Sciences*, doi: 10.1108/ILS-05-2023-0045.

West, D., *et al.* (2020), "Do academics and university administrators really know better? The ethics of positioning student perspectives in learning analytics", *Australasian Journal of Educational Technology*, Vol. 36 No. 2, pp. 60-70, doi: 10.14742/ajet.4653.