Title

The Temptation of Data-Enabled Surveillance: Are Universities the Next Cautionary Tale?

Authorship

Alan Rubel
University of Wisconsin-Madison
Information School
arubel@wisc.edu

Kyle M. L. Jones
Indiana University–Indianapolis (IUPUI)
School of Informatics and Computing, Department of Library and Information Science
kmlj@iupui.edu

Correspondence
All correspondence should be directed to Dr. Rubel via e-mail at arubel@wisc.edu.

Citation

Copyright

I.   Intro

Recent advances in big data have reinvigorated longstanding concerns about data-enabled surveillance. These concerns are often described as a form of "surveillance capitalism," in which data collection and reuse creates enormous corporate profits, but which individuals have little power to resist (Zuboff 2019). There is similarly troubling, but less analyzed, trend of non-profits using data-enabled surveillance in order to advance their organizations' welfare rather than profit. One example is higher education institutions (HEIs).

HEIs provide significant social good, and data analytics has become endemic in HEIs as a means to advance socially desirable ends. For example, HEIs collect, combine, analyze, and use large amounts of student data to help understand learner behaviors and contexts, to improve learning outcomes, and to increase institutional efficiency. This is often referred to as "learning analytics" (LA) (Siemens 2013). ID card swipes and wifi log-ins are used to infer student location, class attendance, use of campus facilities, eating habits, and friend groups. Course management systems capture student interactions with readings, video lectures, and discussion boards. Application materials provide demographic information. These data are combined and analyzed to identify students in need of academic or social support; predict enrollment demands; provide reports to political bodies to which HEIs are accountable; and determine which applicants are likely to enroll to more effectively target recruiting efforts.

While some of these are laudable aims, there are moral concerns with such extensive information collection, analysis, and use. Specifically, these practices raise important questions about who benefits from, and the ethical trade-offs of, large-scale data-enabled surveillance in the name of social good. HEIs often justify LA as a way to advance student interests, but particular projects appear to advance primarily organizational welfare and institutional interests. Moreover, LA seems to advance a narrow conception of student interests while discounting student interests in privacy and autonomy. To illustrate the mechanisms by which data-enabled social good can become surveillance-enabled organizational welfare, we will describe several recent cases in higher education learning analytics, explain why they are important, and then use them to discuss some concerns that should guide LA in higher education and data-enabled surveillance by non-profits.

II.   Four cases

The IOT-enabled movement tracking case. A University of Arizona researcher used student ID card swipe information to create highly detailed maps of student movements and social networks. Researchers used the maps to build a student retention model to help HEIs understand the size of students' social networks, changes in those networks, and strengths and changes in social connections. The university plans to use data from its wifi routers to form even more detailed understandings of student movements and behavior and to share this information with student advisors (Blue 2018). It is unclear whether students are aware of the extent to which their data is collected and analyzed. Note that emerging smart cities and private-public partnerships are developing similar movement tracking and enabling an unprecedented wave of government and corporate surveillance at scale.

The third-party case. In April 2018, researchers from the education and publishing company Pearson presented a paper at the annual meeting of the American Educational Research Association revealing that they had conducted an experiment using one of their learning software programs used by HEIs, which is marketed as a tool to improve student learning. The researchers incorporated encouraging, "growth mindset" messages into the software interface and tested (without students' knowledge or consent) whether placing those messages affected students' performance within the software system (Herold 2018). A number of commentators maintain that this kind of testing is itself objectionable. We won't address that here. Rather, we raise the example because it demonstrates just how valuable student information is to third parties. The premise of LA is that education data can reveal information key to understanding student performance and learning to improve student outcomes. Such information collected for social good will of course be valuable to others with a different set of goals such as corporate profit. Without strong controls, some uses of these data risk becoming part of the broader trend toward surveillance capitalism.

The benchmark case widely touted as a LA success. Georgia State University (GSU) is a public research university with a substantial proportion of students from under-represented groups. About a third of its students are the first in their families to attend college, over half are eligible for federal grants, and about 60 percent are students of color (Hefling 2019). Like many HEIs, GSU has struggled to ensure that students (and in particular students from underrepresented backgrounds) complete their degrees. Ensuring graduation is a social good, since those who do not finish are left with substantial student loan debt but without the degrees that would help secure well-paid employment.

In 2011, GSU developed a system tracking both academic and financial information that alerts students' advisors about risk factors (e.g., an unsatisfactory grade in a key course, failure to take a required course on time). GSU simultaneously hired dozens of new advisors and substantially increased student advising opportunities. GSU's six-year graduation rate rose from 48% in 2011 to 55% in 2018 (Hefling 2019). Moreover, students of color, Pell-eligible, and first-generation students now graduate at higher rates than the student body overall (Ekowo and Palmer 2016).

The GSU case is often described as a LA success, but it is less of an analytics program than a robust advising program informed by better data. Certainly, the increase in retention and graduation rates are important, but it is not clear exactly what GSU's advising interventions actually are and whether they provide the best outcomes for each student or just for GSU. Do they steer students away from challenging courses or majors, reducing student agency and potential for excellence, or do they offer tutoring services for at-risk students— increasing their agency and capabilities in the process? It is unclear that the increased student surveillance was key to improved social good, and it is not clear the extent to which GSU considered the all the relevant moral trade-offs.

The gob-smacking case of data for organizational welfare, not social good. A wildly different example comes from Mount Saint Mary's University (MSMU). The university sought to increase its student retention rates to improve its rankings. Those rankings affect application and enrollment rates, which in turn contribute to an institution's financial well-

being. The university president planned to use a personality test on incoming students to predict individual students' likelihood of dropping out. This in itself could be justified if it were combined with support for at-risk students (widely considered a social good). However, in this case students with poor scores would be encouraged to leave before they were counted as "enrolled students" for ranking systems (and hence reduce the negative effect on the school's rank). When the student newspaper asked the president about it, he suggested that one should not think of first-year students as "cuddly bunnies" and that instead "[y]ou just have to drown the bunnies … put a Glock to their heads." He resigned shortly afterward.

The use of new data sources such as personality tests to predict retention and push students out is abhorrent and raises a host of ethical issues. The point of the example, though, is that the purpose—dis-enrolling students in order to better game a news magazine's college rankings (which are themselves widely criticized as poor measures of program and institutional quality)—favored the institution's interests over that of its students. In other words, using easily-obtained data for institutional interests was (for the president) more attractive than fulfilling the university's moral responsibilities to students.

III.   Why should we care?

The above cases illustrate some of the range of issues in LA, but they can only tweak our intuitions about moral concerns in LA. It is therefore worth making explicit a number of reasons we think LA (depending on how it's developed and implemented) creates moral problems and a tempting, but dangerous course for universities and other non-profits pursuing a social good agenda.

(1) Whose interests? And for what purpose? Data analytics may provide important insights for HEIs, and they may help HEIs fulfill their responsibilities to educate students and marshal resources effectively. However, it is not at all clear that LA will live up to that promise, and it is not at all clear that *student* interests motivate LA research. The Arizona, MSMU, and Pearson cases do not appear tailored to advance student interests at all.

Moreover, universities collect and analyze data somewhat indiscriminately. Administrators developing and implementing LA projects often suggest that they wish to collect and analyze "all the information [they] can" because it is potentially relevant to HEIs' educational mission. Yet (as we have argued elsewhere) this "relevance condition" is by itself insufficient to justify data collection, analysis, and use (Rubel and Jones 2016). Any student data is potentially relevant to educational objectives (e.g., MSMU's personality tests), and it is impossible to tell a priori which will actually be useful. Hence, a collection principle based solely on potential relevance is not a limitation at all.

(2) Fostering trust and trustworthiness. A second reason we should be concerned about LA is the degree of trust that stakeholders place in HEIs. Students attend colleges and universities implicitly believing that those institutions are trustworthy, will respect them as individuals, and will not implement systems that subordinate student rights and interests for the sake of institutional or third-party goals. And yet systems currently being built and deployed create opportunities for greater privacy intrusions (ASU), misuse of information (MSMU), and benefit institutions rather than students (ASU, MSMU, Pearson). Students have little

knowledge of the degree to which their information is collected and analyzed, and they typically have no ability to opt out. It is thus an open question whether their trust is misplaced, or whether they must simply acquiesce because of the social and economic value of a college education.

(3) What benefits? A third concern is that the promise of LA may be further away than advocates suggest. The current model LA system (GSU)—the one that myriad publications point to as the exemplary case—rests substantially on advising resources with small (but important) increases in student success. And even then, it is unclear whether the success is mostly attributable to funneling students into courses that have higher success rates rather than collecting and analyzing troves of student data.

One might object here that HEIs have obligations to advance the educational interests of students at-risk of doing poorly or dropping out. Surely so. However, it does not follow that HEIs should subordinate students' privacy and autonomy interests for the sake of (speculative) retention and achievement rates. Student support (social, advising, tutoring, financial, mental health) should come first, and long before impinging on other interests.

(4) Full account of student interests. We should also be wary of narrowly construing student interests. The above cases focus on student academic achievement. That's certainly important, but students have many other interests as well. They have an interest in privacy, and they should not have to forego that interest for a marginal (and at this point speculative) return in academic achievement. This should seem familiar to any professionals who bristle at overweening surveillance by supervisors.

(5) Full range of higher education's aims. Higher education has a number of aims, including (among others) fostering communication, critical thinking, understanding and appreciation of diversity, ability to live in a global society, and development of rewarding employment and careers (see Bok 2006). The goals in this list are constituted in large part by fostering students' development of their intellectual independence, their own sense of what is valuable, their capacity to act according to their own reasons and values, and their ability to arrive at, and abide by, fair terms of social cooperation. In other words, a key element of the proper aims of higher education is to help students develop and negotiate their autonomy. Yet close monitoring of student movement, social networks, and daily habits is an imposition on student privacy which is itself a key element in developing and exercising individual autonomy. In other words, to the extent that we value student autonomy, we ought to curtail student surveillance.

(6) Third parties. The extent and depth of students' social networks, their travel around campus, their health-, political-, and religious-related activities are inferable from movements, and may well be predictive of some things. But if that information is valuable for HEIs, it is also likely to be of interest to others: potential employers, the National Security Agency (pursuant to business records requests under Patriot Act section 215), software vendors, and others. While there are some legal protections in place for student information, FERPA (e.g.) allows student inspection rights, which enables them to "agree" to give records to employers. Moreover, the greater amount of student data collected creates a

greater opportunity for data breaches. In other words, LA carries risks that we should account for in determining whether it overall advances student interests.

IV. Conclusion

To be clear, we are *not* opposed to learning analytics tout court. There is no question that some data collection about students in HEIs will advance legitimate educational goals while still respecting student privacy and while fostering autonomy. However, as in other settings, LA and other data-enabled forms of surveillance that are developed for social good can all too easily slide into morally suspect territory. Just as with smart cities and in corporations, institutional interests and student interests are not identical, and we should not assume that they align. Governments, corporations, and HEIs should avoid data collection and analysis as a matter of convenience (e.g., University of Arizona) and for specious reasons (MSMU). Similarly, when they aim to advance student educational interests, they should primarily do so in ways that are consistent with (rather than contrary to) student interests in privacy and autonomy. Governments, corporations, and HEIs should also be transparent about data collection and provide opportunities to opt out. In short, HEIs should conduct LA in ways that justify the substantial trust that students place in them. A proactive and justifiable ethical stance is key to keeping systems for data-enabled surveillance focused on social good.

References

Blue, A. Researcher looks at 'digital traces' to help students. University of Arizona News. (Mar. 7, 2018); http://bit.ly/31rCzVP.

Bok, D.C. *Our Underachieving Colleges: A Candid Look at How Much Students Learn and Why They Should be Learning More.* Princeton University Press, Princeton, NJ, 2006.

Ekowo, E. and Palmer, I. The promise and peril of predictive analytics in higher education: A landscape analysis. *New America* (Oct. 24, 2016).

Harwell, D. Colleges are turning students' phones into surveillance machines, tracking the locations of hundreds of thousands. *Washington Post.* (Dec. 24, 2019); https://wapo.st/3bjnsSX

Hefling, K. The 'Moneyball' solution for higher education. Politico. (Jan. 16, 2019).

Herold, B. Pearson tested 'social-psychological' messages in learning software, with mixed results. *Education Week* (Apr. 17, 2018); http://bit.ly/39bKuJw.

Jones, K.M.L., Rubel, A., and LeClere, E. A matter of trust: Higher education institutions as information fiduciaries in an age of educational data mining and learning analytics. *Journal of the Association for Information Science & Technology* (2019), 1–15.

Rubel, A. and Jones, K.M.L. Student privacy in learning analytics: An information ethics perspective. *The Information Society 32*, 2 (Feb. 2016), 143–159.

Zuboff, S. *The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power.* Public Affairs, New York, NY, 2019.