

Datafication of Education and Algorithm (Un)fairness in Education: Another Form of Oppression and Repression

by
Ulaş Başar Gezgin

Abstract

Critical data literacy and datafication of education are reviewed. The wider context of datafication is surveillance capitalism. Educational data and learning analytics are commercialized to make a profit, which poses a challenge to the public interests of the educational system. We propose that educational phenomena that cannot be datafied are just as important as datafied phenomena. Although data utopians celebrated the emergence of big data practice, the tragic reality is closer to what data dystopians feared. The paper concludes with suggestions for future research.

Keywords: *Critical data literacy, datafication of education, surveillance capitalism, big data, algorithm fairness.*

Ulas Basar Gezgin is a social science researcher and an academic with more than 20 years of teaching experience in various institutions in Turkey, Vietnam, Thailand and Malaysia, and research experience in New Zealand (PhD work), Australia (joint project) and Latin America (journalism). He holds degrees in education, psychology, cognitive science, and urban planning. As of 2020, he has 109 books, a high number of book chapters and journal articles in various fields. A novel, an opera libretto, edited collections, poetry, a compilation of his stories, textbooks, theatre plays, scenarios and translations etc. are among his works. Some of his works have been translated to 12 languages (Turkish, English, French, German, Spanish, Italian, Russian, Japanese, Vietnamese, Thai, Georgian and Azerbaijani). He can be reached at ulasbasar@gmail.com

Introduction

Datafication in and of education is both old and new: “The datafication of education ... is part of a series of historical developments in statistics, state power, quantification, computation, and valuation culminating with the expansion and intensification of digital information systems and ‘big data’” (Williamson, 2019, p. 3). Williamson, Bayne, and Shay (2020) define datafication as “the rendering of social and natural worlds in machine-readable digital format.” Stevenson (2017) defines datafication of teaching as the educational process “increasingly transformed into numbers that allow measurement, comparison, and the functioning of high-stakes accountability systems linked to rewards and sanctions” (p. 537). Datafication in education is so pervasive that even preschool education is not immune to it. Roberts-Holmes (2018) discusses how datafication and accompanying “schoolification” undermine preschool education.

Datafication in Education

In the current education system, every student behavior is recorded and datafied, which leads to stigmatization in many cases. According to Pierlejewski (2019), data “functions as a regulatory device to objectify and control both teachers and children” (p. 1). Surveillance technologies have oppressive potential as well, regardless of how benignly they are described: “Digital data practices may allow for new forms and possibilities of monitoring and surveillance, while at the same time promoting transparency” (Jarke & Breiter, 2019, p. 4).

Against such datafication, Hillman (2022) suggests data privacy literacy education which can be subsumed under critical data literacy education (see Raffaghelli et al., 2020). Markham (2019) reconsiders critical pedagogy “as a response to datafication” (p. 754) and Pronzato (2021) considers it as “a practice of resistance to algorithms” (np). On the opposite

side, Mertala (2020) defines “data (il)literacy” “as an uncritical, one-dimensional understanding of data and datafication” (p. 30).

Current datafication policies are ignorant of epistemic quality of education (Hayes & Cheng, 2020). All the data is quantified to allow for widescale computations. It has to be noted that “datafication brings the risk of pedagogic reductionism as only that learning that can be datafied is considered valuable” (Williamson et al., 2020, p. 358). In contrast, pedagogy is not reducible to information transactions (Lundie, 2016). As a result of datafication, the teachers are “devalued, demoralized and disappearing” (Daliri-Ngametua & Hardy, 2022, p. 102). The negative effects of datafication are mainly data surveillance and algorithmic biases (Raffaghelli & Stewart, 2021). According to Alevizou (2017), educational technology moved from mediation to datafication. Fawns, Aitken, and Jones (2021) caution that “giving too much weight to outcome measures, such as grades, retention, employment, or salary, risks marginalizing valuable, yet less visible, forms of student and teacher practices” (p. 71).

Datafication of education is biased, as the underlying data is biased and nothing is done to avoid bias in datafication. Discrimination on the basis of age, gender, ethnicity, nationality, religion, gender orientation, region, etc. contaminates the educational data. What is not datafied is equally important with what is datafied. For example, how many students are denied their right to education in their native language is not considered worthy of datafication. In this context, critical race theory is not sufficient for critical data literacy,¹ as it ignores other forms of repression and discrimination such as class. Undemocratic practices as such cloud the past data which is fed backward into the datafying system. The irony is the following: “Data and algorithms are invested with promises of objectivity and impartiality, at a time when human

¹ For an attempt at connecting critical race theory and critical data literacy, see Johnson et al., 2021.

experts are not necessarily to be trusted because they are too clouded by subjective opinion, bias and partiality” (Williamson, 2019, p. 11).

Educational data is “in the form of marks, student satisfaction ratings, workload surveys, attendance monitoring, sickness absence reporting, administration of student support systems, marking turnaround times, learning analytics, and more” (Fawns et al., 2021, p. 67). In that sense, data excludes reasons for non-attendance and failure due to social injustices. Success as well is explained as an individual achievement, while some students are more privileged due to their social standing. The system appears as if failure is not systemic but individualized.

The system, therefore, has self-serving biases. Capitalism as a system is considered to be behind success stories, while the majority who can’t succeed is explained by individual incompetencies. Likewise, student satisfaction surveys (which are much like consumer satisfaction tools) are problematic since they are open to subjective influences (Gezgin, 2011). Thus, assuming the objectivity—and therefore the supremacy—of data is not justified. It is obvious that “datafication reinforces and reproduces historical inequalities” through “digital redlining,” that is, the denial of services or the targeting of particular marginalized groups (Williamson et al., 2020, p. 360). It leads to a new digital divide and thereby increases inequalities (Jarke & Breiter, 2019).

Despite of the fact that “universities are key drivers to foster data literacies and must enable students and educators to challenge biased metrics, unethical uses of data, violations of privacy and the interaction of the datafied society with the quantified self” (Atenas et al., 2020, p. 8), critical approaches to learning analytics are lacking. Loftus and Madden (2020) claim that they potentially provide cognitivistic critique of the behavioristic model underlying learning analytics. However, this requires a full-fledged discussion.

Another concept to consider is data activism in education (cf. Gezgin, 2019, 2020). In general, data activism is presented and discussed in a number of works (e.g., Aguerre & Tarullo, 2021; George & Leidner, 2019; Kennedy, 2018; Meng & DiSalvo, 2018), but implications for education are understudied. An exception is Raffaghelli, Atenas, and Havemann (2019).

Critical Data Literacy in Education

Critical data literacy involves both using and understanding data as well as a form of technology criticism (Van Audenhove et al., 2020). There is a widespread (mis)understanding that critical data literacy involves mostly user privacy (Sander, 2020a). However, context is important in critical data literacy (Agesilaou & Kyza, 2021; Gebre, 2022; Verständig, 2021): “Data are always embedded into social, cultural and historical contexts. Examining the contexts of data can be linked to critical thinking and reflection on the circumstances under which the data has been collected” (Verständig, 2021, p. 6). The wider context of data is surveillance capitalism (see Raffaghelli, 2020; Zuboff, 2019).

Critical (big) data literacy has its roots in critical media literacy and critical digital literacy research (Sander, 2020b). Tygel and Kirsch (2016) connect Paulo Freire’s critical pedagogy with critical data literacy (cf. Freire & Macedo, 1998). In parallel with Paulo Freire’s critical pedagogy, critical data literacy allows context interpretation like the rate of concentrated land ownership, questioning of common-sense concepts as GDP as a non-indicator of income inequalities, and development of new concepts such as uses of data mapping (Tygel & Kirsch, 2016). Critical data literacy is *sine qua non* as “controversy has arisen over automated recruitment systems, where applications for jobs are screened without human oversight, because they are found to disadvantage applicants from already under-

represented groups, based on previous training data showing that predominantly white male applicants perform more highly” (Williamson et al., 2020, p. 360).

When we talk about critical data literacy in education, usually we mean students’ literacy. In a rare study, Raffaghelli and Stewart (2020) investigate teachers’ critical data literacy on which the research is quite limited, despite of a number of research studies about teachers’ perspectives on datafication (Ali, 2022; Roberts-Holmes & Bradbury, 2016; Neumann, 2019; Takayama & Lingard, 2019). In fact, critical data literacy should involve all stakeholders of education, including parents. On the other hand, we need to note that most data research concerning teachers is not critical at all, but mainstream (Raffaghelli & Stewart, 2020).

Tygel and Kirsch (2016) state that

Data are not neutral. The seducing precision and objectivity of data grounded statements almost always hide ideologies and intentions about anything one wants to prove. Thus, it is fundamental to problematize the origin of data. Are data from the government or from civil society organizations? What was the political position of that organization at the time when data were generated? If it is about scientific data, who funded the research? More complex, but also of great importance, is the knowledge of the methodology used to gather data. Lack of awareness of the methodological approach can lead to misunderstandings and flawed conclusions (p. 116).

Accordingly, we have to ask various questions for an educational context. The major data collector is the government, but there may be misrepresentations as well as omissions. Educational data are rarely collected by civil society organizations, which is an obvious gap in

the field. Funding source makes a difference as well. Methodological considerations are usually out of scope of the discussions.

Converging with and in addition to Tygel and Kirsch (2016) and D'Ignazio (2017) is the idea of data biography, which involves the following questions:

- Who collected the data?
- How did they collect it?
- For what purpose?
- How is it used? By whom?
- What are its impacts? On whom?
- What are the known limitations? (p. 10).

So, we must ask how educational data is used. As it is used to make macro-level educational decisions, this is especially important. The impacts are important as well. Non-representation of ethnic as well as educational minorities makes a difference, while limitations have a lot to do with the methodology employed.

Raffaghelli and Stewart (2020) classify three types of epistemologies with regard to data: (1) Reactive, in which datafication is seen in a dystopic way, where the people take action to defend privacy and avoid personal data collection or tracking. It is a defensive positioning. (2) Proactive, in which datafication is seen as utopia, and people make every effort to capture data value. The focus of activity is data handling, collection, extraction, visualization, communication and integration into technical and technological innovation. (3) Complex, in which datafication is considered as layered system, where reactive and proactive epistemologies are embedded, but there is an effort of social and cultural contextualization that leads to actors understanding of all possible scenarios and to search and decide own best way (p. 443).

For data utopians, data (and AI) would solve all social problems. This is sometimes called as “data mythology” and it is the mainstream approach. Data dystopians, however, are in the minority (Gezgin, 2021). They are the most vocal critiques of datafication, although the resources to propagate their ideas are much more limited compared to the utopians’. Louie (2022) rightly points out that “People with critical data literacy are alert to the personal and social harms that powerful interests can inflict with data” (p. 2). Van Audenhove, Van den Broeck, and Mariën (2020) propose that “the role of media and data literacy is threefold: (1) to empower the individual; (2) to protect the individual; and (3) to ensure that the individual has the choice and space to act” (p. 4).

Gebre (2022) reviews four conceptualizations of data literacy:

- “Developing competency”: “A focus on skill set and often technical/procedural aspects of data literacy”
- “Data-driven inquiry process / thinking with data”: “Focuses on using data in inquiry processes (asking research questions and answering them using data)”
- “Personal data awareness”: “Raising awareness, concern about personal data and security”
- “Civic engagement”: “Empowering citizens and communities. Using data to address issues of community and personal relevance” (p. 1084).

Usually, technical aspects are taught at STEM courses in high schools and engineering courses at universities. Personal privacy issues, as mentioned before, are the best-known issue with regard to data literacy, whereas civic engagement involves some forms of data activism. On the other hand, asking data-relevant research questions may be coupled with

misconceptions if the data is not offered with criticism, accepting it as it is. The question is how to empower stakeholders in education through data.

D'Ignazio and Bhargava (2016) propose that

data literacy includes the ability to read, work with, analyze and argue with data as part of a larger inquiry process. Reading data involves understanding what data is, and what aspects of the world it represents. Working with data involves acquiring, cleaning, and managing it. Analyzing data involves filtering, sorting, aggregating, comparing, and performing other such analytic operations on it. Arguing with data involves using data to support a larger narrative intended to communicate some message to a particular audience. (p. 84)

In that sense, numbers should make social sense, rather than raw characterizations of the social phenomena. Arguing with educational data should start with the awareness of data and data collection, as many of the stakeholders of educations are unaware of the data process. Furthermore, data literacy in this understanding of the term requires advanced computer skills which students, parents and even teachers often do not have.

According D'Ignazio, and Bhargava (2015), (big) data literacy involves:

- Identifying when and where data is being passively collected about your actions and interactions.
- Understanding the algorithmic manipulations performed on large sets of data to identify patterns.

- Weighing the real and potential ethical impacts of data-driven decisions for individuals and for society (p. 2).

They state that usually the weakest among the data scientists is the third one (D'Ignazio & Bhargava, 2015). As the users are passive in datafication process, we need to activate them, mobilize their actions toward a higher-level understanding of the educational data. One example is university rankings, which affect parental and student decisions for university entrance. The ranking criteria should be transparent so that stakeholders can have an insider view of such rankings. One of lessons to be drawn from such a process is the fact that universities are getting increasingly less of public interest and are becoming more entrepreneurialized.

Hautea et al. (2017) find that

- (1) data collection and retention have privacy implications,
- (2) data analysis requires skepticism and interpretation,
- (3) data can come with assumptions and hidden decisions,
- (4) data-driven algorithms cause exclusion, and
- (5) measuring and reporting on data can affect the system that created the data (p. 920).

The question is who is excluded in educational data. Quite often, educational data involves schools' ignoring informal and non-formal forms of education (Gezgin, 2015). Education is an immanent human activity in all spheres of life. Data definitely requires skepticism as the idea of data biography showed us.

Other terms to be applicable to our discussion on datafication of education are algorithm (un)fairness and algorithm (in)justice. However, research on educational implications of these is virtually non-existent. The work of Vallejos et al. (2017) on young people's approaches to

algorithm fairness is relatively early and exceptional; however, pedagogical discussion is lacking. Likewise, McDonald and Pan (2020) investigate students' views on AI ethics including algorithm fairness.

In contrast, we have many research studies on algorithmic literacy, but not necessarily concerning education. Bakke (2020) develops a reflection assignment to teach algorithmic literacy, but pedagogical discussion of algorithm unfairness is lacking again. In a rare discussion, Ciccone (2021) presents the place and role of algorithmic literacy in K-12 curriculum and why it is not always covered. Ciccone (2021) states that we should start from teachers' algorithmic literacy before teaching it to students. However, since it is a political matter, it is not popular among teachers (Ciccone, 2021). In this vein, in their seminal work, Dasgupta and Hill (2021) call "the intellectual tools that allow children to understand and critique the algorithmic systems that affect their lives critical algorithmic literacies" (p. 2).

Conclusion

More research is necessary for different aspects of datafication in education, especially about the notion of algorithmic unfairness in education. More research on “dataveillance”—surveillance through data—is needed. In critical data literacy research, stakeholders of education other than students, especially teachers and parents, need to be considered and targeted for literacy education.

References

- Agesilaou, A., & Kyza, E. A. (2021). Empowering students to be data literate: The design and implementation of a learning environment to foster critical data literacy. In Proceedings of the 15th International Conference of the Learning Sciences-ICLS. International Society of the Learning Sciences.
- Aguerre, C., & Tarullo, R. (2021). Unravelling resistance: Data activism configurations in Latin American civil society. *Palabra Clave*, 24(3).
- Alevizou, Giota (2017). From mediation to datafication: theorizing evolving trends in media, technology and learning. In Ferreira, G. M. S.; Rosado, L. A. S. and Carvalho, J. S., Eds. *Education and technology: Critical approaches*. Rio de Janeiro: Editora UNESA/SESES, pp. 332–357.
- Ali, M. M. (2022). Datafication, teachers' dispositions and English language teaching in Bangladesh: A Bourdieuan Analysis. *Tesol Quarterly*.
- Atenas, J., Havemann, L., & Timmermann, C. (2020). Critical literacies for a datafied society: Academic development and curriculum design in higher education. *Research in Learning Technology*, 28.
- Bakke, A. (2020). Everyday Googling: Results of an observational study and applications for teaching algorithmic literacy. *Computers and Composition*, 57, 102577.
- Ciccione, M. (2021). Algorithmic literacies: K-12 realities and possibilities. In *Algorithmic Rights and Protections for Children*. PubPub.
- Daliri-Ngametua, R., & Hardy, I. (2022). The devalued, demoralized and disappearing teacher: The nature and effects of datafication and performativity in schools. *Education Policy Analysis Archives*, 30, 102-102.

- Dasgupta, S., & Hill, B. M. (2021). Designing for critical algorithmic literacies. In *Algorithmic Rights and Protections for Children*. PubPub.
- D'Ignazio, C. (2017). Creative data literacy: Bridging the gap between the data-haves and data-have nots. *Information Design Journal*, 23(1), 6-18.
- D'Ignazio, C., Bhargava, R. (2016). DataBasic: Design principles, tools and activities for data literacy learners. *The Journal of Community Informatics*, 12(3), 83-107.
- D'Ignazio, C., & Bhargava, R. (2015, September). Approaches to building big data literacy. In *Proceedings of the Bloomberg Data for Good Exchange Conference* (p. 6).
- Fawns, T., Aitken, G., & Jones, D. (2021). Ecological teaching evaluation vs. the datafication of quality: Understanding education with, and around, data. *Postdigital Science and Education*, 3(1), 65-82.
- Freire, A. M. A., & Macedo, D. (1998). *The Paulo Freire Reader*. Cassell and Continuum.
- Gebre, E. (2022). Conceptions and perspectives of data literacy in secondary education. *British Journal of Educational Technology*, 53,1080–1095.
- George, J. J., & Leidner, D. E. (2019). From clicktivism to hacktivism: Understanding digital activism. *Information and Organization*, 29(3), 100249.
- Gezgin, U.B. (2021). Zeka Feşizmine Karşı: Yapay Zeka, Büyük Veri, Bilişsel Bilim ve Gelecekbilim Üzerine [Against Intelligence Fetishism: On Artificial Intelligence, Big Data, Cognitive Science and Futurology]. İstanbul: Yar.
- Gezgin, U. B. (2020). An invitation to critical social science of big data: From critical theory and critical research to omni-resistance. *AI & SOCIETY*, 35(1), 187-195.
- Gezgin, U. B. (2019). Data activism: Reviving, extending and upgrading critical citizenship education and consumer rights movements. *Connectist: Istanbul University Journal of Communication Sciences*, (56), 67-86.

- Gezgin, U. B. (2015). Çokkültürlü Eğitim [Multicultural Education]. Ankara: Ütopya Yayınları.
- Gezgin, U. B. (2011). Potential problems of student evaluation of teaching (SET) in off-shore campuses in southeast and East Asia and suggestions. *Journal of Higher Education Theory and Practice*, 11(2), 90-102.
- Hautea, S., Dasgupta, S., & Hill, B. M. (2017, May). Youth perspectives on critical data literacies. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems* (pp. 919-930).
- Hayes, A., & Cheng, J. (2020). Datafication of epistemic equality: Advancing understandings of teaching excellence beyond benchmarked performativity. *Teaching in Higher Education*, 25(4), 493-509.
- Hillman, V. (2022). Data privacy literacy as a subversive instrument to datafication. *International Journal of Communication*, 16, 22.
- Jarke, J., & Breiter, A. (2019). The datafication of education. *Learning, Media and Technology*, 44(1), 1-6.
- Johnson, B., Rydal Shapiro, B., DiSalvo, B., Rothschild, A., & DiSalvo, C. (2021, May). Exploring approaches to data literacy through a critical race theory perspective. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* (pp. 1-15).
- Kennedy, H. (2018). Living with data: Aligning data studies and data activism through a focus on everyday experiences of datafication. *Krisis: Journal for Contemporary Philosophy*, 2018(1), 18-30.
- Loftus, M., & Madden, M. G. (2020). A pedagogy of data and artificial intelligence for student subjectification. *Teaching in Higher Education*, 25(4), 456-475.

Louie, J. (2022). Critical data literacy: Creating a more just world with data.

<https://www.nationalacademies.org/documents/embed/link/LF2255DA3DD1C41C0A42D3BEF0989ACAECE3053A6A9B/file/D16254F310D01BBDA873920E4EFB8151F2D8334181AA>

Lundie, D. (2016). Authority, autonomy and automation: The irreducibility of pedagogy to information transactions. *Studies in philosophy and education*, 35(3), 279-291.

Markham, A. N. (2019). Critical pedagogy as a response to datafication. *Qualitative Inquiry*, 25(8), 754-760.

McDonald, N., & Pan, S. (2020). Intersectional AI: A study of how information science students think about ethics and their impact. *Proceedings of the ACM on Human-Computer Interaction*, 4(CSCW2), 1-19.

Meng, A., & DiSalvo, C. (2018). Grassroots resource mobilization through counter-data action. *Big Data & Society*, 5(2), 2053951718796862.

Mertala, P. (2020). Data (il)literacy education as a hidden curriculum of the datafication of education. *Journal of Media Literacy Education*, 12(3), 30-42.

Neumann, E. (2021). Setting by numbers: datafication processes and ability grouping in an English secondary school. *Journal of Education Policy*, 36(1), 1-23.

Perez Vallejos, E., Koene, A., Portillo, V., Douthwaite, L., & Cano, M. (2017, June). Young people's policy recommendations on algorithm fairness. In *Proceedings of the 2017 ACM on web science conference* (pp. 247-251).

Pierlejewski, M. (2020). The data-doppelganger and the cyborg-self: Theorising the datafication of education. *Pedagogy, Culture & Society*, 28(3), 463-475.

Raffaghelli, J. E. (2020). Is data literacy a catalyst of social justice? A response from nine data literacy initiatives in higher education. *Education Sciences*, 10(9), 233.

- Raffaghelli, J. E.; Atenas, J. and Havemann, L. (2019). Data activism as scholarly pursuit: Nurturing open connections and collaboration against the domination of datafication in higher education. In OE Global Conference 2019, 26-28 Nov 2019, Milan.
- Raffaghelli, J. E., Manca, S., Stewart, B., Prinsloo, P., & Sangrà, A. (2020). Supporting the development of critical data literacies in higher education: Building blocks for fair data cultures in society. *International Journal of Educational Technology in Higher Education*, 17(1), 1-22.
- Raffaghelli, J. E., & Stewart, B. (2020). Centering complexity in “educators’ data literacy” to support future practices in faculty development: A systematic review of the literature. *Teaching in Higher Education*, 25(4), 435-455.
- Raffaghelli, J., & Stewart, B. (2021, December). The educators’ datafied present and future: Complexity as an approach to developing educators’ data literacies. In *The Open/Technology in Education, Society, and Scholarship Association Conference* (Vol. 1, No. 1, pp. 1-8).
- Roberts-Holmes, G. (2018). The “datafication” of early years pedagogy: “If the teaching is good, the data should be good and if there’s bad teaching, there is bad data.” In *Governing by Numbers* (pp. 4-17). Routledge.
- Roberts-Holmes, G., & Bradbury, A. (2016). Governance, accountability and the datafication of early years education in England. *British Educational Research Journal*, 42(4), 600-613.
- Sander, I. (2020a). Critical big data literacy tools—Engaging citizens and promoting empowered internet usage. *Data & Policy*, 2.
- Sander, I. (2020b). What is critical big data literacy and how can it be implemented? *Internet Policy Review*, 9(2), 1-22.

- Stevenson, H. (2017). The “datafication” of teaching: Can teachers speak back to the numbers? *Peabody Journal of Education*, 92(4), 537-557.
- Takayama, K., & Lingard, B. (2019). Datafication of schooling in Japan: An epistemic critique through the “problem of Japanese education.” *Journal of Education Policy*, 34(4), 449-469.
- Tygel, A. F., & Kirsch, R. (2016). Contributions of Paulo Freire for a critical data literacy: A popular education approach. *The Journal of Community Informatics*, 12(3), 108-121.
- Van Audenhove, L., Van den Broeck, W., & Mariën, I. (2020). Data literacy and education: Introduction and the challenges for our field. *Journal of Media Literacy Education*, 12(3), 1-5.
- Verständig, D. (2021, August). Critical data studies and data science in higher education: An interdisciplinary and explorative approach towards a critical data literacy. In *Seminar.net* (Vol. 17, No. 2).
- Williamson, B. (2019). Datafication of education: A critical approach to emerging analytics technologies and practices. In *Rethinking pedagogy for a digital age* (pp. 212-226). Routledge.
- Williamson, B., Bayne, S., & Shay, S. (2020). The datafication of teaching in higher education: Critical issues and perspectives. *Teaching in Higher Education*, 25(4), 351-365.
- Zuboff, S. (2019). *The age of surveillance capitalism: The fight for a human future at the new frontier of power*. Public Affairs.