Journal of Ethics and Emerging Technologies

IEET

Book Review

Becoming Data Equals: Democratic Equality, Educational AI, and the Politics of Data Design

Colin Koopman, *Data Equals: Democratic Equality and Technological Hierarchy* (Oxford, UK: Oxford University Press, 2025).

Marietjie Botes[1]*

[1] University of Texas at Austin, Department of Psychology, United States, marietjie.botes@austin.utexas.edu.
* Correspondence: marietjie.botes@austin.utexas.edu

Abstract: In, *Data Equals: Democratic Equality in Data Design*, Colin Koopman investigates the impact and downstream consequences of artificial intelligence (AI) on data infrastructures in educational systems, democratic governance, and normative politics. Koopman's main argument is that democratic equality must be actively built into data formats, classificatory systems and technical architectures from the outset, as opposed to treating them as neutral tools that often require downstream ethical correction. To illustrate this, Koopman focuses on educational AI and how personalized learning systems, that rely on "learner models" to render students computationally legible through selective features, simply manage to exacerbate existing inequalities and pedagogical assumptions. He then offers a "format anatomy" as a method to shift this critique upstream to the level of database schema and data ontologies to show how these structures privilege individualized rather than relational concepts of learning. This book review will discuss the distinction Koopman draws between technologies of separation and technologies of collaboration and how this offers a powerful framework for rethinking educational technology beyond current narrow concern such as bias, privacy, and efficacy. Although Koopman's data design agenda remains predominantly conceptual, it does make a compelling contribution by advancing a democratic theory of data infrastructure.

Keywords: educational AI; democratic equality; data design; learner models; personalized learning; data ethics.

Citation: Botes, Marietjie. 2026. Becoming Data Equals: Democratic Equality, Educational AI, and the Politics of Data Design. *Journal of Ethics and Emerging Technologies* 36: 2. https://doi.org/10.55613/jeet.v36i2.247

Received: 09/06/2026
Accepted: 10/06/2026
Published: 01/07/2026

Publisher's Note: IEET stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Much of the current conversation around AI and data ethics revolves around a familiar set of concerns: bias, privacy, fairness, and transparency (Khan et al. 2021; Whittlestone et al. 2019). Colin Koopman's *Data Equals: Democratic Equality in Data Design* asks a deeper, and in many ways more unsettling, question. Rather than asking only whether data systems can be made more ethical, Koopman asks what happens when data infrastructures begin to shape the very terms on which people are known, classified, and acted upon. His answer is both ambitious and original. If data systems increasingly structure what people can do, how institutions treat them, and even who they can become, then democratic equality must be pursued not only in law, politics, and social life, but within data itself. This is what he calls data equality. In a field saturated with calls for "ethical AI," *Data Equals* stands out because it refuses to treat data as a neutral substrate awaiting moral correction. Instead, Koopman asks how the very formats through which people are rendered computationally knowable can either harden hierarchy or help sustain more egalitarian forms of collective life.

The book is at its most persuasive when Koopman develops what he calls a "format anatomy," a method for showing how data formats, classificatory structures, and technical representations do not merely describe social life but actively help organize it. This is where *Data Equals* moves beyond familiar critiques of bias and procedural fairness and into more fertile territory. Koopman's claim is not simply that data can be used unjustly, but that the forms of data themselves help shape the conditions under which people and institutions encounter one another. That argument is grounded in his account of democratic equality as a fundamentally relational ideal. In one of the book's most important theoretical chapters, he argues that equality is not exhausted by distribution or formal participation but depends on whether people can stand in relations of equality with one another and inhabit institutions structured around mutual standing rather than hierarchy (Koopman 2024, 79–80). For readers working at the intersection of law, ethics, and governance, this is where the book becomes especially valuable. It dissolves any easy separation between technical design and institutional form. Once equality is understood relationally, data design can no longer be treated as a merely technical matter. It becomes part of the architecture of democratic life itself.

That broader theoretical move is what makes Koopman's later chapters on educational AI so compelling. *Data Equals* refuses to let debates about educational technology remain confined to the usual questions of bias, efficiency, or improved learning outcomes. Instead, Koopman asks more demanding democratic questions: What kinds of social relations do these systems build into educational environments; and What kinds of learners do they make possible? This is where the book is at its most original. His concern is not simply whether educational data systems generate inaccurate or unfair outputs, but whether they organize learners, and ultimately learning itself, in ways that erode the conditions for democratic equality. The central issue, in other words, is not whether AI-driven systems improve performance or narrow measurable disparities, but whether so-called "datafied" educational infrastructures cultivate learners as participants in shared relations of equality. That shift in emphasis is one of the book's most valuable contributions. It moves the discussion of educational AI away from downstream performance metrics and toward the more fundamental question of whether data-intensive institutions can sustain genuinely democratic forms of education.

The book becomes even more interesting when Koopman turns from this broader democratic framing to the internal logic of learner models themselves. What counts as a meaningful learner characteristic, he shows, is never simply a matter of technical optimization. It depends on prior assumptions about what can be measured, what kinds of educational theories are taken seriously, and what sorts of learners the system is designed to recognize. His discussion of feature selection is especially sharp. The continued presence of concepts such as "learning styles" within personalized learning systems is more than an amusing oversight. It points to a deeper problem, namely that the scientific sheen of personalization can rest on categories far shakier than their computational presentation suggests. Koopman is particularly good at showing how the language of precision and adaptivity can sometimes promise more than the underlying pedagogy can comfortably sustain.

To my mind, however, the book's most original contribution lies in what Koopman calls a "format anatomy" of learner models. Rather than locating the politics of educational AI only in downstream algorithmic decisions, *Data Equals* directs attention upstream to the data formats that make those decisions possible. This builds directly on his broader methodological claim that models are shaped by prior choices about variables, feature construction, and what is included or omitted from statistical representation. The key concept here is discretization or the tendency of learner data to be structured around the assumption that each student is a separate, bounded, and individually legible unit. This is

one of the book's most elegant insights. Personalized learning systems are individualizing not merely because of ideology or interface design, but because their underlying data ontologies already privilege the discrete learner over the relational one. In this context, Koopman sketches the possibility of more "synechistic" forms of data, being forms capable of registering continuity, collaboration, shared activity, and interaction between learners, rather than merely assigning properties to isolated individuals. The distinction between discrete and relational data is one of the book's most memorable contributions, because it shows that the politics of educational technology may be embedded in the database schema long before any recommendation engine begins to operate.

This argument deepens further when Koopman turns from separation to reproduction. If personalized learning separates learners through discretization, it can also reproduce inequality through its reliance on prior-performance data. Systems built on grades, test scores, platform logs, and other records of past performance do not simply discover who students are but also risk reinscribing who unequal institutions have already made them appear to be. Here Koopman links relational and distributive inequality into a mutually reinforcing cycle: 1) learners are separated; 2) that separation shapes access and opportunity; 3) unequal outcomes are fed back into the system as evidence; and 4) the resulting infrastructure differentiates learners still further. In this respect, his argument resonates strongly with the book's broader claim that data systems can reproduce hierarchy precisely by presenting contingent and historically loaded categories as if they were neutral descriptions.

One of the book's great strengths is that it does not stop at diagnosis. Koopman repeatedly resists the temptation, which is common enough in some critical writing on technology, to imply that the answer is simply less data, less modelling, or less technology. Instead, he argues for more equal forms of data and for a reorientation of educational technology away from individualized optimization and toward collaborative forms of learning. This is one of the most constructive and refreshing parts of the book. Drawing on the distinction developed by Wise, Knight, and Buckingham Shum, Koopman usefully separates the "analytics of collaborative learning" from "collaborative-learning analytics." The latter does not merely study collaborative learning, but it uses analytics to support and improve it. This distinction matters. It suggests that the alternative to personalized learning is not a mere rejection of analytics, but a different design paradigm altogether, one that shifts the unit of analysis from the isolated learner to the group, the interaction, the network, or the collective task.

This emphasis on collaboration is not merely pedagogical it is also deeply normative. It follows directly from Koopman's earlier account of democratic equality as a relational ideal. What matters is not simply whether individuals receive equal treatment in the abstract, but whether institutions enable them to stand in relations of equality with one another (Koopman 2024, 79–80). This is what makes the discussion of collaborative learning more than an educational preference. It becomes a paradigmatic site in which democratic equality is enacted. This is a bold claim, but I think it is also a very convincing one. It helps explain why the book's critique of personalized learning is ultimately about much more than pedagogy or software design. It is about whether educational infrastructures cultivate democratic capacities or quietly erode them. That question feels especially urgent in deeply unequal educational contexts, including many in the Global South, where data-driven differentiation can too easily reproduce older institutional hierarchies under the polished language of innovation.

Koopman's conclusion brings these threads together with clarity. Particularly effective is his insistence that implementation matters, a phrase that neatly captures the book's overall movement from diagnosis to method to reconstruction. His tripartite model of democratic

reconstruction, legal regulation, technological redesign, and human education, is especially helpful because it avoids the familiar tendency to reduce technological problems either to policy or to design alone. The conclusion is also where some of the book's most striking conceptual language appears, especially in its treatment of database structures, interface requirements, and algorithm selection as forms of "technological disposition." This is a subtle but powerful claim. Data systems do not merely describe users. They position, orient, and shape them. They dispose subjects as they enter the system.

The discussion of regulation is likewise refreshingly concrete. Rather than repeating the now predictable claim that digital technologies exist in some lawless frontier beyond the reach of the administrative state, Koopman argues that the largest technology firms are often not simply under-regulated but selectively protected. They benefit from immunities, expansive trade secrecy regimes, weak competition enforcement, and inadequate product-safety oversight. This is, in many ways, a recognizably legal insight. Forms, classifications, and regulatory omissions are never merely technical or administrative; they are part of how institutional power is structured and exercised. The book is equally strong when it turns to technical redesign. Here, redesign is not treated as a matter of adding ethics after the fact, but as a reconstruction of values already sedimented in data formats, interfaces, and classificatory structures. Koopman's example of requiring users to select categories such as race or gender in a database is especially effective, because it makes visible what *Data Equals* insists upon throughout is that the politics of data are often located upstream of the model in the variable, the field, the schema, and the format.

If I have a reservation, it is that the book's constructive program is more persuasive as a normative and diagnostic intervention than as a fully worked design agenda. The contrast between "technologies of separation" and "technologies of collaboration" is analytically powerful and often illuminating, but at times it feels sharper than the messy realities of educational practice may warrant. Many educational settings will require hybrid arrangements that combine individualized support with genuinely collaborative structures, and the book is less interested in mapping those compromises than in clarifying the stakes of the opposition. Likewise, the call for more equal forms of data is compelling, but the practical criteria by which such infrastructures might be built, governed, and evaluated remain somewhat programmatic. What, concretely, would distinguish a genuinely egalitarian collaborative-learning system from a platform that simply rebrands familiar forms of surveillance under a more participatory vocabulary? Koopman gives readers the conceptual tools to ask that question, but less by way of operational criteria for answering it.

I would treat these less as flaws than as the limits of a book whose primary achievement is conceptual reorientation rather than technical blueprinting. Data Equals is at its strongest when it shows that the politics of educational AI are not exhausted by downstream questions of bias, accuracy, or performance. They are already embedded upstream in the formats, variables, interfaces, and institutional assumptions that shape what data can count as, what learners can become, and whether they can encounter one another as equals. In that sense, the book offers something more demanding than a fairness critique, and more useful than a generic suspicion of technology. It offers a democratic theory of educational data infrastructures that asks us to take seriously the possibility that equality must be built not only into laws and institutions, but into the data formats through which those institutions increasingly know and govern us. That is a demanding standard. It is also precisely what makes *Data Equals* such a valuable book. If, as Koopman insists, we now increasingly live through our data, then the future of democratic education may well depend on learning how to become, in his resonant phrase, **data equals**.

Author Contributions: Reading and reviewing of book; writing, review, and editing of original draft preparation, M.B.

Acknowledgments: The author used OpenAI's ChatGPT (GPT-5 series) solely for assistance with language editing, stylistic refinement, structural reorganization, and formatting consistency during the preparation of this book review. All substantive interpretation, critical analysis, selection of cited material, and final editorial decisions were undertaken by the author. The author independently read the book under review, reviewed and revised all AI-assisted output, and accepts full responsibility for the accuracy, integrity, and originality of the manuscript.

References

Khan, Arif Ali, Sher Badshah, Peng Liang, Muhammad Waseem, Bilal Khan, Aakash Ahmad, Mahdi Fahmideh, Mahmood Niazi, and Muhammad Azeem Akbar. "Ethics of AI: A systematic literature review of principles and challenges." In Proceedings of the 26th international conference on evaluation and assessment in software engineering, pp. 383-392. 2022. https://doi.org/10.1145/3530019.3531329.

Koopman, Colin. Data Equals: Democratic Equality in Data Design. Chicago: University of Chicago Press, 2024.

Whittlestone, Jess, Rune Nyrup, Anna Alexandrova, Kanta Dihal, and Stephen Cave. "Ethical and societal implications of algorithms, data, and artificial intelligence: a roadmap for research." London: Nuffield Foundation (2019): 1-59. https://www.nuffieldfoundation.org/wp-content/uploads/2019/02/Ethical-and-Societal-Implications-of-Data-and-AI-report-Nuffield-Foundat.pdf.