



## King's Research Portal

### *Document Version*

Publisher's PDF, also known as Version of record

[Link to publication record in King's Research Portal](#)

### *Citation for published version (APA):*

Meunier, A., Gray, J., & Ricci, D. (2021). A New AI Lexicon: Algorithm Trouble: Troublesome Encounters with Algorithms that go Beyond Computational Processes. In *A New AI Lexicon* AI Now Institute.  
<https://ainowinstitute.org/publication/a-new-ai-lexicon-algorithm-trouble>

### **Citing this paper**

Please note that where the full-text provided on King's Research Portal is the Author Accepted Manuscript or Post-Print version this may differ from the final Published version. If citing, it is advised that you check and use the publisher's definitive version for pagination, volume/issue, and date of publication details. And where the final published version is provided on the Research Portal, if citing you are again advised to check the publisher's website for any subsequent corrections.

### **General rights**

Copyright and moral rights for the publications made accessible in the Research Portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognize and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the Research Portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the Research Portal

### **Take down policy**

If you believe that this document breaches copyright please contact [librarypure@kcl.ac.uk](mailto:librarypure@kcl.ac.uk) providing details, and we will remove access to the work immediately and investigate your claim.

# A New AI Lexicon: Algorithm Trouble

Troublesome Encounters with Algorithms that go Beyond Computational Processes

[AI Now Institute](#)

Dec 16, 2021

.

10 min read

.

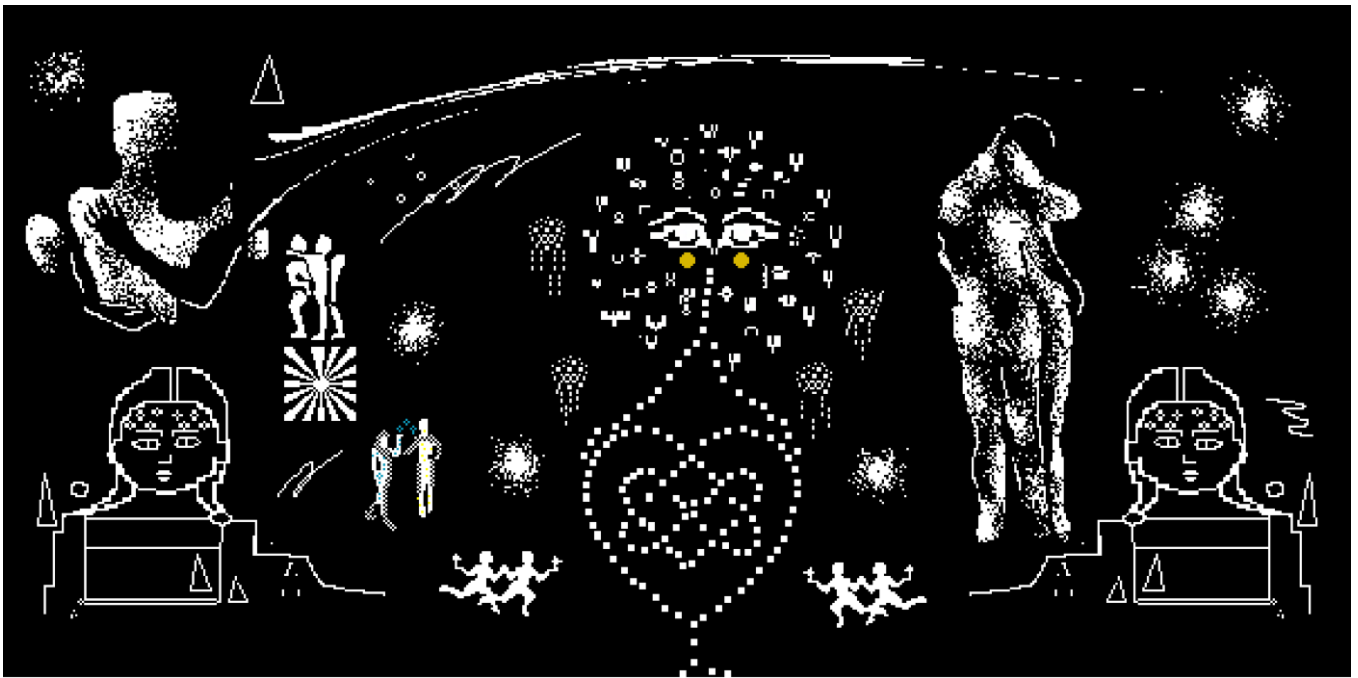


Illustration by Somnath Bhatt

A guest post by Axel Meunier, Jonathan Gray and Donato Ricci. Axel is a PhD candidate in design at Goldsmiths, University of London. Jonathan is Senior Lecturer in Critical Infrastructure Studies at Department of Digital Humanities, King's College London. Donato is designer and researcher at Sciences Po médialab, Paris.

*This essay is part of our ongoing "[AI Lexicon](#)" project, a call for contributions to generate alternate narratives, positionalities, and*

*understandings to the better known and widely circulated ways of talking about AI.*

For decades, social researchers have argued that there is much to be learned when things go wrong.<sup>1</sup> In this essay, we explore what can be learned about algorithms when things do not go as anticipated, and propose the concept of **algorithm trouble** to capture how everyday encounters with artificial intelligence might manifest, at interfaces with users, as unexpected, failing, or wrong events. The word *trouble* designates a problem, but also a state of confusion and distress. We see algorithm troubles as failures, computer errors, “bugs,” but also as unsettling events that may elicit, or even provoke, other perspectives on what it means to live with algorithms — including through different ways in which these troubles are experienced, as sources of suffering, injustice, humour, or aesthetic experimentation (Meunier et al., 2019). In mapping how problems are produced, the expression *algorithm trouble* calls attention to what is involved in algorithms beyond computational processes. It carries an affective charge that calls upon the necessity to *care* about relations with technology, and not only to fix them (Bellacasa, 2017).

Looking for proxies to explore the troubles that people experience in their everyday encounters with algorithms, we conducted a search of posts containing [#Alfail](#) or [#algorithmfail](#) on Twitter, Instagram or Facebook.<sup>2</sup> What we found included examples of misclassification (“Facebook thinks my channel is a restaurant”); misplaced recommendations (“The NRA is recommended to a depression and suicide awareness person?”); unrelated suggestions (“If we’re out of plant-based cheese, would you like some plant-based flushable wipes?”); content moderation issues (Fig. 1 “Rizzo, btw, is a horse”); humorous mis-targeting (“the internet does not get me at all”, “clearly they know me so well”, “thanks but no thanks”); problematic automated translations (“[She is beautiful. He is clever. He understands math. She is kind ...](#)”).<sup>3</sup>

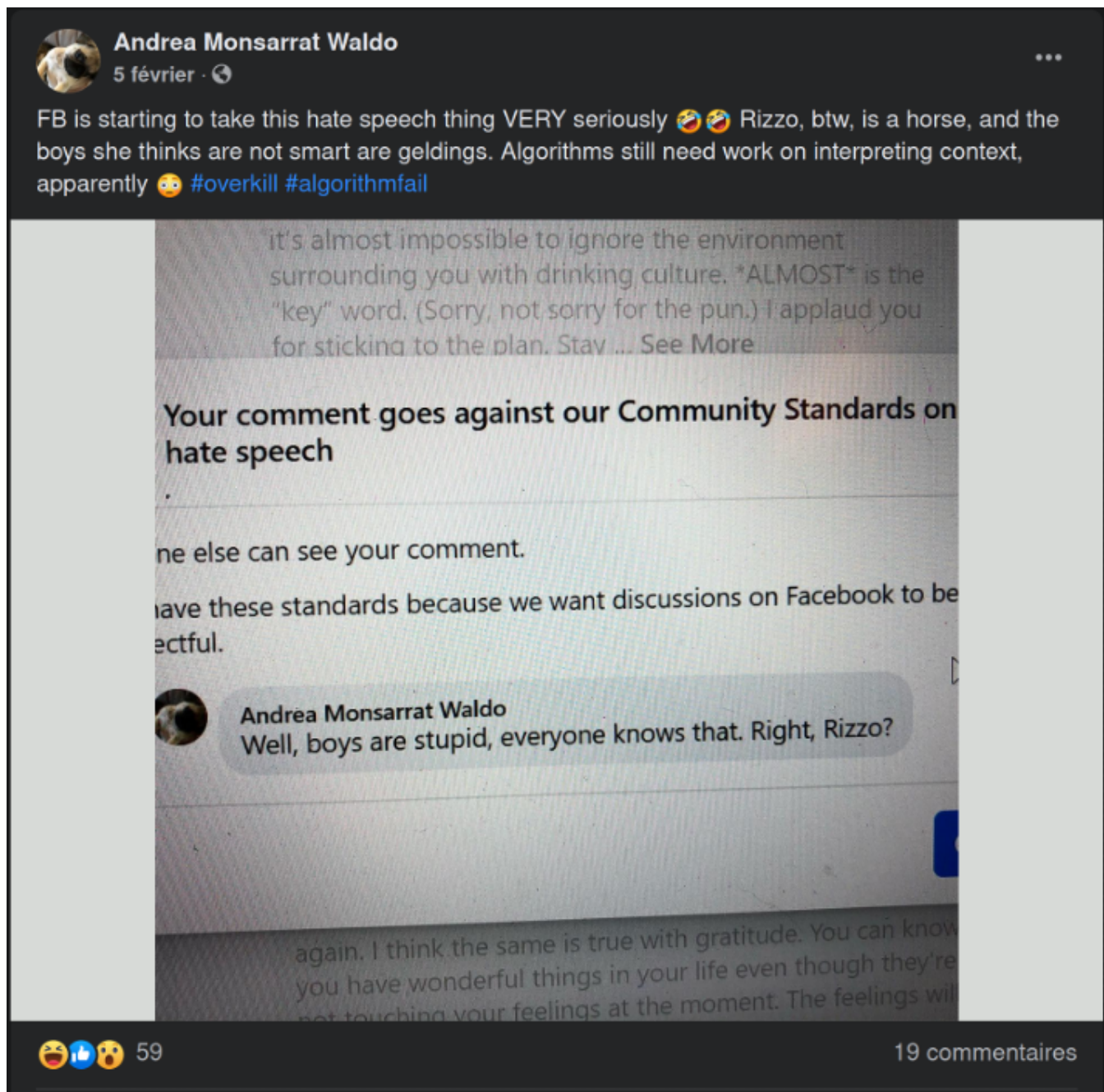


Fig. 1 Content moderation trouble on Facebook.

As Safiya Noble highlights in her *Algorithms of Oppression*, screenshots of problematic search results are as much a research technique as a part of digital culture, surfacing how algorithm trouble is encountered and experienced in everyday settings (2018). She gives the example of teenager Kabir Ali's viral post with a video showing his search for "three black teenagers" showing Google Image results dominated by mug shots, in contrast to "three white teenagers" who are represented as "represented as wholesome and all-American" (Noble, 2018). Similarly in

2016, concerns around image search results emerged through the screenshots of a student querying for “unprofessional hairstyles for work” which mainly showed images of Black women with natural hair.<sup>4</sup> Five years later, the same google search produces the same outcome, albeit referring to the *trouble* itself and its subsequent commentaries on racism in relation to Afro hair styles, which have now *extended* the algorithmic system and thickened the accounts and stories that can be told in relation to the questions: *Which are the sensitive and affective effects of being miscalculated? How do we live and negotiate with them? What is the outcome of such interplay?*

Algorithm troubles materialise through a range of different digital objects, devices, and media. Even though they can be accounted for through bug reports, technical charts, or computational tracking, they are best — or so we believe — *expressed* in the vernacular formats of internet culture: social media threads, screenshots, GIFs, video recordings, composite images, and memes. Their sensible expression bears and embeds the residual traces of an equally wide range of emotions: humor, irony, scorn, sadness, anger, or outrage. Furthermore, digital objects such as hashtags may serve as sites where algorithm troubles are not only expressed by internet users, but also gathered and collected so as to surface, thread together, and “issuefy” algorithms, complementing initiatives such as the AI Incident Database<sup>5</sup> or the reports and monitoring efforts of dedicated organisations such as AlgorithmWatch and the Algorithmic Justice League.<sup>6</sup> As Sara Ahmed comments in her *Living a Feminist Life*, collective processes of gathering together troubling incidents into lists or catalogues can be a way to attend to the specificity of incidents as well as “hearing continuities and resonances” across them, so that *“this or that incident is not isolated but part of a series of events: a series as a structure.”* (Ahmed, 2017, p. 30)

The twitter account [@internetofshit](#) came up during our previous research on algorithmic glitches, not only as a popular account to mock less-than-

smart IoT but also as a serious attempt to document systematically the convergence between things and computers in everyday life, during the fleeting moments when the seams between the digital technologies and daily circumstances become visible. It also reminds us that algorithm troubles are quickly moving beyond the realm of computer interfaces. For example in Figure 2 below, a user with a face recognition-enabled home security system user must go around their snow-covered garden to look for, and stomp with their feet, any random pattern that might look like the face of a burglar, thus triggering text alerts from their surveillance camera that thinks that the snow is a real human face. The system's propensity to pareidolia — i.e., the tendency to perceive a specific, often meaningful image in a random or ambiguous visual pattern — has created an entirely new realm of algorithmic trouble for users, and coping here entails learning new daily gestures to deal with edge cases of face recognition technologies.



Fig. 2 Face recognition trouble on Facebook, retweeted.

Algorithmic technologies build upon statistical regularity to anticipate and predict users' expectations, needs, and satisfaction, and indeed *do* discriminate, in the sense that "*they afford greater degrees of recognition and value to some features of a scene than they do to others*" (Amoore, 2019, p. 8). Now, when the algorithmic trouble unfolds, this seamless assemblage of statistics, interaction design, and "relevant" features, gives way to an encounter, where the entities and relations<sup>7</sup> that *matter* in the situation are uncertain. Calculations that stop being congenial to



expectations, might just require a fix, but might be the start of a collective inquiry into a vast socio-technical infrastructure, in search of a readjustment between the two entangled sides of any algorithmic process: the calculated world and the sensitive one. *Is it a new edge case? Should the code be changed? Has the technology been misused? Should the service be redesigned or deactivated? Are we just over-projecting capacities over these computational services and processes?* For example, in Figure 3 below, Twitter's image cropping algorithm — which automatically chooses how an image embedded in a tweet is displayed — generated concerns around racial bias after the following case was posted.



Fig. 3 Twitter image crop algorithms trouble.<sup>8</sup>

The trouble led to a first round of collective inquiry, through user interface tests, to explore whether image crops focused on white faces rather than black faces, that included not only photographs of people but also illustrations of cartoon characters where similar troubling patterns were observed. A few months later, after implementing a new basic version of cropping without machine learning, Twitter launched an open competition to search for biases and “debug” their algorithm.<sup>9</sup> However, what the

competition did was not solve the trouble with the cropping algorithm, quite the opposite: it articulated the trouble with new sets of "representational harms" affecting "*groups beyond those addressed in our previous work, such as veterans, religious groups, people with disabilities, the elderly, and individuals who communicate in non-Western languages.*"<sup>10</sup> Through collective inquiry, the challenge put in question the saliency model underlying the cropping algorithm, and, thus doing, probably *diminished* the chance of "solving" the problem, at least in the short term. But it also did something else: it made the cropping algorithm appear as an *ontological trouble* which was progressively given substance and turned into a matter of concern for different unanticipated communities. Communities outside the global north have been engaged in this process as well, which also supports the fact that the uncanny *sensation* associated with our being calculated can be the trigger for the comprehension of a whole computational experience as much as messy and tentative — emotive and biographical — plots rather than as purified and objective — analytical and abstracted — bug. Here we follow Marres (2016) in her argument that underlines a "*dynamic political ontology in which the process of the specification of issues and the organization of actors into issue assemblages go hand in hand*" (Marres, 2016, p. 55), a conception she traces back to early 20th-century American pragmatist philosopher Dewey.

Slowing down and staying with algorithm trouble might suggest different collective responses and political approaches, beyond problem-solving and policy fixes. The examples of troubles we have used to illustrate this piece have been collected from global platforms, taking English language keywords and hashtags as starting points. This only gives a very partial picture, and more inquiries are needed to attend to how algorithms become troubling in beyond English language keywords, beyond the most studied platforms, and outside the Global North.

In complement to approaches to algorithmic auditing and accountability

which start with, for example, reverse-engineering and the use of computational techniques to detect biases, algorithm troubles are the temporary accounts of a situation that has not yet bifurcated into a technological component on one side, and a social component on the other; into code and context; into an individual use case to be better taken into account, and a general social critique of algorithms. This is precisely why the algorithm trouble, as fragile and fleeting as it may seem, is worth attending to: it presents itself as a situated, mundane and ordinary experience of living with algorithms, stripped of the extraordinary promises and future perils through which algorithms and artificial intelligences are usually envisaged; yet it destabilizes and engages, intensifies and deflects the inertia of beliefs into an active inquiry.

Before closing this essay, it might be worth linking the algorithmic trouble to the critique of fairness solutions in Overdorf et al. (2018) where they argue that fairness solutions are unable to address *externalities* of Optimization Systems, in particular the dynamic relationship these systems possess with the environment in which they are deployed. They liken possible response strategies to those coming from *outside* the system, for example when Uber drivers collectively game the matching algorithm to induce surge pricing. It could be worth also pointing out how externalities are dynamically redistributed through the articulation of issues we have described, a process which requires adhoc sociotechnical inquiries into the social world rather than the modeling of impacts.

Donna Haraway, who advocates the virtues of "staying with the trouble" of living together with non-humans that make the world as much as humans do (Haraway, 2016), also reminds us of yet another meaning of trouble in French, that describes the material quality of liquids: cloudy and blurred, as opposed to clear and transparent. The algorithm troubles redirect the attention from the ideal of transparency to solve the political problem of algorithmic systems, to the muddy and messy reality of composing a collective life with (and without) them.

## Footnotes

[1] For example, see Latour, 2007; Star, 1999; Graham and Thrift, 2007.

[2] The use of web-native objects such as hashtags or hyperlinks as a method for social inquiry, as well as precautions to take when doing so, is based on scholarship in media studies (Rogers, 2013), digital sociology (Marres, 2017) or controversy mapping (Venturini and Munk, 2021).

[3] See: <https://www.instagram.com/explore/tags/algorithmfail/> and <https://twitter.com/search?q=%23algorithmfail>

[4] Alexander, Leigh. *Do Google's 'unprofessional hair' results show it is racist?* The Guardian. See:

<https://www.theguardian.com/technology/2016/apr/08/does-google-unprofessional-hair-results-prove-algorithms-racist->

[5] See : <https://incidentdatabase.ai/>

[6] See : <https://algorithmwatch.org/en/> and <https://www.ajl.org/>

[7] The relations between 'values, assumptions, and propositions about the world' (Amoore, 2019, p. 6)

[8] Romero, Aja. *Open for a Surprise : The endearing results of Twitter's new image crop.* Vox. See : <https://www.vox.com/22430344/what-is-twitter-crop-new-image-ratio-memes>

[9] See: <https://twitter.com/TwitterEng/status/1424819778397511680?s=20>

[10] Chowdhury, Rumman. *Sharing learnings about our image cropping algorithm.* Twitter Engineering Blog. See: [https://blog.twitter.com/engineering/en\\_us/topics/insights/2021/sharing-learnings-about-our-image-cropping-algorithm](https://blog.twitter.com/engineering/en_us/topics/insights/2021/sharing-learnings-about-our-image-cropping-algorithm)

## Bibliography

Ahmed, S. (2017) *Living a Feminist Life*. Durham: Duke University Press.

Amoore, L. (2020) *Cloud Ethics: Algorithms and the Attributes of Ourselves and Others*. Durham: Duke University Press.

Bellacasa, M.P. de la (2017) *Matters of Care: Speculative Ethics in More Than Human Worlds*. Minneapolis: University of Minnesota Press.

Graham, S. and Thrift, N. (2007) 'Out of Order', *Theory, Culture and Society*, 24(3), pp. 1–25. doi:10.1177/0263276407075954.

Haraway, D.J. (2016) 'Staying with the Trouble: Making Kin in the Chthulucene'. doi:10.1215/9780822373780.

Latour, B. (2007) *Reassembling the Social: An Introduction to Actor-Network-Theory*. Oxford, New York: Oxford University Press (Clarendon Lectures in Management Studies).

Marres, N. (2016) *Material Participation: Technology, the Environment and Everyday Publics*. Springer.

Marres, N. (2017) *Digital Sociology: The Reinvention of Social Research*. Malden, MA: Polity Press.

Meunier, A. et al. (2019) 'Les glitches, ces moments où les algorithmes tremblent', *Techniques & Culture. Revue semestrielle d'anthropologie des techniques* [Preprint]. Available at: <https://journals.openedition.org/tc/12594> (Accessed: 16 September 2021).

Noble, S.U. (2018) *Algorithms of Oppression: How Search Engines Reinforce Racism*. New York: New York University Press.

Overdorf, R. et al. (2018) 'Questioning the assumptions behind fairness solutions', arXiv:1811.11293 [cs] [Preprint]. Available at:

<http://arxiv.org/abs/1811.11293> (Accessed: 10 December 2021).

Rogers, R. (2013) *Digital Methods*. Cambridge, MA, USA: MIT Press.

Star, S.L. (1999) 'The Ethnography of Infrastructure', *American Behavioral Scientist*, 43(3), pp. 377–391. doi:10.1177/00027649921955326.

Venturini, T. and Munk, A.K. (2021) *Controversy Mapping: A Field Guide*. Cambridge, UK ; Medford, MA, USA: Polity Press.