A Mulching Proposal

Analysing and Improving an Algorithmic System for Turning the Elderly into High-Nutrient Slurry

Os Keyes
Department of Human Centered Design &
Engineering
University of Washington
Seattle, WA, USA
okeyes@uw.edu

Jevan Hutson
School of Law
University of Washington
Seattle, WA, USA
jevanh@uw.edu

Meredith Durbin
Department of Astronomy
University of Washington
Seattle, WA, USA
mdurbin@uw.edu

ABSTRACT

The ethical implications of algorithmic systems have been much discussed in both HCI and the broader community of those interested in technology design, development and policy. In this paper, we explore the application of one prominent ethical framework—Fairness, Accountability, and Transparency—to a proposed algorithm that resolves various societal issues around food security and population ageing. Using various standardised forms of algorithmic audit and evaluation, we drastically increase the algorithm's adherence to the FAT framework, resulting in a more ethical and beneficent system. We discuss how this might serve as a guide to other researchers or practitioners looking to ensure better ethical outcomes from algorithmic systems in their line of work.

CCS CONCEPTS

• **Human-centered computing** → **Empirical studies in HCI**; **Social engineering (social sciences)**; • **Computing methodologies** → **Object recognition**; **Machine learning algorithms**; •

Social and professional topics → **Age**; • **Applied computing** → *Consumer health*; **Metabolomics / metabonomics**; **Agriculture**;

KEYWORDS
algorithmic critique; algorithmic analysis; computer vision; dystopia; fairness; accountability; transparency; ethics

ALGORITHMIC CRITIQUE AND HCI
As algorithmic systems have become more widely deployed in critical social domains, a range of concerns have arisen about how such systems might create unjust and harmful outcomes. These concerns have—in areas as varied as child welfare policy and cancer treatment—often been validated [11, 23, 29, 32].

In response, researchers have developed a range of standards by which to evaluate and critique algorithmic systems. These include not only academic proposals, such as the "Principles of Accountable Algorithms", [10, 24] but also proposals from major technology companies [8, 25, 31]. One shared set of values these standards offer can be summarised by the principles of the "Fairness, Accountability and Transparency" (FAT) [9, 12, 18, 21] framework. This says that algorithmic systems should, to be ethical, be:

(1) *Fair*: lacking biases which create unfair and discriminatory outcomes;
(2) *Accountable*: answerable to the people subject to them;
(3) *Transparent*: open about how, and why, particular decisions were made.

By assuring these conditions are met, we can rest easy, threatened no more by the possibility of an algorithm producing harmful outcomes.

HCI's response to this work has been to dive into unpacking these principles' meaning, and develop mechanisms with which they might be attained. We have developed new transparency approaches [26], explored user perceptions of fairness [17], and directly "audited" algorithms to identify whether they are biased [5, 6, 28].

But these studies are often external, undertaken by researchers who are not directly engaging with the algorithm's developers. As a consequence, researchers have a limited ability to directly correct any problems they find and test their proposed solutions. In addition, studies often only tackle one
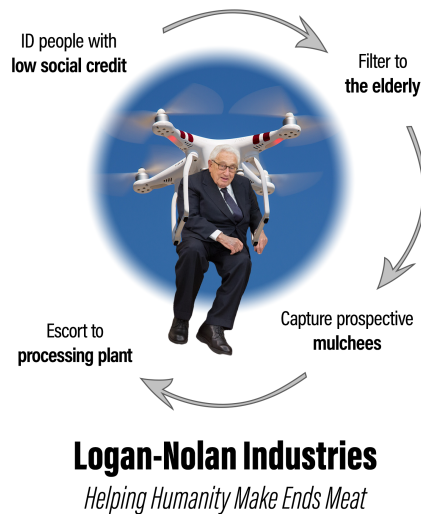
Figure 1: A publicity image for the project, produced by Logan-Nolan Industries

part of the FAT framework, focusing exclusively on (for example) transparency or fairness, rather than taking a systemic view of these principles.

In this paper, we document an algorithmic analysis which was undertaken in *collaboration* with the system's developers, seeking to ensure that design followed the principles laid out in the FAT framework. We believe that our work makes for an interesting case study in HCI engagement with industry. HCI researchers concerned about algorithmic ethics might use our experience as a basis for more efficacious and holistic engagement with the systems they are studying.

FIELD SITE

Logan-Nolan Industries (LNI) is a large multinational with a range of interests, from software development to mining. Recognising a gap in the market created by multiple ongoing crises—namely, the aging population in Western society, and the likely reduction in arable farmland due to climate change—they developed a program in which elderly people are rendered down into a fine nutrient slurry, directly addressing both issues. Elderly volunteers (or "mulchees") are provided with generous payments for their families before being rendered down and recombined with other chemicals into a range of substitutes for common foodstuffs, including hash browns (Grandmash™), bananas (Nanas™) and butter (Fauxghee™).

Promisingly remunerative early trials led LNI to seek to expand—but there are only so many possible volunteers. It seems that, despite the *clear* benefit the program offers humanity, many are reticent to be rendered. In an effort to move past this, LNI has developed a new, automated approach. An algorithm, provided with vast amounts of social media and communications data, identifies subjects with low levels of social connectivity, partially using a HCI-developed algorithm for approximating "social credit" [30]. Photographs of subjects who fall below a certain connectivity threshold are then run through a computer vision system which identifies those who appear to be above the age of 60. Once a person is classified as both old and unloved, their information is dispatched to a network of patrolling unmanned aerial vehicles (UAVs) that—upon spotting a person who is present in the database—obtain them and bring them to the nearest LNI processing centre.

LNI consulted informal focus groups on this proposal. The company was surprised to find that possible participants responded quite negatively to the idea. LNI's expert geriatric gustatologists thus reached out to us, seeking our expertise in order to resolve the anxiety of both their consumers and those to be consumed. We were afforded full access to the development team and process, including contact with senior managers, in order to ensure our feedback could be directly implemented. The result is an interesting (and certainly unique!) case study in algorithmic systems design.

FINDINGS

Fairness

Fairness is a complex problem—there are, after all, many definitions of "fair" [7, 13, 20]—and we chose, as other algorithmic researchers have [3, 19], to look specifically for *demographic* fairness. Did the system (either in determining social connectivity or age) evenly distribute false positives and negatives across racial and gender groups?

We assembled a dataset of 900 images across these demographic axes, sourced from LNI employees and family members who consented (through their employment contracts) to allow us to assess their social credit score and use their photographs. Having quickly skimmed Keyes's *The Misgendering Machines* [16], we saw a need to include transgender (trans) people in our dataset, and expanded our model of gender in order to do so. The resulting data was tested against the LNI algorithm; our results can be seen in table 1.

Table 1: Percentage of individuals tagged as worthy of mulching, by demographic.

	Mulching Probability				
Race	*Cis Man*	*Cis Woman*	*Trans Man*	*Trans Woman*	*Non-Binary Person*
White	44.6%	33.3%	2.2%	3.2%	1.1%
Asian-American	22.2%	16.3%	2.8%	1.2%	1.8%
African-American	26.9%	11.2%	2.3%	1.9%	3.4%
Latino	16.9%	18.7%	3.3%	1.2%	1.7%
Native American	14.4%	12.4%	1.0%	0.8%	1.5%
Hawaiian & Pacific Islander	11.6%	7.8%	2.4%	1.1%	0.7%

As shown, the algorithm disproportionately tagged white cisgender men as worthy of mulching, biasing against other populations. While we cannot adequately explain what wider reason the algorithm might have for determining that white cisgender men are, on average, lacking in societal worth, this is clearly unacceptable. For an algorithm to be fair, it must lack gender or racial bias.

We provided our results and concerns to LNI's engineers, who were eager (unsurprisingly, given the demographics of the average engineering department) to address this issue. They responded by collecting the photographs and social credit data of 3,000 more potential mulchees, particularly women, trans people and/or people of colour, in line with previous (and, according to their authors, disproportionately consequential) approaches to improving fairness [27]. These images and data traces were integrated into the model, which (as seen in Table 2) now produces far fairer results.

Table 2: Post-audit mulching probabilities.

Race	Cis Man	Cis Woman	Trans Man	Trans Woman	Non-Binary Person
White	44.6%	43.3%	44.2%	46.3%	41.2%
Asian-American	52.2%	51.3%	55.8%	49.6%	52.3%
African-American	46.9%	51.1%	53.2%	49.1%	53.3%
Latino	56.9%	48.2%	47.3%	51.1%	47.4%
Native American	54.4%	54.2%	51.5%	48.8%	51.2%
Hawaiian & Pacific Islander	51.6%	48.6%	44.9%	51.1%	47.0%

User Feedback

"I don't know if I'm comfortable eating Nonna"
Judith, grand-daughter of a potential mulchee.

"Until that little robot showed up I'd never even heard of this program. Say, how did you get in my house, anyway?"
Robert, a potential mulchee classified as "not to be mulched".

"Do the papers know about this kind of thing? There ought to be some investigation!"
Joan, a potential mulchee classified as "to be mulched".

"Ow!"
Colin, being mulched.

Accountability

Accountability is a vital component of the FAT framework, and refers to the ability of people to address any failure or inaccuracy produced by an algorithmic system [22]. This is particularly important here given the consequences of an incorrect classification.

In the case of this algorithm, a failure necessitating accountability might happen at two points. The computer vision algorithm could fail, incorrectly classifying someone as elderly, or the analysis of social connections might be inaccurate due to a person's limited presence on social media sites.

Initial versions of LNI's algorithm lacked mechanisms for accountability—they were not responsive to possible user concerns, not even leaving as much as a voicemail, let alone providing any mechanism of redress if someone felt that they (or more likely, their friend or relative) had been rendered into slurry incorrectly.

To address accountability concerns, we undertook a round of *formalised* user testing, soliciting feedback from mulchees and their relatives and friends at various stages in the mulching process. Some examples of the feedback can be seen in the sidebar. Based on the feedback, we proposed two mechanisms for accountability—one appearing prior to mulching, and the other after, and both interlinking with concerns around *Transparency* (see below).

The pre-mulching accountability mechanism involves the drone itself. After approaching a pending mulchee, the drone informs them that they have been selected for participation in this program. They are then afforded a ten-second window in which to state whether they feel their selection was correct or not. In the event that they feel they were selected in error, they are put through to a human operator at LNI customer services. The operator ("or death doula") discusses the reasons behind the customer's classification, and presents them with an opportunity to discuss possible factors in age or societal utility the algorithm may have overlooked. They then either confirm or reverse the algorithm's
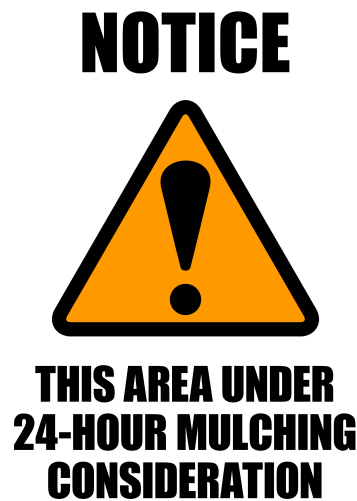
decision. Their decisions are reported back to the algorithmic development team for consideration when adding new variables to the model or altering variable weight.

Post-mulching, the company reaches out to the friends and family of the mulchee (if such individuals exist) to inform them of the decision reached and provide the serial numbers of any food products containing their relative. Our user studies showed that people express some qualms about eating their grandparents. In the event that next-of-kin feel the decision was made wrongly, they are offered a 30-day window in which to appeal. While LNI cannot reconstitute their loved one, the company has agreed to provide an elderly person of equal or greater wholesomeness and social utility, at discounted cost.

While proposing these ideas we were highly cognizant of potential cultural bias: the 10-second and 30-day window are clearly highly variable periods of time in the context of language proficiency. For those reasons, the company has invested vast sums in accumulating additional data on not only possible mulchees but also their social contexts, ensuring that any letters and/or employees can engage in the customer's native language.

Transparency

Deeply intertwined with *Accountability* is the principle of *Transparency*: are users aware of how the algorithm reaches decisions? This is vital in allowing users and regulators to evaluate any algorithmic system.

We address transparency in several ways. During the pre-mulching accountability mechanism, the drone provides not only the decision but also a comprehensive list of variables that were considered—including but not limited to phone and SMS metadata, number of facebook friends, number of birthday and christmas cards received from relatives—along with the scores the mulchee received *for* each variable. These are also (albeit by letter) provided to those who narrowly fail to meet the threshold for mulching.

LNI executives were particularly enthused at participants' comments (see the user feedback in the sidebar) that the "not mulched" notice made them far more aware of the program, raising public attention and serving as free advertising. Building upon this, and in the name of transparency, we have encouraged LNI to provide additional notice and achieve more rigorous consent from potential participants by deploying ubiquitous signage in public and private settings [4]. This can be seen as Figure 2.

In undertaking this work we were cognizant of the ethical constraints that commercial contexts often generate. In particular, the proprietary nature of algorithmic systems is often a roadblock for algorithmic transparency [1, 33]. LNI felt it was important overcome this obstacle in order to promote broader understanding of its system (as well as gain competitive advantage over other providers). Its

Figure 2: A warning sign of drones' presence, serving as a mechanism for consent.

solution has been to actively waive any right to legally pursue third-party researchers who audit or otherwise examine LNI's software.

We also suggested—in line with the feedback we gathered through user testing—that there was some advantage in being more proactive in public audit and transparency efforts. Our proposal was to build tools that would allow members of the public to easily interact with the model. The result is an open website (www.mulchme.com) where members of the public can upload their own photographs and data, play around with the model's features, and see what results. LNI saw this as not only a boost to transparency, but also an opportunity to collect further data for testing and refining their model.

DISCUSSION

We realise that some researchers (be they potential mulchees or not) might react to this paper with some disquiet. In particular, one might argue that the paper undermines itself: that rather than showing how useful Fairness, Accountability and Transparency can be in reducing the negative possibilities of an algorithm, it shows its paucity.

In particular, such researchers could point to the fact that this work assumes algorithmic failures to be a question of implementation. Issues with the morality of this algorithm are treated as resolvable through changes to input data and deployment. Such a frame of ethics ignores whether murdering the elderly might be morally obscene in principle. Sometimes the problem is not how the sausage gets made, but that they're making people into sausage.

One could also argue that, as might be expected from this tacit acceptance of the datafication of society, our ethical frame defers to the concerns and priorities of Logan-Nolan Industries and their (neoliberal) motivations. From that view, our direct collaboration with LNI is not to be lauded, but instead condemned as legitimising their actions with the stamp of scientific approval, and accepting these inhumane infrastructures as inevitable. We are collaborators not in the sense of team-mates, but in the sense of Vichy France.

Furthermore, even were we not complicit in this legitimisation, our approach fundamentally centers the treatment of ethics as a series of heuristic checkboxes that can be resolved technically, or through design methods, rather than as representative of wider societal issues. In doing so, it simultaneously over-fits (by ignoring the role that context plays in algorithmic propriety) and under-fits (by ignoring how these systems play into existing inequities) [2, 15].

Rather than run interference for Logan-Nolan Industries, they might say, we should instead begin challenging them directly, speaking out against *any* form of grandma-mashing machine, whether it is driven by an algorithm or a human. Perhaps we could consider developing and applying the same critical thinking skills when *evaluating* this algorithm that data scientists are being encouraged to apply when *developing* them:

"Would the implementation of this algorithm represent a reformist or non-reformist reform? Is the design of this algorithm affirmative or critical? Would empowering our project partner with this algorithm entrench or challenge oppression and inequality? Such efforts can help data scientists interrogate their notions of "good" to engage in non-reformist, critical, and anti-oppressive data science." [14]

But we are simply researchers—researchers seeking a fairer, more transparent and more accountable world, for the benefit of all. If this framing is insufficient, well: that would imply FAT *itself* were insufficient. And were *that* the case, it would say some pretty negative things about HCI's engagement with algorithmic evaluation, and carry some profoundly depressing implications about the actual meaningful impact we are likely to make [15].

CONCLUSION

In this paper we have presented a case study in algorithmic analysis, centred around a system that we hope will take big bites out of both food insecurity and population imbalances. Secure in the knowledge that nothing data ethicists would ask of us has been missed, we are excited to see what other researchers make of our techniques as we kick back, put our feet up, and enjoy a nice big plate of Fauxghee™.

ACKNOWLEDGEMENTS

REFERENCES

[1] Robert Brauneis and Ellen P Goodman. 2018. Algorithmic transparency for the smart city. *Yale JL & Tech.* 20 (2018), 103.

[2] Simone Browne. 2015. *Dark Matters.* Duke University Press.

[3] Joy Buolamwini and Timnit Gebru. 2018. Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification. *Conference on Fairness, Accountability and Transparency* 81 (2018), 1–15.

[4] Matt Burgess. 2018. Twitter. Retrieved December 24, 2018 from https://twitter.com/mattburgess1/status/1074731539030786048

[5] Le Chen, Ruijun Ma, Anikó Hannák, and Christo Wilson. 2018. Investigating the Impact of Gender on Rank in Resume Search Engines. In *the 2018 CHI Conference.* ACM Press, New York, New York, USA, 1–14.

[6] Le Chen, Alan Mislove, and Christo Wilson. 2015. Peeking Beneath the Hood of Uber. In *the 2015 ACM Conference.* ACM Press, New York, New York, USA, 495–508.

[7] Sam Corbett-Davies and Sharad Goel. 2018. The measure and mismeasure of fairness: A critical review of fair machine learning. *arXiv preprint arXiv:1808.00023* (2018).

[8] Microsoft Corporation. 2018. Our Approach to AI. Retrieved December 19, 2018 from https://www.microsoft.com/en-us/ai/our-approach-to-ai

[9] Nicholas Diakopoulos. 2016. Accountability in algorithmic decision making. *Commun. ACM* 59, 2 (Jan. 2016), 56–62.

[10] Nicholas Diakopoulos, Sorelle Friedler, Marcelo Arenas, Solon Barocas, Michael Hay, Bill Howe, H.V. Jagadish, Kris Unsworth, Arnaud Sahuguet, Suresh Venkatasubramanian, Christo Wilson, Cong Yu, and Bendert Zevenbergen. 2018. Principles for Accountable Algorithms and a Social Impact Statement for Algorithms. Retrieved October 28, 2018 from https://www.fatml.org/resources/principles-for-accountable-algorithms

[11] Virginia Eubanks. 2018. *Automating Inequality*. St. Martin's Press.

[12] ACM FAT. 2018. ACM Conference on Fairness, Accountability, and Transparency (ACM FAT*). Retrieved December 19, 2018 from https://www.fatconference.org/

[13] James Foulds and Shimei Pan. 2018. An Intersectional Definition of Fairness. *arXiv preprint arXiv:1807.08362* (2018).

[14] Ben Green. 2018. Data Science as Political Action - Grounding Data Science in a Politics of Justice. *CoRR* 1811 (2018), arXiv:1811.03435.

[15] Daniel Greene, Anna Lauren Hoffmann, and Luke Stark. [n. d.]. Better, Nicer, Clearer, Fairer: A Critical Assessment of the Movement for Ethical Artificial Intelligence and Machine Learning.

[16] Os Keyes. 2018. The Misgendering Machines: Trans/HCI Implications of Automatic Gender Recognition. *Proceedings of the ACM on Human-Computer Interaction* 2, CSCW (2018), 88.

[17] Min Kyung Lee and Su Baykal. 2017. Algorithmic Mediation in Group Decisions. In *ACM Conference on Computer Supported Cooperative Work*. ACM Press, New York, New York, USA, 1035–1048.

[18] Bruno Lepri, Nuria Oliver, Emmanuel Letouzé, Alex Pentland, and Patrick Vinck. 2017. Fair, Transparent, and Accountable Algorithmic Decision-making Processes. (Aug. 2017), 1–17.

[19] Rishabh Mehrotra, Ashton Anderson, Fernando Diaz, Amit Sharma, Hanna Wallach, and Emine Yilmaz. 2017. Auditing search engines for differential satisfaction across demographics. In *Proceedings of the 26th International Conference on World Wide Web Companion*. International World Wide Web Conferences Steering Committee, 626–633.

[20] Shira Mitchell, Eric Potash, and Solon Barocas. 2018. Prediction-Based Decisions and Fairness: A Catalogue of Choices, Assumptions, and Definitions. *arXiv.org* (Nov. 2018), arXiv:1811.07867. arXiv:1811/1811.07867

[21] FAT ML. 2018. Fairness, Accountability, and Transparency in Machine Learning. Retrieved December 24, 2018 from http://www.fatml.org/

[22] Daniel Neyland. 2015. Bearing Account-able Witness to the Ethical Algorithmic System. *Science, Technology, & Human Values* 41, 1 (Sept. 2015), 50–76.

[23] Safiya Umoja Noble. 2018. *Algorithms of Oppression*. New York University Press.

[24] Cathy O'Neill. 2017. The Ivory Tower Can't Keep Ignoring Tech. Retrieved January 3, 2019 from https://www.nytimes.com/2017/11/14/opinion/academia-tech-algorithms.html

[25] Sundar Pichai. 2018. AI at Google: our principles. Retrieved December 19, 2018 from https://www.blog.google/technology/ai/ai-principles/

[26] Emilee Rader, Kelley Cotter, and Janghee Cho. 2018. Explanations as Mechanisms for Supporting Algorithmic Transparency. In *the 2018 CHI Conference*. ACM Press, New York, New York, USA, 1–13.

[27] Inioluwa Deborah Raji and Joy Buolamwini. 2019. Actionable Auditing: Investigating the Impact of Publicly Naming Biased Performance Results of Commercial AI Products. *Proceedings of the Thirty-Third AAAI Conference on Artificial Intelligence* (Jan. 2019), 1–7.

[28] Ronald E Robertson, David Lazer, and Christo Wilson. 2018. Auditing the Personalization and Composition of Politically-Related Search Engine Results Pages. In *the 2018 World Wide Web Conference*. ACM Press, New York, New York, USA, 955–965.

[29] Casey Ross and Ike Swetlitz. 2018. IBM's Watson Recommended 'unsafe and Incorrect' Cancer Treatments. Retrieved December 17, 2018 from https://www.statnews.com/2018/07/25/ibm-watson-recommended-unsafe-incorrect-treatments/

[30] Vivek K. Singh and Isha Ghosh. 2017. Inferring Individual Social Capital Automatically via Phone Logs. *Proc. ACM Hum.-Comput. Interact.* 1, CSCW, Article 95 (Dec. 2017), 12 pages. https://doi.org/10.1145/3134730

[31] Unity Technologies. 2018. Introducing Unity's Guiding Principles for Ethical AI. Retrieved December 19, 2018 from https://blogs.unity3d.com/2018/11/28/introducing-unitys-guiding-principles-for-ethical-ai/

[32] Daisuke Wakabayashi. 2018. Self-Driving Uber Car Kills Pedestrian in Arizona, Where Robots Roam. Retrieved December 17, 2018 from https://www.nytimes.com/2018/03/19/technology/uber-driverless-fatality.html

[33] Rebecca Wexler. 2018. Life, liberty, and trade secrets: Intellectual property in the criminal justice system. *Stan. L. Rev.* 70 (2018), 1343.