Algorithmic Appeals

Kristen Vaccaro and Karrie Karahalios
University of Illinois Urbana-Champaign
kvaccaro, kkarahal@illinois.edu

INTRODUCTION

Recent work has emphasized the crucial importance that algorithms are playing in all aspects of life: from hiring decisions to credit scores to criminal sentencing [2, 6, 19, 33]. These algorithmic results, often built on opaque machine learning machinery, determine people's futures. This rapidly growing impact has inspired a great deal of work into topics of transparency of algorithms [14, 27], algorithmic awareness [15], and understanding and capturing algorithmic bias [5].

This growing recognition of the impact of algorithmic decision making, and recent legislation mandating explanations for machine learning results [17], has inspired machine learning work focused on explaining outputs [7, 18, 21, 22]. Being able to explain decisions deeply and meaningfully has been proposed as essential for maintaining legitimacy (particularly of governmental decision-making) [6, 12, 28]. There are risks as well as benefits to providing explanations, however. Providing explanations might make it easier for antagonists to game systems [13], with potentially devastating consequences [8].

Even when machine learning decisions can be explained, decision-subjects may not agree with the outcome. Many researchers have highlighted the importance of auditing algorithms [31], to improve them or provide warnings to others. But particularly in high-impact decision scenarios like criminal sentencing, subjects may need to appeal algorithmic decisions. In one case, for example, a judge overruled a plea deal based on an algorithmically generated score: "*when I look at the risk assessment, it is about as bad as it could be*" [2].

We briefly use the scenario of credit scoring to highlight some of the challenges of appeals. If someone is given a low credit score, how can they appeal the decision? In fact, consumers cannot dispute the credit score itself. But they can still manage their score. First, the rough set of features used in computing traditional credit scores is well publicized [23]. Thus users might make sure to pay their credit card bills on time to improve their scores.[1] Credit reporting agencies also allow people to dispute incorrect content in their credit report. If information from another person or a duplicate account is included, for example, consumers can contact both the credit reporting agency and the information provider (e.g. loan company) with a letter describing each inaccurate data point [10]. While the FTC provides a sample letter [9], it is up to each decision-subject to explain and argue their case. And while the credit reporting company must investigate, the dispute might not be resolved – in which case, the company must simply include a dispute statement in future reports.

[1] A perhaps unusual case where users "gaming" the system actually improves outcomes both for the consumer and the credit companies.

This example highlights a few of the challenges presented in appeals – there are many elements of an algorithmic decision-making system (input data, training data, algorithm itself, and output decision) each of which might conceivably be appealed. For example, while consumers might believe that having a higher credit limit *should not* be used as a feature to indicate better creditworthiness (or might question the training data from which that connection was learned), they cannot challenge that aspect of credit scoring. Similarly, the credit scoring scenario indicates that rarely are consumers the only audience for appeals. Indeed, regulators and other businesses may be more important (or powerful) stakeholders than the consumer; after all, credit scores are primarily designed to provide value for credit lending companies (not the consumer).

As this example suggests, making the algorithmic results interpretable is necessary to support algorithmic appeals, but not sufficient. Even if we assume that such explanations are perfect, despite the likely challenges,[2] further issues remain. For example, who should make algorithmic appeals (consumers, regulators, other stakeholders)? What element of the algorithmic decision-making system should appeals operate on – the output, the algorithm itself, or input information (including training data)? And fundamentally, what algorithmic decision-making systems should be subject to appeals?

We briefly overview some of the systems that currently support appeals to understand what appeals typically look like. We then highlight a few cases where algorithmic systems likely *should* support appeals and note challenges for adding appeals.

CURRENT SYSTEMS SUPPORTING APPEALS

There are a number of existing systems that support appeals. By surveying these systems we hope to better understand how appeals typically work offline – and better understand how they might operate for algorthmically generated decisions.

The canonical example of appeals is in the legal system. The American Bar Association notes that it is a "popular misconception" that all cases are eligible for appeals [3]. Instead, appeals occur only when there is a legal basis, typically an alleged material error in the trial. For that reason, appeals are also not a new trial of the case (and do not consider new

[2] The learning sciences literature has not established a framework for providing "good" explanations – instead finding that designing good explanations depends highly on topic (e.g. science vs. literature) and context [Eleanor O'Rourke, personal communication]

evidence), instead focusing on arguments about trial procedure or interpretation of the law. It is only if a judgement is reversed in the appeal that a new trial might occur. Thus, legal appeals showcase both the limitations placed on appeals (not every case can be appealed) and a case where it is primarily the decision-making process itself that is challenged (where input information is held fixed).

We have already briefly outlined appeals in credit scoring. A third area where people often encounter appeals is in insurance claims. In the health care system, for example, a decision-subject frequently appeals after an insurance decision has been made (unlike credit scoring where a consumer might preemtively check). The forms provided for Medicare appeals provide an example; the Department of Health and Human Services asks decision-subjects to explain: "*I do not agree with the determination decision on my claim because* _____" [26]. In this case, the decision-subject can appeal both the input data (having their medical team provide new information to the insurance provider) as well as how the decision was made.

SYSTEMS THAT SHOULD SUPPORT APPEALS

As the current appeal systems make clear, typically appeals are supported in high-impact situations (like health care decisions or criminal proceedings). Given the potential cost of appeals to system providers (as in audit studies [32]), a conservative assessment of which systems *should* support appeals might be prudent. One potential model for evaluating where appeals should be supported is the current legal framework that forbids discrimination in housing, employment, and credit decisions.[3]

Algorithmic decision-making is already growing in these areas. For example, algorithms are playing a growing role is in hiring decisions. While past work has shown that algorithmic cues tend to be isolated to early in the hiring process [4], the growing use of algorithmic hiring tools (even if used only to filter potential candidates) suggests their importance. While many companies celebrate the potential of algorithms to improve diversity [16], many are also aware that bias liability remains [30]. In this case, there might be reasons for the applicant to appeal, but also for internal audits or appeals of system results (e.g., a hiring manager overruling a system decision). However, if a system eliminates an applicant automatically, they may not have the opportunity to appeal that decision, particularly if the reason is never shared with them.

Algorithms also already play an important role in criminal justice. Officials use algorithmic decisions about defendant's risk of recidivism to inform decisions about bail, sentencing and early release [2]. These algorithmic tools use machine learning to infer which input features are more associated with recidivism. As Angwin et al. note, "*defendants rarely have an opportunity to challenge their assessments*" as the calculations that produce the score are considered proprietary [2].

Finally, researchers have explored the variety of new data being added to credit scoring algorithms [19]. Indeed, companies use social network data, e-commerce shopping behavior, device data (e.g. mobile applications installed), and even behavioral analytics including how quickly users scroll through terms of service to produce credit scores. While these approaches are "*already provoking alarm among regulators and consumer-advocacy groups*" the use of new data and technologies make it difficult to apply existing regulatory laws [19], suggesting that appeals may be an important alternative.

CHALLENGES OF APPEALING ALGORITHMIC RESULTS

Appeals may be important for algorithmic decision-making processes, though lack of awareness of algorithmic decision making systems stands in the way. However, there are a number of additional challenges for developing appeals of algorithmic systems, particularly those using machine learning. Appealing the algorithm itself is likely particularly challenging, as algorithms are frequently treated as propriety intellectual property [11]. Thus, revealing algorithmic processes is unlikely because it might allow theft, gaming of the system or it is fundamentally impossible to recreate the decision process.

Appealing input data can also be difficult in systems driven by machine learning. One might want to appeal either the training data used or the input data for an individual decision – however, the scale of data can make this hard. For example, if credit agencies use location data (which is particularly prone to inaccuracy [20]), it may be "practically impossible" to review the information due to the sheer volume of location traces for any given consumer [19]. As Ananny & Crawford note, "*transparency can intentionally occlude*" [1].

Even appealing the output may be hard for users. Users often trust algorithmic systems by default [24, 25, 29]. Recent work has even found that users can be surprisingly reticent to exert control over algorithmic systems, even when they disagree with the result and options for change are well-scaffolded [34].

SUGGESTED PATH FORWARD

While the challenges for developing a system supporting algorithmic appeals seem great, it is an important topic. We thus briefly suggest a path forward. We suggest initially focusing on algorithmic decision-making systems in protected categories, like housing, employment and credit decisions. Indeed, a 2016 White House report focused on just such case studies to highlight the opportunities and challenges of "big data" techniques [33]. Given the existing work on auditing algorithmic systems in such high impact domains, perhaps appeals should be tied to audits (e.g., the audit of criminal risk assessment tools [2]). The algorithm auditing literature also has explored a number of ways of exposing bias (including crowdsourcing), which might be leveraged. Given the aforementioned challenges to users of managing the scale of possible data and the reticence users might experience to intervene on algorithmic outputs, we also recommend controlled experiments designing, testing, and evaluating appeal systems that are usable and give consumers, regulators, and other stakeholders power to intervene in these algorithmic systems that are growing in scope and power.

[3] Via the Civil Rights Act, Fair Housing Act, and Equal Credit Opportunity Act

REFERENCES

1. Mike Ananny and Kate Crawford. 2016. Seeing without knowing: Limitations of the transparency ideal and its application to algorithmic accountability. *new media & society* (2016), 1461444816676645.

2. Julia Angwin, Jeff Larson, Surya Mattu, and Lauren Kirchner. 2016. Machine Bias. `https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing`. (23 May 2016).

3. American Bar Association. 2017. How Courts Work: Steps in a Trial. `https://www.americanbar.org/groups/public_education/resources/law_related_education_network/how_courts_work/appeals.html`. (2017).

4. Moshe Barach, Ming De Leung, and Sibo Lu. 2017. Man Versus Machine: The Effect of Algorithms on Employer Search and Hiring Behavior. In *Academy of Management Proceedings*, Vol. 2017. Academy of Management, 16697.

5. Solon Barocas. 2014. Data mining and the discourse on discrimination. In *Data Ethics Workshop, Conference on Knowledge Discovery and Data Mining*.

6. Reuben Binns. 2017. Algorithmic Accountability and Public Reason. *Philosophy & Technology* (2017), 1–14.

7. Aaron Bornstein. 2016. Is Artificial Intelligence Permanently Inscrutable? `http://nautil.us/issue/40/learning/is-artificial-intelligence-permanently-inscrutable`. (2016).

8. Massimo Calabresi. 2017. Inside Russia?s Social Media War on America. *Time Magazine* (2017).

9. Federal Trade Commission. 2013. Sample Letter for Disputing Errors on Your Credit Report. `https://www.consumer.ftc.gov/articles/0384-sample-letter-disputing-errors-your-credit-report`. (2013).

10. Federal Trade Commission. 2017. Disputing Errors on Credit Reports. `https://www.consumer.ftc.gov/articles/0151-disputing-errors-credit-reports`. (2017).

11. Matthew Crain. 2016. The limits of transparency: Data brokers and commodification. *new media & society* (2016).

12. John Danaher. 2016. The threat of algocracy: reality, resistance and accommodation. *Philosophy & Technology* 29, 3 (2016), 245–268.

13. Nicholas Diakopoulos. 2016. Accountability in Algorithmic Decision Making. *CACM* (2016).

14. Nicholas Diakopoulos and Michael Koliska. 2017. Algorithmic Transparency in the News Media. *Digital Journalism* (2017). DOI: `http://dx.doi.org/10.1080/21670811.2016.1208053`

15. Motahhare Eslami, Aimee Rickman, Kristen Vaccaro, Amirhossein Aleyasen, Andy Vuong, Karrie Karahalios, Kevin Hamilton, and Christian Sandvig. 2015. I always assumed that I wasn't really that close to [her]: Reasoning about Invisible Algorithms in News Feeds. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*. ACM, 153–162.

16. Kelsey Gee. 2017. In Unilever's radical hiring experiment, resumes are out, algorithms are in. `http://www.foxbusiness.com/features/2017/06/26/in-unilevers-radical-hiring-experiment-resumes-\are-out-algorithms-are-in.html`. (26 June 2017).

17. Bryce Goodman and Seth Flaxman. 2016. European Union regulations on algorithmic decision-making and a" right to explanation". *arXiv preprint arXiv:1606.08813* (2016).

18. Patrick Hall, Wen Phan, and SriSatish Ambati. 2017. Ideas on interpreting machine learning. `https://www.oreilly.com/ideas/ideas-on-interpreting-machine-learning`. (2017).

19. Mikella Hurley and Julius Adebayo. 2016. Credit Scoring in the Era of Big Data. *Yale Journal of Law & Technology* 18 (2016), 148.

20. Steven Jacobs. 2015. Report: More Than Half of Mobile Location Data is Inaccurate. `https://perma.cc/43L2-4ULH`. (14 May 2015).

21. Pang Wei Koh and Percy Liang. 2017. Understanding black-box predictions via influence functions. *arXiv preprint arXiv:1703.04730* (2017).

22. Todd Kulesza, Margaret Burnett, Weng-Keen Wong, and Simone Stumpf. 2015. Principles of explanatory debugging to personalize interactive machine learning. In *Proceedings of the 20th International Conference on Intelligent User Interfaces*. ACM, 126–137.

23. Time Magazine. 2014. What Is My Credit Score, and How Is It Calculated? `http://time.com/money/2791957/what-is-my-credit-score/`. (26 May 2014).

24. Bonnie M Muir. 1994. Trust in automation: Part I. Theoretical issues in the study of trust and human intervention in automated systems. *Ergonomics* 37, 11 (1994), 1905–1922.

25. Bonnie M Muir and Neville Moray. 1996. Trust in automation. Part II. Experimental studies of trust and human intervention in a process control simulation. *Ergonomics* 39, 3 (1996), 429–460.

26. Department of Health and Human Services Centers for Medicare & Medicaid Services. *Medicare Redetermination Request Form – 1st Level of Appeal*. Form CMS-20027 (12/10).

27. Frank A Pasquale. 2011. Restoring transparency to automated authority. *Journal on Telecommunications and High Technology Law,* (2011).

28. Frank A Pasquale. 2017. Toward a Fourth Law of Robotics: Preserving Attribution, Responsibility, and Explainability in an Algorithmic Society. *Ohio State Law Journal* (2017).
29. Paul Robinette, Wenchen Li, Robert Allen, Ayanna M Howard, and Alan R Wagner. 2016. Overtrust of robots in emergency evacuation scenarios. In *Human-Robot Interaction (HRI)*.
30. John Rossheim. 2017. Recruiting Algorithms: Appraising their Limits and Benefits. `https://hiring.monster.com/hr/hr-best-practices/recruiting-hiring-advice/strategic-workforce-planning/recruiting-algorithms.aspx`. (2017).
31. Christian Sandvig, Kevin Hamilton, Karrie Karahalios, and Cedric Langbort. 2014a. An algorithm audit. *Data and discrimination: collected essays. New York, NY: New America, Open Technology Institute* (2014), 6–10.
32. Christian Sandvig, Kevin Hamilton, Karrie Karahalios, and Cedric Langbort. 2014b. Auditing algorithms: Research methods for detecting discrimination on internet platforms. *ICA Preconference: Data and discrimination: converting critical concerns into productive inquiry* (2014).
33. Megan Smith, DJ Patil, and Cecilia Muñoz. 2016. *Big Data: A Report on Algorithmic Systems, Opportunity, and Civil Rights*. Technical Report.
34. Jeffrey Warshaw, Tara Matthews, Steve Whittaker, Chris Kau, Mateo Bengualid, and Barton A Smith. 2015. Can an Algorithm Know the Real You?: Understanding People's Reactions to Hyper-personal Analytics Systems. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*. ACM, 797–806.