Contestability in Algorithmic Systems

Kristen Vaccaro
University of Illinois
Urbana-Champaign
kvaccaro@illinois.edu

Daniel Kluttz
University of California Berkeley
dkluttz@berkeley.edu

Karrie Karahalios
University of Illinois
Urbana-Champaign
kkarahal@illinois.edu

Tad Hirsch
Northeastern University
tad.hirsch@northeastern.edu

Deirdre K Mulligan
University of California Berkeley
dmulligan@berkeley.edu

Abstract

As algorithmic (and particularly machine learning) decision making systems become both more widespread and make more important decisions, there are growing concerns about their embedded values and ability to establish legitimacy among decision subjects. We argue that designing for contestability in these systems can assist in surfacing values, aligning system design and use with context, and building legitimacy. However, designing for contestability can be challenging, particularly in systems that are designed to be opaque: systems need to accurately surface embedded values, expose decision making processes in ways that are meaningful for users, support engagement with and allow influence over system performance, and so on. In addition to these technical aspects, designing for contestability may by challenged by the need to protect intellectual property and prevent gaming of the system. In this workshop, we will address goals, audiences, and designs for contestability in algorithmic systems. We hope to develop a taxonomy of contestable systems and understand the value provided by contestability, while bringing together a community to work on this multidisciplinary problem.

Author Keywords

contestability, algorithmic experience

Workshop Goals

Meet colleagues and build collaborations: Make connections across disciplinary boundaries (machine learning, HCI, law, ethics, etc.)

Establish a set of definitions and precedents: Establish a set of definitions and understand the ethical foundations of the concept of contestability.

Construct a taxonomy of contestable systems: Understand what characterizes contestable systems, as well as which systems can/should be designed for contestability.

Decide what's important: Set a research agenda for future work on contestability, developing both a set of questions to be addressed and principles for how to do so.

Communicate beyond HCI: Understand how industry, regulators, and policy can operationalize findings from this area, as well as how to draw contributions and share findings with these larger communities.

Workshop Themes

The concept of contestability in technology systems has deep roots in human-computer interaction. For example mixed-initiative systems were designed for users to compose final outcomes via negotiation with a system [12]. And even the earliest experiments in expert systems revealed that experts must be able to "correct one or more of [the deductive steps and/or facts used] if necessary" [2, 4]. This history of designing for a "conversation" between a user and system has become ever more important as machine decisions have taken on greater and greater impact.

More recently, in order to address potential power asymmetries and distortions of decision making roles in the development and deployment of machine learning systems, Hirsch et al. introduced contestability as a design objective [5]. This design objective is addressed via four goals and accompanying design strategies: 1) *accuracy* via iterative deployment and incentivizing feedback, 2) *legibility* by providing explanations, confidence levels, and traces of system predictions, 3) *training* that explicitly addresses system limitations and allows experimentation to develop shared understandings, and 4) *mechanisms for questioning and disagreeing* with system behavior whether at the individual or aggregate scale. Kluttz and Mulligan generalize these as "mechanisms for users to understand, construct, shape and challenge model predictions" [8].

Importantly, contestability has been defined as being "in band" for the system; unlike simple contestation in which disagreement or attempts to shape the decision making process may be asynchronous, pursued through outside channels, or otherwise externalized, contestability is built into the system to support iteration on the decision making process. This makes contestability a deep system property: the ability to interrogate, investigate, scrutinize the system throughout the process of coming to a joint decision between human and algorithm. It must surface information to the user but also support interaction with and co-construction of the decision making process.

However, in a world with complex sociotechnical systems – in which, for example, machine learning systems may constantly integrate feedback from crowdsourced human-in-the-loop labels – bounding such systems can be challenging. In this workshop, we seek to understand the boundaries: what kinds of systems can and should be designed for contestability? What instrumental value does contestability offer within these systems? And finally, how can contestability be operationalized in different ways and different contexts: timing, expected interaction, and relative authority can vary across contestable systems in ways that matter.

From Contestability to System and Societal Goals

Perhaps one of the most important themes for this workshop is understanding the instrumental value that contestability can offer. When and how can contestability as a system property support contestation, autonomy, safety, the transfer of expert knowledge, construction of domain specific knowledge, maintaining visibility of embedded values [11], even the shaping the allocation of accountability? Different domains, contexts, and tools have different goals (in terms of ethics, safety, legal mandates, professional responsibility, and so on); we hope to explore how contestability can relate to and support these diverse objectives.

Many existing systems are designed for contestation, but even this often fails; recent work on content moderation systems found that users experience the appeal process as "speaking into a void" [10]. Interest has also grown in contesting decision making processes at scale, often in the form of auditing and detecting bias in these systems (for example, collectively auditing Twitter's content moderation [9]).

Participants The workshop will aim for 20-30 participants, to balance the goals of focused discussion and participant diversity (in terms of field, background, methodology, etc.).

Recruitment The workshop hopes to draw together a community of interest with as many perspectives as possible: the HCI community, practitioners of machine learning, ethicists, designers, and professionals in domains with contestability practices (like law, credit scoring or insurance). The call for participation (and website) will be shared with mailing lists, on social media including with the #cscw2019 and other (#mlux, #HumanCenteredAI) hashtags, and via direct invitation.

Submissions Participants are encouraged to submit any form of document that suggests their current thinking on this topic: case studies, position papers, design fictions, etc. Submissions should be <10,000 words, excluding references. We recommend aiming for 4-6 pages. Previously published work is also welcome but should include an ~1000 word cover letter identifying the relevance to the themes of the workshop.

This workshop seeks to carefully explore differences between contestation and contestability, how systems can shift from contestation to contestability, and how these relate to system-specific as well as societal goals for automated decision making.

Who Does Contestability Serve?
These examples of contesting decisions at scale draw in a related question: who contestability can and should serve. How does contestability differ when the co-participants are developers, domain experts, purchasers, regulators, or even decision subjects? Are there other audiences that should be taken into account? Decisions about contestability signal important judgements about relative authority, but defining the audience for contestability of decision making systems also impacts the timing and degree of interaction. For example, regulators might interact with a decision making systems at different times and in different ways than developers, even if all designs aim for contestability. Further, while most work has focused on experts [5, 7], contestability can also serve important functions for decision subjects.

> In the current design of airport security scanners, the system is set for each passenger to either male or female. The TSA or CBP agent decides the setting; if the machine identifies any anomalies relative to the selected gender, the passenger must be examined by an officer of the same gender as the selected setting. In the past, this has been a source of "horrifying experiences" for transgender and gender-nonconforming individuals [3, 6]. However, this is also a setting that could be a point of contestability for the public. TSA policies could instead allow individuals to, in real time, inform the constructs of the decision making.

Ethical Foundations
In addition to understanding the intrinsic and instrumental values that designing for contestability can serve, we hope to explore the ethical foundations of contestability and how it connects to (or provides tensions for) notions of fairness, accountability, and trustworthiness. Contestability has often been conceptualized as a form of procedural justice: feeling one's voice has been heard addresses a fundamental need for perceptions of fairness [13], with particularly strong effects for marginalized or disempowered populations [1]. However, are there alternative frameworks that can drive and shape contestability design? In addition, current designs for contestation don't always improve perceived fairness [10]. Is this because they are designed for contestation instead of deeper notions of contestability? Or simply due to failures in the system design?

Finally, does contestability and the attendant adaptability and impermanence of the decision making process challenge notions of fairness overall, particularly if decision subjects are treated inconsistently over time? In this workshop, we hope to engage deeply from practical questions of how contestability can support ethical practices and expose embedded values to fundamental questions about the ethical frameworks that underlie contestability as a principle.

Activities and Contributions

Activities
The full workshop schedule can be found on the website (contestability.org). Activities will include presentations, discussions and an interactive activity. The early morning will feature brief presentations from a subset of submitted papers from participants and organizers. Discussions will be focused around a set of guiding questions, drawn from the submitted papers but also solicited using hashtags on social media and directly from invited participants before

Interactive Activity
The interactive activity addresses two goals: achieving experiential learning and having fun. This activity should both showcase frustration experienced by many users, while at the same time driving participants (though the iterative nature of the activity) to discover novel and innovative ways of engaging with a decision making system.

To that end, participants will split into two teams (A and B). Team A is confronted with an adverse decision and should contest that decision. The decision will be developed based on a case study drawn from the submissions or provided by the workshop organizers, but will also integrate elements from the earlier discussion. Team B, rather than improving the decision, should instead try to make it worse. Team A will again try to contest the decision, and Team B will again response by making the decision worse. Then the two teams will switch sides and draw on a second case study decision.

the start of the workshop. The goal of this period will be to encourage broad participation and a widely ranging discussion. The early afternoon will include an interactive activity inspired by adversarial learning (left).

Contributions
The outcomes of the workshop will be: 1) the development of a community of interest around the topic of contestability in algorithmic experience design, drawing from a variety of disciplines (HCI, machine learning, ethics, law, medicine, and so on); 2) a website documenting the taxonomies of contestable systems that we develop within the workshop and 3) a collaborative paper summarizing the findings of the workshop – the form of this collaborative paper will depend on the workshop outcomes, in particular, suggestions for communicating findings beyond the HCI community.

Workshop Organizers

Kristen Vaccaro is a PhD candidate in Computer Science at the University of Illinois Urbana-Champaign. Her research focuses on designing algorithmic decision making systems for user agency and control, with two primary mechanisms of interest: control settings and contestability. Her recent work has explored designing for contestability via participatory design workshops, as well as by running large-scale online surveys that assess the effect of common contestability designs.

Karrie Karahalios is a Professor of Computer Science at the University of Illinois in Urbana-Champaign, the director of the Social Spaces Group, and the co-director of the Center for People and Infrastructures. Her work focuses on the signals that people emit and perceive in social computer mediated communication. More recently, she has explored how algorithmic curation alters these signals and people's perception of communication. Karahalios studies existing systems and builds infrastructures for new communication systems (that move control to people, allow for inferences of bias and fairness, and evaluate algorithm explainability).

Deirdre K. Mulligan is an Associate Professor in the School of Information at UC Berkeley, a faculty Director of the Berkeley Center for Law & Technology, a co-organizer of the Algorithmic Fairness & Opacity Working Group, an affiliated faculty on the Hewlett funded Berkeley Center for Long-Term Cybersecurity, and a faculty advisor to the Center for Technology, Society & Policy. Mulligan's research explores legal and technical means of protecting values such as privacy, freedom of expression, and fairness in emerging technical systems.

Daniel Kluttz is a Postdoctoral Scholar at the UC Berkeley School of Information's Algorithmic Fairness and Opacity Working Group (AFOG). Drawing from intellectual traditions in organizational theory, law and society, and technology studies, Kluttz's research is oriented around two broad lines of inquiry: 1) the formal and informal governance of economic and technological innovations, and 2) the organizational and legal environments surrounding such innovations. He holds a PhD in sociology from UC Berkeley and a JD from the UNC-Chapel Hill School of Law.

Tad Hirsch is a Professor of Art + Design at Northeastern University, where he conducts research and creative practice at the intersection of design, engineering, and social justice. He is currently developing automated assessment and training tools for addiction counseling and mental health; prior work has tackled such thorny issues as human trafficking, environmental justice, and public protest. His pioneering work on "Designing Contestability" identified contestability as a new principle for designing systems that evaluate human behavior.

REFERENCES

1. Emile G Bruneau and Rebecca Saxe. 2012. The power of being heard: The benefits of 'perspective-giving' in the context of intergroup conflict. *Journal of experimental social psychology* 48, 4 (2012), 855–866.
2. Bruce G. Buchanan and Edward H. Shortliffe. 1984. *Rule Based Expert Systems: The Mycin Experiments of the Stanford Heuristic Programming Project (The Addison-Wesley Series in Artificial Intelligence)*. Addison-Wesley Longman Publishing Co., Inc., Boston, MA, USA.
3. Dawn Ennis. 2015. Her Tweets Tell One Trans Woman's TSA Horror Story. The Advocate. (September 2015). `https://www.advocate.com/transgender/2015/9/22/one-trans-womans-tsa-horror-story`
4. G Anthony Gorry and others. 1973. Computer-assisted clinical decision making. *Methods of Information in Medicine* 12, 1 (1973), 45.
5. Tad Hirsch, Kritzia Merced, Shrikanth Narayanan, Zac E Imel, and David C Atkins. 2017. Designing contestability: Interaction design, machine learning, and mental health. In *Proc. DIS*. ACM, 95–99.
6. Carina Julig. 2018. How Airport Security Makes Travel Traumatic for Butches and Trans Folks, Whether or Not They Pass. Slate. (June 2018). `https://slate.com/human-interest/2018/06/gendered-airport-security-makes-travel-traumatic\for-butches-and-trans-folks.html`
7. Daniel Kluttz, Nitin Kohli, and Deirdre K Mulligan. 2018. Contestability and Professionals: From Explanations to Engagement with Algorithmic Systems. *Available at SSRN 3311894* (2018).
8. Daniel Kluttz and Deirdre K Mulligan. 2019. Automated decision support technologies and the Legal Profession. *Berkeley Technology Law Journal* (2019).
9. J. Nathan Matias, Amy Johnson, Whitney Erin Boesel, Brian Keegan, Jaclyn Friedman, and Charlie DeTar. 2015. Reporting, reviewing, and responding to harassment on Twitter. *Available at SSRN 2602018* (2015).
10. Sarah Myers West. 2018. Censored, suspended, shadowbanned: User interpretations of content moderation on social media platforms. *New Media & Society* 20, 11 (2018), 4366–4383.
11. Helen Nissenbaum. 2001. How computer systems embody values. *Computer* 34, 3 (2001), 120–119.
12. David G Novick and Stephen Sutton. 1997. What is mixed-initiative interaction. In *Procedings of the AAAI Spring Symposium on Computational Models for Mixed Initiative Interaction*. 114–116.
13. Tom R Tyler, Kenneth A Rasinski, and Nancy Spodick. 1985. Influence of voice on satisfaction with leaders: Exploring the meaning of process control. *Journal of Personality and Social Psychology* 48, 1 (1985), 72.