Realizing AI in Healthcare: Challenges Appearing in the Wild

Tariq Osman Andersen
University of Copenhagen, Denmark
Vital Beats ApS, Denmark

Francisco Nunes
Fraunhofer Portugal AICOS

Lauren Wilcox
Google, Inc, Mountain View, CA, USA
Georgia Tech, Atlanta, GA, USA

Elizabeth Kaziunas
AI Now, New York University, USA

Stina Matthiesen
University of Copenhagen, Denmark

Farah Magrabi
Macquarie University, Australia

ABSTRACT

The last several years have shown a strong growth of Artificial Intelligence (AI) technologies with promising results for many areas of healthcare. HCI has contributed to these discussions, mainly with studies on explainability of advanced algorithms. However, there are only few AI-systems based on machine learning algorithms that make it to the real world and everyday care. This challenging move has been named the "last mile" of AI in healthcare, emphasizing the sociotechnical uncertainties and unforeseen learnings from involving users in the design or use of AI-based systems. The aim of this workshop is to set the stage for a new wave of HCI research that accounts for and begins to develop new insights, concepts, and methods, for transitioning from development to implementation and use of AI in healthcare. Participants are invited to collaboratively define an HCI research agenda focused on healthcare AI in the wild, which will require examining end-user engagements and questioning underlying concepts of AI in healthcare.

CCS CONCEPTS

• **Information systems**; • **Information systems applications.**;

KEYWORDS

artificial intelligence, human computer interaction

ACM Reference Format:
Tariq Osman Andersen, Francisco Nunes, Lauren Wilcox, Elizabeth Kaziunas, Stina Matthiesen, and Farah Magrabi. 2021. Realizing AI in Healthcare: Challenges Appearing in the Wild. In *CHI Conference on Human Factors in Computing Systems Extended Abstracts (CHI '21 Extended Abstracts), May 08–13, 2021, Yokohama, Japan.* ACM, New York, NY, USA, 5 pages. https://doi.org/10.1145/3411763.3441347

1 BACKGROUND

The integration of Artificial Intelligence (AI) into multiple areas of our lives is changing the world as we know it, with profound social, political, and economic effects. New forms of human–computer interaction based on predictive models and datafication at large are actively co-configuring everyday lives and professional practices. However, in healthcare, very few AI-systems based on machine learning algorithms have been deployed and successfully embedded in real life. This stands in stark contrast to the many examples of lab experiments that demonstrate technical feasibility and promising AI-performance. In fact, during the past two years, publications on medical AI have increased almost tenfold [3] and individual short-term disease predictions seem to present appropriate results for multiple conditions including, diabetes, cancer, heart -, and mental illness.

In HCI, research on AI has benefitted from a focus on improving the intelligibility of advanced algorithms. XAI —explainable AI research—has advanced our understanding and come a long way in developing methods and toolkits to tackle interaction issues related to interpretable, fair, accountable and transparent algorithms [1]. There is extensive research on XAI and approaches to achieving explainability are diverse, ranging from visualizing internal information processing to making one neural network explain another [12]. However, there have been relatively few experimental studies of whether these models achieve their intended effects when deployed in real-world settings [19]. Most work in explainable AI uses only the researchers' intuition of what constitutes a 'good' explanation [18]. In recent CHI publications, Abdul et al. [1] and Wolf [11] also raise this concern that there is a lack of understanding of how explainability emerges in real-world deployments of AI systems. This tells us that there has been a race for getting the technology right before exposing human end-users to the new promising AI tools— in or close to their situated, everyday practices.

The integration of AI in healthcare has been named the "last mile" [3, 7], emphasizing a considerable effort in dealing with the sociotechnical uncertainties and unforeseen learnings that emerge from studies of design and use of AI-based tools in healthcare. For different reasons, "near-live" evaluation [16] and real-world deployment is often considered out of scope for studies examining AI-based technologies in healthcare. While there are good reasons for postponing integration in care settings, such as ensuring computational precision and responsible deployment, there is an inherent need for research that studies the experimental and empirical effects of embedding AI into sociotechnical healthcare environments and discusses implications for design of AI in healthcare.

While qualitative empirical research and experimental studies of AI systems are still limited, recent studies have begun to highlight key sociotechnical issues. Sendak et al. [13], for example, detailed the development, implementation, and use of Sepsis Watch, a machine learning-driven system designed to help clinicians in the early diagnosis and treatment of sepsis in a hospital setting. Their paper highlights the importance of the implementation phase and the work of integrating AI-enabled health systems into the organizational context of the hospital and work practices of clinicians – topics of long interest to researchers in the HCI/CSCW community.

CHI '21 Extended Abstracts, May 08–13, 2021, Yokohama, Japan

ACM ISBN 978-1-4503-8095-9/21/05.
https://doi.org/10.1145/3411763.3441347

Elish and Watkins [8] emphasizes this point further by stressing that Sepsis Watch is a "sociotechnical system, not just a machine learning model" and that "repair work" - the work of implementation and involving clinicians in creating a new set of practices – is critical for avoiding healthcare AI systems remain potential solutions. Another recent example is the work of Beede et al. [2], who conducted a human-centered study of how a deep-learning-based eye screening system affected local workflows and nurse–patient interactions in a busy hospital setting. They describe how seemingly trivial environmental factors and system design decisions affected the nurses' workflows, and how social and environmental factors at the deployment site affected the system performance. In their qualitative study, several well-known AI-related concepts get unpacked and learnings from medical AI in-the-wild raise important questions about how to take end-users' contexts and workflows into consideration early on when designing and adapting intelligent algorithms for healthcare. This teaches us that concepts from data science like "prediction accuracy" may have a certain meaning in a lab environment but multiple alternative or contradictory meanings in real-world situations where clinicians or consumers wrestle with connecting AI-based technologies to situated action. Moreover, Cai et al. [4] show that seemingly trivial expectations of what AI can do for physicians in their practices have profound implications for making real-world AI deployment successful and Kocaballi et al. [15] show that professional autonomy gets challenged by AI when trying to intelligently automate clinical documentation work. Through this work, new concepts like "expectations" and "professional autonomy" emerge as important tools for an HCI perspective on developing AI technologies in healthcare.

Although fully deployed AI systems based on machine learning in clinical contexts are still rare, AI technologies are increasingly shaping people's experiences of health and wellness across a growing number of everyday technologies, from chatbot apps to wearable fitness devices to workplace wellness programs. Furthermore, AI is increasingly being used in a wide range of healthcare administrative processes, such as classifying high risk patients to automating the scheduling of clinical appointments. Of particular concern is need for HCI researchers to consider how systemic issues around social determinants of health such as racism, sexism, and unequal access to medical resources are addressed in AI-Health design processes.

Also, of relevance to these emerging discussions on AI and Health is a growing HCI, CSCW, and STS literature on people's varied interactions with health data and algorithms across health and wellness contexts. Importantly, this work has drawn attention to the materiality of data tracking, social and emotional aspects of new health technology adoption and use, everyday practices and labor around health data, complex infrastructural relations around data and medical care, and offered critical frameworks for understanding the social impacts of health and wellness technologies [9, 10, 14, 17]. Within this collective work, technofeminist scholarship and theories of care have emerged as a generative analytic framework for conceptualizing new types of human-nonhuman interactions and data-body-machine entanglements. As AI-based technologies in healthcare are increasingly integrated into a wider range of medical settings and everyday activities, such literature offers a starting point for rethinking purely technical approaches to AI and opening up space for diverse approaches to AI in healthcare that center a wider range of stakeholder needs, types of interactions, and social concerns.

Other relevant resources to look for and to be inspired by is the wider social science engagement with algorithmic research. Between the technological excitement, scholars have pointed to the urgency of understanding how exactly people engage and interact with algorithmic technologies in practice. Sociologists have pointed to how professionals in different domains engage in situated actions and apply various strategies of "algorithmic resistance" to minimize the impact of algorithmic tools in their daily work. Sociologist Angéle Christin, for example, has recently called for more ethnographic studies on actual practices around and interactions with algorithmic technologies as she argues that the current approaches to understand algorithmic systems remain decontextualized [5, 6] and that missing from the discussion are *"the actual practices, applications, and uses surrounding algorithmic tools in the fields and organizations where they unfold"* [5].

2 GUIDING THEMES AND RESEARCH AGENDA

With this workshop, we wish to set the stage for a new wave of HCI research that accounts for and begins to develop new insights, concepts, and methods, for going the "last mile" with AI in health. The aim of the workshop is to develop a conceptual framework that maps out a theoretical space from which new studies can be inspired. We will do this by inviting participants to empirically ground and qualitatively unpack existing key concepts such as "explainability", "trust", and "accuracy" as well as by developing new concepts such as "expectations" and "repair work" that can support researchers and developers in taking a more sociotechnical and human-centered HCI perspective, when working to realize the benefits of AI into transforming care delivery.

2.1 Topics of interest include:

Empirical studies of healthcare AI in the wild: What does empirical and qualitative studies of AI in the wild or "near live" in healthcare teach us? What insights about "explainability", "trust", "accuracy", or other concepts are found when studied close to end users' real world practices? What new concepts are more useful for understanding what is involved when going the "the last mile" of AI in health?

Methodological challenges for studying AI systems in the wild: How have HCI researchers and designers navigated site-specific challenges in studying AI-mediated health systems? What research techniques and design methods have been useful for gaining insight into the everyday use of AI systems?

Human-centered AI in the HCI design process: What Human-AI interaction design and user research training requirements are needed for safe and effective clinical and consumer systems? What is unique about AI-based systems' user interfaces and what concepts are appropriate to describe or explain them?

Theories and framings for broadening HCI approaches to AI and health: What disciplinary perspectives and analytic frameworks are needed, and which theories support a better understanding of the social, organizational, and ethical dimensions of AI and Health?

Sociotechnical impacts and critical perspectives: Are there health contexts where we should not use AI? What lessons can healthcare researchers and designers learn from the critical push-back of AI systems in other social domains, such as the use of facial recognition technology in law enforcement? How can we benefit from understanding and exploring various forms of sociotechnical issues like algorithmic resistance as it takes place in the wild? How can we study algorithmic resistance as a way to inform design?

Expanding the conception of what AI-tools in healthcare should offer: How can we take seriously the argument that algorithmic output is only as good as the input we have available and thus, develop AI tools that focus on supporting practice as it takes place in the wild rather than achieving the best metrics for precision and recall? How can we challenge the design of AI-systems that are driven by the "mythology of big Data" that assumes the larger the data set the higher the intelligence/knowledge?

3 ORGANIZERS

This group of workshop organizers are currently engaged with studies at the intersection of AI and human-centered HCI and have expertise from several earlier studies that unpacks the sociotechnical issues emerging from in-the-wild, design-oriented, and experimental studies of intelligent algorithms for health. Collectively, the group represents the diversity of scholars we hope to recruit and connects research from different continents across the world. The organizers also have experience running successful workshops at CHI, CSCW, and other HCI conferences such as the Participatory Design conference and the Workshop on Interactive Systems in Healthcare.

Elizabeth Kaziunas is a postdoctoral researcher at AI Now, a research institute at New York University investigating the social impacts of artificial intelligence. Her ethnographic research examines social, organizational, and ethical contexts of health information systems, and is currently focused on understanding the practices and politics of care in relation to personal data and AI-mediated technologies. She received a Ph.D. from the University of Michigan's School of Information, where her thesis investigated designing technology to support the lived experience of chronic illness.

Farah Magrabi is an Associate Professor at the Australian Institute of Health Innovation, Macquarie University. She has a background in Electrical and Biomedical Engineering with over 15 years' experience in Health Informatics focusing on the design and evaluation of digital health technologies for clinicians and consumers. Farah is currently investigating the patient safety risks of artificial intelligence (AI) in healthcare. She co-chairs the Australian AI Alliance's working group on safety, quality and ethics; the International Medical Informatics Association's (IMIA) working group on Technology Assessment & Quality Development (2013-present).

Francisco Nunes is a Senior Researcher at Fraunhofer Portugal AICOS, working on the human-centered Design team. His research is concerned with the user research, design, and evaluation of mobile and AI-based self-care technologies. Francisco has a PhD in Human-Computer Interaction from TU Wien (2017).

Lauren Wilcox is an associate professor in the School of Interactive Computing at Georgia Tech and research lead in the Google Wellbeing Lab. She brings over thirteen years of experience conducting human-centered computing research in service of human health and well-being. Previously at Google Health, Wilcox led initiatives to align AI advancements in healthcare with the needs of clinicians, patients, and their family members and recently co-authored some of the first papers at SIGCHI conferences describing human-centered issues when deploying AI-based systems in clinical practices. Wilcox was an inaugural member of the ACM Future of Computing Academy and frequently serves on the organizing and technical program committees for premier conferences in the field (e.g., ACM CHI).

Stina Matthiesen is an assistant professor in Software, Data, People & Society at the Department of Computer Science, University of Copenhagen. She is inspired by critical studies on race, technology and datafication and has previously explored how stereotypes and implicit bias manifest itself in the everyday practices of global software development (GSD). Stina is currently investigating the emotional labor of chronic patients, as well as analyzing and co-designing data-driven and AI-based technologies for patient-clinician collaboration and clinical decision support in cardiac care.

Tariq Osman Andersen is an assistant professor in Software, Data, People, and Society at the Department of Computer Science, University of Copenhagen and he is co-founder and head of research in a scale-up medical-AI company called Vital Beats. His research is concerned with large-scale and long-term co-design of digital health and revolves around experimental studies of AI-based clinical decision-making and patient-clinician interaction in cardiac care. Tariq holds a PhD from the University of Copenhagen (2012).

4 LINK TO WEBSITE

http://bit.ly/RealizingAIinHealthcareWS

5 POSITION PAPER SUBMISSION AND SELECTION

5.1 Position paper submission and selection

Workshop participants will submit position papers based on their work. Our goal is to explore empirical work and question underlying assumptions, so we will ask authors to frame their position papers to discuss specific insights, concepts, and methods, related to implementation and use of AI in healthcare. Authors will be able to draw on their artifacts, empirical results, theoretical reflections, and explorations to contribute to advance an agenda of an HCI perspective for AI in healthcare. Moreover, and since the reflection of the authors starts prior to the workshop, we will be able to have more advanced discussions when at the workshop.

We encourage submissions from academics, researchers, engineers, designers, data scientists, social scientists, medical professionals, and patients, who are interested in broadening reflections around human-centered AI in healthcare and the sociotechnical challenges appearing in the wild. Position papers should be up to 5 pages (excluding references) and submitted in the CHI 2021 Extended Abstracts Format. Non-academics will be able to send a motivation letter explaining their interest to participate in the workshop. The submitted papers should be sent to chi2021wsaihealth@gmail.com and will be lightly reviewed by the

workshop organizers. Selection will be based on their quality, originality, diversity, and relevance to the workshop topic and vision.

5.2 Short video submission

The authors with accepted position papers will prepare and share a 2-3 minute video that explains their contribution. The video will be submitted some weeks before the workshop and will be included in the workshop website. The goal of the video is to ensure that all participants are familiar with reflections from other participants before they join the workshop, which shall increase the quality of the discussions, reflections, and critique.

5.3 Preparation of case studies and materials for discussion

The workshop organizers will prepare a set of materials to facilitate discussion among groups at the workshop. This includes selecting artifacts, quotes, or short videos that can help reflections about concepts and narratives that demonstrates some aspect of healthcare AI in the wild. Some of these materials will be based on the contributions from the workshop participants, while others will come from the fieldwork and work of the organizers themselves.

6 WORKSHOP ACTIVITIES AND STRUCTURE

We expect 20-30 participants at the workshop. If more participants have relevant submissions, we will consider adapting the format for supporting a larger number of attendants. All workshop activities will take place online to ensure that all participants can engage safely. We will draw on videoconference technologies (e.g., Zoom or Teams) and collaborative discussion tools (e.g., Miro or Mural) to support synchronous activities during the workshop. The website of the workshop will also take an important role in the workshop, by familiarizing participants with the contributions from other workshop participants.

Show and tell: After welcome messages and brief introductions, each participant will have 3-5 minutes to present their work. We will invite participants to talk about their work as if they were doing show and tell, framing their reflections with artefacts, quotes, or pictures that can help other participants remember the contributions when engaging in workshop discussions. Following each presentation, there will be time for one or two questions. Organizers will make notes during the session to compile a list of themes to be discussed in the remaining sessions. These notes will be shared with participants (e.g., Google Docs Document).

World Café method: The workshop organizers will derive topics of interest from the participants' presentations and discussions and assign each topic to a break-out room. Participants will be invited to discuss these themes in small groups, and, after some time, rotate them to another table that discusses a different topic. Collaborative discussion tools (e.g., Miro, Mural) will support note taking and reflections with the group. Moderating each table will be the workshop organizers, which will, in addition to facilitating conversations, bring materials (e.g., quotes, pictures, videos) to the discussion that were prepared before the workshop. While participants rotate, moderators will stay in the same virtual room ensuring a connection between the work of different groups.

Whole group reflection: The insights from the four sessions of World Café will be shared in whole group sessions, making all participants familiar with the insights from the groups. Moreover, we will discuss future research visions and ways of continuing discussions during these activities.

Workshop timetable: See workshop website.

7 POST-WORKSHOP PLANS

We will upload all papers and videos from workshop participants to the workshop website. Having obtained authorization from the participants, we will also share pictures and short videos of some discussions to help document the findings of the workshop. The conversations and reflections started at the workshop will proceed to a journal special issue. Some of the organizers are about to submit a closely related special issue to TOCHI and workshop participants will be invited to contribute with expanded versions of their work.

8 CALL FOR PARTICIPATION

We invite submissions for a two-day virtual workshop, which is focused on the sociotechnical challenges of embedding AI in real-world healthcare settings. Participants are invited to collaboratively define an HCI research agenda for a new wave of HCI research on AI in healthcare, which will require examining end-user engagements with AI and questioning underlying concepts of AI in healthcare. The workshop is motivated by the difficulties of integrating AI into everyday care, known as the "last mile" of AI in healthcare. Position papers should be 2-5 pages long in the CHI 2021 Extended Abstract format and may address topics related to the intersections of HCI, AI, Health Informatics, and social science. This includes but is not limited to: ongoing work; reflections on past work; concept development; combining methods from HCI and design to AI; and emergent sociotechnical, ethical, and political challenges. The due date for submissions is no later than February 21, 2021 by email to chi2021wsaihealth@gmail.com. Participants will be selected based on the quality and clarity of their submissions as they reflect the interests of the workshop. Notifications will go out no later than March 19, 2021. At least one author of each accepted position paper must virtually attend the workshop. All participants need to register for both the workshop and at least one day of the (virtual) conference.

ACKNOWLEDGMENTS

Francisco Nunes was funded by project COTIDIANA funded under the AAL JP, by the European Commission and the National Funding Authorities of Portugal, Austria, and Switzerland.

REFERENCES

[1] Ashraf Abdul, Jo Vermeulen, Danding Wang, Brian Y. Lim, and Mohan Kankanhalli. 2018. Trends and Trajectories for Explainable, Accountable and Intelligible Systems: An HCI Research Agenda. In Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems (CHI '18). Association for Computing Machinery, New York, NY, USA, Paper 582, 1–18. DOI:https://doi.org/10.1145/3173574.3174156

[2] Emma Beede, Elizabeth Baylor, Fred Hersch, Anna Iurchenko, Lauren Wilcox, Paisan Ruamviboonsuk, and Laura M. Vardoulakis. 2020. A Human-Centered Evaluation of a Deep Learning System Deployed in Clinics for the Detection of Diabetic Retinopathy. In Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems (CHI '20). Association for Computing Machinery, New York, NY, USA, 1–12. DOI:https://doi.org/10.1145/3313831.3376718

[3] Federico Cabitza, Andrea Campagner, and Clara Balsano. 2020. Bridging the "last mile" gap between AI implementation and operation: "data awareness" that matters. Ann Transl Medicine 8, 7 (2020), 501. DOI:https://doi.org/10.21037/atm.2020.03.63

[4] Carrie J. Cai, Samantha Winter, David Steiner, Lauren Wilcox, and Michael Terry. 2019. "Hello AI": Uncovering the Onboarding Needs of Medical Practitioners for Human-AI Collaborative Decision-Making. Proc. ACM Hum.-Comput. Interact. 3, CSCW, Article 104 (November 2019), 24 pages. DOI:https://doi.org/10.1145/3359206

[5] Angèle Christin. 2017. Algorithms in practice: Comparing web journalism and criminal justice. Big Data Soc, 4, 2 (2017) DOI:https://doi.org/10.1177/2053951717718855

[6] Angèle Christin. 2020. The ethnographer and the algorithm: beyond the black box. Theor Soc (2020), 1–22. DOI:https://doi.org/10.1007/s11186-020-09411-3

[7] Enrico Coiera. 2019. The Last Mile: Where Artificial Intelligence Meets Reality. J Med Internet Res 21, 11 (2019), e16323. DOI:https://doi.org/10.2196/16323

[8] Madeleine Clare Elish and Elizabeth Anne Watkins. 2020. Repairing Innovation: A Study of Integrating AI in Clinical Care. Data & Society. Retrieved from https://datasociety.net/wp-content/uploads/2020/09/Repairing-Innovation-DataSociety-20200930-1.pdf

[9] Kathleen H. Pine, Claus Bossen, Yunan Chen, Gunnar Ellingsen, Miria Grisot, Melissa Mazmanian, and Naja Holten Møller. 2018. Data Work in Healthcare: Challenges for Patients, Clinicians and Administrators. In Companion of the 2018 ACM Conference on Computer Supported Cooperative Work and Social Computing (CSCW '18). Association for Computing Machinery, New York, NY, USA, 433–439. DOI:https://doi.org/10.1145/3272973.3273017

[10] Sarah E. Fox, Amanda Menking, Jordan Eschler, and Uba Backonja. 2020. Multiples Over Models: Interrogating the Past and Collectively Reimagining the Future of Menstrual Sensemaking. ACM Trans. Comput.-Hum. Interact. 27, 4, Article 22 (September 2020), 24 pages. DOI:https://doi.org/10.1145/3397178

[11] Christine T. Wolf. 2019. Explainability scenarios: towards scenario-based XAI design. In Proceedings of the 24th International Conference on Intelligent User Interfaces (IUI '19). Association for Computing Machinery, New York, NY, USA, 252–257. DOI:https://doi.org/10.1145/3301275.3302317

[12] Riccardo Guidotti, Anna Monreale, Salvatore Ruggieri, Franco Turini, Fosca Giannotti, and Dino Pedreschi. 2018. A Survey of Methods for Explaining Black Box Models. ACM Comput. Surv. 51, 5, Article 93 (January 2019), 42 pages. DOI:https://doi.org/10.1145/3236009

[13] Mark Sendak, Madeleine Clare Elish, Michael Gao, Joseph Futoma, William Ratliff, Marshall Nichols, Armando Bedoya, Suresh Balu, and Cara O'Brien. 2020. "The human body is a black box": supporting clinical decision-making with deep learning. In Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency (FAT '20). Association for Computing Machinery, New York, NY, USA, 99–109. DOI:https://doi.org/10.1145/3351095.3372827

[14] Elizabeth Kaziunas, Mark S Ackerman, Silvia Lindtner, and Joyce M Lee. 2017. Caring through Data: Attending to the Social and Emotional Experiences of Health Datafication. In (Proceedings of the SIGCHI Conference on Human Factors in Computing Systems), 2260–2272. DOI:https://doi.org/10.1145/2998181.2998303

[15] A Baki Kocaballi, Kiran Ijaz, Liliana Laranjo, Juan C Quiroz, Dana Rezazadegan, Huong Ly Tong, Simon Willcock, Shlomo Berkovsky, and Enrico Coiera. 2020. Envisioning an artificial intelligence documentation assistant for future primary care consultations: A co-design study with general practitioners. J Am Med Inform Assn (2020), ocaa131. DOI:https://doi.org/10.1093/jamia/ocaa131

[16] Alice C. Li, Joseph L. Kannry, Andre Kushniruk, Dillon Chrimes, Thomas G. McGinn, Daniel Edonyabo, and Devin M. Mann. 2012. Integrating usability testing and think-aloud protocol analysis with "near-live" clinical simulations in evaluating clinical decision support. Int J Med Inform 81, 11 (2012), 761–772. DOI:https://doi.org/10.1016/j.ijmedinf.2012.02.009

[17] Stine Lomborg, Henriette Langstrup, and Tariq Osman Andersen. 2020. Interpretation as luxury: Heart patients living with data doubt, hope, and anxiety. Big Data Soc 7, 1 (2020), 205395172092443. DOI:https://doi.org/10.1177/2053951720924436

[18] Tim Miller. 2018. Explanation in artificial intelligence: Insights from the social sciences. Artif Intell 267, (2018), 1–38. DOI:https://doi.org/10.1016/j.artint.2018.07.007

[19] Forough Poursabzi-Sangdeh, Daniel G Goldstein, Jake M Hofman, Jennifer Wortman Vaughan, and Hanna Wallach. 2018. Manipulating and Measuring Model Interpretability. arXiv preprint arXiv:1802.07810 (2018).