Explaining AI Fairly (Well)

Margaret Burnett
Department Name
Oregon State University
Corvallis, Oregon USA
burnett@eecs.oregonstate.edu

ABSTRACT

How can the field of Explainable AI (XAI) get from where we are now, explaining *some* aspects of AI fairly well, to where we need to be—explaining AI *fairly and well*? In this keynote, I'll talk about three critical challenges to our field, focusing especially on the third of these: explaining AI fairly.

CCS CONCEPTS

• **Computing methodologies** → *Intelligent agents;* • Human-centered computing → **Human-Computer Interaction (HCI)**

KEYWORDS

Explainable AI; explaining to diverse populations; biased explanations; XAI challenges

ACM Reference format:

Margaret Burnett. 2019. Explaining AI Fairly (Well): In Joint Proceedings of the ACM IUI 2019 Workshops, Los Angeles, USA, March 20, 2019, 1 pages

1 Overview

Explainable AI (XAI) has started experiencing explosive growth, echoing the explosive growth that has preceded it of AI becoming used for practical purposes that impact the general public. This spread of AI into the world outside of research labs brings with it pressures and requirements that many of us have perhaps not thought about deeply enough.

In this keynote address, I will explain why I think we have a long way to go before we'll be able to achieve our long-term goal: to explain AI *well*.

One way to characterize our current state is that we're doing "fairly well", doing *some* explaining of *some* things. In a sense, this is reasonable: the field is young, and still finding its way. However, moving forward demands progress in (at least) three areas.

(1) *How* we go about XAI research: Explainable AI cannot succeed if the only research foundations brought to bear on it are AI foundations. Likewise, it cannot succeed if the only foundations used are from psychology, education, etc. Thus, a challenge for our emerging field is how to conduct XAI research in a truly effective multi-disciplinary fashion, that is based on not only what we can make algorithms do, but also on solid, well-founded principles of explaining the complex ideas behind the algorithms to real people. Fortunately, a few researchers have started to build such foundations.

(2) *What* we can succeed at explaining: So far, we as a field are doing a certain amount of cherry picking as to what we explain. We tend to choose what to explain by what we can figure out how to explain—but we are leaving too much out. One urgent case in point is the societal and legal need to explain fairness properties of AI systems.

The above challenges are important, but the field is already becoming aware of them. Thus, this keynote will focus mostly on the third challenge, namely:

(3) *Who* we can explain to. Who are the people we've even tried to explain AI to, so far? What are the societal implications of who we explain to well and who we do not?

Our field has not even begun to consider this question. In this keynote I'll discuss why we have to explain to populations to whom we've given little thought—diverse people in many dimensions, including gender diversity, cognitive diversity, and age diversity.

Addressing all of these challenges is necessary before we can claim to explain AI fairly *and* well.

ACKNOWLEDGMENTS

This work has been supported in part by DARPA #N66001-17-2-4030 and NSF #1528061. Any opinions, findings and conclusions or recommendations expressed are the authors' and do not necessarily reflect the views of NSF, DARPA, the Army Research Office, or the US government.

IUI Workshops'19, March 20, 2019, Los Angeles, USA.