

Co-design and Ethical Artificial Intelligence for Health: Myths and Misconceptions

Joseph Donia, MSc¹ & James A. Shaw, PhD^{1,2}

Background & Introduction

The contemporary field of artificial intelligence / machine learning (AI/ML) is dynamic and rapidly growing. Although AI/ML technologies are multi-purpose, they are particularly consequential in health care. One strategy for anticipating and addressing the potential benefits and harms of AI/ML for health is patient and public involvement in the design of those technologies, often referred to as co-design. **Co-design and its variants have a diverse intellectual and practical history**, however, and have been conceptualized in many different ways. Moreover, **the meaning and value of co-design is challenged by AI/ML systems**. Informed by perspectives from critical data studies, and critical digital health studies, **we outline five myths and misconceptions arising from co-design discourse related to AI/ML for health care**. We use ‘co-design’ as an umbrella term for approaches that involve end-users, patients, or publics in *any* stage of the design process.

Myth #1: ‘Better’ involvement strategies result in ‘better’ design outcomes

The central point advanced with Myth #1 is that **‘better’ involvement (indicated by breadth, depth, or impact of involvement on decision-making) does not imply a stronger focus on the entirety of the socio-technical system**, much of which is out of view for both users and designers of AI/ML technologies. The consequences of newly designed AI/ML technologies (e.g. the linked effects of data commodification; surveillance; and individualization of care), remain marginal in most approaches to co-design.

Any claims to the ethical standing of co-design must therefore be evaluated against the sociotechnical configurations it produces, rather than the proximate effects of co-design processes on stand-alone products (which arguably, never stand alone)

Myth 2: ‘Good’ co-design increases the agential capacities of patients and publics

The central point advanced with Myth #2 is that **‘better’ involvement does not mean that people are entirely free from agential constraints that inevitably shape their participation in design activities**. These constraints do not only apply to patients and publics, but others implicated in design processes, too. Ethical co-design of AI/ML technologies for health must engage with this broader ecosystem of design, expanding the view of who and what is considered relevant.

Myth #3: Representation and inclusion reduces risk of harms of designed artifacts

The central point advanced with Myth #3 is that **the inclusion of communities in design processes does not necessarily address the upstream causes that lead to marginalization in the first place**. As such, they risk supplanting consideration of those causes with easy-to-use technological solutions that may exacerbate societal inequities. Rather than emphasizing representation and inclusion for its own sake, ethical co-design ought to include provisions for reflecting on *why* particular individuals or groups are being pursued to begin with, how AI/ML and co-design can help, and how it can’t.

Myth #4: Co-design is an inherently ethical approach to design

The central point advanced with Myth #4 is that **co-design is only as ethical as the consequences of the artifacts and socio-technical systems it produces**. These consequences are often further away in time than designers tend to look, which raises the importance of accountability for consequences of design beyond its immediate effects, to those that are geographically and temporally more distant. If the inherently contingent and social act of co-design fails to achieve ethical outcomes beyond the design process, can we really call it co-design? And who benefits most when we do?

Myth #5: All problems can be (co-)design problems

The central point advanced with Myth #5 is that **co-design is not always the best practice to mobilize to address a problem**. Furthermore, when design is chosen as a strategy to solve an identified problem, its limitations must be acknowledged. Such a design humility might ask questions like: when does co-design substitute other expressions of public interest and action?; what are the epistemic limits of design research as it is currently practiced?; who is sidelined by professional design practice and what might we learn from them? and perhaps most importantly, when shouldn’t we design?

A New Era for Co-design?

Based on these challenges, in our full paper, we suggest that a new era for co-design requires theories and methods that are context-appropriate, amenable to black-box nature of AI/ML technologies, and anticipatory in scope.