

# Abstract

## Towards human rights 2.0?

A meta-theoretical analysis of the disruptions to human rights foundations by artificial Intelligence

The use of artificial intelligence ('AI') is ubiquitous in society today. It has contributed towards advancements in diverse fields such as healthcare, transportation, public administration and in helping humanity to solve pressing challenges such as climate change. As AI increasingly plays a larger role within our daily lives and in society, it is inevitable that human rights issues will arise in relation to its use. However, scholarship on AI and human rights typically focus on the impact of AI on discrete enumerated rights. These are first-order concerns that engage and pertain to the content of the existing human rights, such as nondiscrimination, right to private life, freedom of assembly, freedom of expression and other rights found within human rights instruments.

Yet, this thesis identifies that the harms posed by AI systems and its impact upon discrete rights constitute a necessary but insufficient way to account for how AI challenges human rights. AI affordances introduce novel forms of harms and the grasp of existing human rights vernacular of such harms remain slippery as new technologically mediated realities call into question how human rights are being violated, what counts as human rights violations and who are violating human rights, posing second-order challenges towards the fitness of the human rights framework. This thesis takes a meta-theoretical approach to examine the ways in which human rights foundations are being challenged from three aspects: namely at the level of the conceptual, contextual and normative foundations of the international human rights law framework.

At the conceptual level, the individualist, state-oriented and discrete legal rights orientation of human rights are revealed to be conceptually brittle in accounting for risks and harms from AI systems that go beyond these frameworks. Its contextual foundations are in turn undermined through implicit social and material conditions that informed the contours of discrete human rights afforded in the first place. In turn, the human rights normative foundation of human dignity, while itself a flexible concept that has been expanded through case law and treaty interpretation, is nonetheless being challenged in novel ways by AI systems. The decentering of the human being and the disruptions to the conditions of possibilities for the exercise of autonomy are identified as being insufficiently theorized within the literature on human rights and AI. The thesis finds that the collective challenge towards these foundations undermines human rights as an effective mechanism to address challenges by AI systems that impact upon individuals, communities and societies.

By excavating these systemic foundational challenges through a ground-clearing and problem-finding approach, the research aims lay the key steps for a human rights framework that is fit(ter) for the age of AI. This requires the engagement of three key steps: a reframing of the problem space, a reorientation of the nature of the challenge by emerging technologies such as AI into one that takes seriously its material affordances and a re-theorization of the fundamental concept of human dignity informed by the lens of human vulnerability.