
From Workers to Oracles: Tracing Human-Machine Assemblages in the Production of AI

Iraklis Vogiatzis*^{1,2}

¹Hellenic Open University [Patras] (HOP) – Greece

²DiPlab research team – DiPlab – France

Abstract

Artificial intelligence (AI) systems are often portrayed as autonomous; yet, their development hinges on vast, invisible human labor ecosystems. This research investigates a paradox: while human labor is presented as redundant or unskilled, some advances in Machine Learning require new methods of integrating human activities into AI systems. I examine a set of techniques and frameworks from various disciplines of computer science, such as human-in-the-loop (HITL) and Active Learning, questioning how microtasking platforms simultaneously rely on and marginalize human workers. Drawing on case studies from annotation firms (i.e., Pareto.AI, CloudFactory), this study reveals how the AI industry exploits geopolitical inequities and precarious labor to sustain its "automated" systems, while obscuring these workers' contributions.

Under these systems, which integrate human judgment into AI training loops, expose the irreplaceability of human labor in tasks such as data annotation, ambiguity resolution, and model validation. Active Learning (AL) strategies-where workers act as "oracles," labeling uncertain data points-underscore this dependency. However, corporate narratives increasingly emphasize "AutoML" tools that promise human-free workflows, despite evidence that humans remain indispensable for refining AI accuracy, particularly in high-stakes domains like medical imaging and autonomous vehicles.

Addressing the question of how AI technology production is organized requires a focus on the technical details of how human labor is imported, organized, and assessed within these systems. This paper examines the organization of invisible human labor behind machine learning, specifically emphasizing microtasking techniques. These techniques, utilized by software engineers, decompose complex tasks into smaller units that platform workers can complete. The process of segmentation entails an automated approach to managing, evaluating, and compensating human labor.

This paper aims to present the different ways that human labor is 'deeply embedded' (Tubaro 2021) in AI systems and the technical means through which this trend is emerging in several subfields of computer science. Secondly, by following the relevant strategies, including those that promise the automation of the data annotation process, this paper seeks to demonstrate that recent developments over the last couple of years, along with the proliferation of generative AI technologies, have led to an ever-growing demand for human labor in the core logic behind the creation of these systems.

Tubaro, P. (2021). Disembedded or deeply embedded? A multi-level network analysis of online labor platforms. *Sociology*, 55(5), 927-944.

*Speaker

Keywords: Active Learning, Microwork, Microtasking