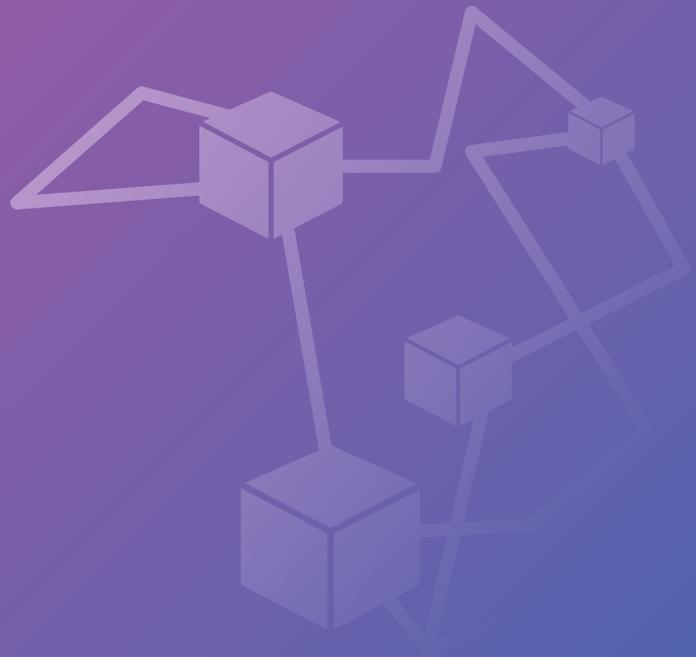


Responsible Sourcing of Data Enrichment Services

Executive Summary



Executive Summary

As AI becomes increasingly pervasive, there has been growing and warranted concern over the effects of this technology on society. To fully understand these effects, however, one must closely examine the AI development process itself, which impacts the world both directly and through the models it creates. This white paper addresses an often overlooked aspect of the development process and what AI practitioners can do to help improve it: the working conditions of data enrichment professionals, without whom the value being generated by AI would be impossible.

High-precision AI models are dependent on clean and labeled datasets. While obtaining and enriching data so it can be used to train models is sometimes perceived as a simple means to an end, this process is highly labor-intensive and often requires data enrichment workers to review, classify, and otherwise manage massive amounts of data. Given that this process of labeling and enriching data inherently embeds human judgement and lived experiences into data, AI's intelligence is highly dependent on human intelligence. Despite the foundational role played by these data enrichment professionals, a growing body of research reveals the precarious working conditions these essential, but largely unseen, workers face.¹ There is, however, an opportunity to make a difference. The decisions AI developers make while procuring enriched data have a meaningful impact on the working conditions of data enrichment professionals. This paper focuses on how these decisions during the procurement process impact workers and proposes avenues for AI developers to meaningfully improve these working conditions.

This paper draws upon existing literature on the experience of data enrichment professionals and insights gathered from AI developers and key stakeholders through conversations and a series of workshops. Acknowledging the existing complexity and lack of standards around how to build equitable data supply chains, we aim to critically evaluate the impact of the industry's current practices on workers, explore practices the industry can adopt to improve worker well-being, and advance the discourse around the future of data enrichment work and the indispensable role it plays in AI development. While more work and research is needed, we have outlined key worker-oriented considerations that practitioners can use as a starting point to raise conversations with internal teams and vendors. Specifically, this paper covers worker-centric considerations for AI companies making decisions in: selecting data enrichment providers, running pilots, designing data enrichment tasks and writing instructions, assigning tasks, defining payment terms and pricing, establishing a communication cadence with workers, conducting quality assurance, and offboarding workers from a project.

Our intention with this paper is to aid the industry in accounting for well-being when making decisions about data enrichment and to set the stage for further conversations within and across AI organizations. Recognizing the critical role that data enrichment professionals play in building AI is imperative, both for ensuring that their work is fairly recognized and compensated and for understanding that the resulting models are a product of human intelligence. We hope this paper serves as a step forward, bringing us closer to a world where data enrichment professionals are recognized and rewarded by the industry for their central role in enabling AI advancement.

1 Gray, M.L and Suri, S. 2019. Ghost Work: How to Stop Silicon Valley From Building a New Global Underclass. Houghton Mifflin Harcourt. Google-Books-ID: 8AmXDwAAQBAJ