

Long hours and low wages: The human labor powering AI's development

November 16 2023, by Ben Lee Taylor



Credit: Pixabay/CC0 Public Domain

The Finnish tech firm Metroc [recently began using prison labor to train a large language model to improve artificial intelligence \(AI\) technology.](#) For 1.54 euros an hour prisoners answer simple questions about snippets

of text in a process known as data labeling.

Data labeling is often outsourced to labor markets in the Global South where companies can find workers who are fluent in English and willing to work for low wages.

Due to the lack of Finnish speakers in these countries, however, Metroc has tapped into a local source of cheap labor. Were it not for the prison labor program, Metroc would likely be hard-pressed to find Finns willing to take data-labeling jobs that pay a fraction of [the average salary in Finland](#).

These cost-cutting strategies not only highlight the significant amount of human labor still required to fine tune AI, but they also raise important questions about the long-term sustainability of such business models and practices.

AI's labor problem

The ethical ambiguity of prison labor-sourced AI is part of a larger story about the [human cost behind AI's significant growth](#) in recent years. One issue that has become more evident over the past year revolves around the [question of labor](#).

Leading AI firms are not denying their use of outsourced and low-wage labor to do work like [data labeling](#). However, the hype around tools like OpenAI's ChatGPT has drawn attention away from [this aspect of the technology's development](#).

As researchers, including myself, are trying to understand the perceptions and use of AI in [higher education](#), the ethical problems associated with current AI models continue to pile up. These include the biases that AI is prone to reproducing, the [environmental impact of AI](#)

[data centers](#), and [privacy and security concerns](#).

Current practices of outsourcing data labeling work expose [an uneven global distribution of AI's costs and benefits](#), with few proposed solutions.

The implications of this situation are twofold.

First, the massive amount of human labor that is still required to shape the "intelligence" of AI tools should give users pause when evaluating the outputs of these tools.

Second, until AI firms take serious steps to address their exploitative labor practices, users and institutions may want to reconsider the so-called values or benefits of AI tools.

What is data labeling?

The "intelligence" component of AI still requires significant human input to develop its data processing capabilities. Popular chatbots like ChatGPT are pre-trained (hence, the PT in GPT). A critical phase in the pre-training process consists of [supervised learning](#).

During supervised learning, AI models learn how to generate outputs from [data sets](#) that are labeled by humans. Data labelers, like the Finnish prisoners, perform different tasks. For example, labelers might need to confirm whether an image contains a certain feature or to flag offensive language.

In addition to improving accuracy, data labeling is necessary to improve the "safety" of AI systems. Safety is defined according to the goals and principles of each AI firm. A "safe" model for one company might mean [avoiding the risk of copyright infringement](#). For another, it might entail

minimizing false information or [biased content and stereotypes](#).

For most popular models, safety means that the [model](#) should not generate content based on prejudiced ideologies. This is partly achieved through a properly labeled training data set.

Who are data labelers?

The job of combing through thousands of potentially graphic images and snippets of text has fallen on data labelers largely concentrated in the Global South.

In early 2023, [Time magazine reported on OpenAI's contract](#) with Sama, a data labeling firm based in San Francisco. The report revealed that employees at a Kenyan satellite office were paid as little as US\$1.32 per hour to read text that "appeared to have been pulled from the darkest recesses of the internet."

[Wired also investigated the global economic realities of data labelers](#) in South America and East Asia, some of whom worked more than 18 hours per day to earn less than their country's minimum wage.

The [Washington Post has taken a close look at ScaleAI](#) which employs at least 10,000 workers in the Philippines. The newspaper revealed the San Francisco-based company "paid workers at extremely low rates, routinely delayed or withheld payments and provided few channels for workers to seek recourse."

The data labeling industry and its required workforce [is set to expand drastically in the coming years](#). Consumers who increasingly use AI systems need to know how they are built as well as the harm and inequities being perpetuated.

Transparency needed

From prisoners to gig workers, the potential for exploitation is real for all entwined in big AI's thirst for data to fuel [bigger \(and possibly more unpredictable\) models](#).

As institutions and individuals are swept up by the momentum of AI and all of its promises, the public tends to pay less attention to ethical aspects of the technology's development.

Researchers at Stanford University recently launched a [website showcasing their Foundation Model Transparency Index](#). The index provides metrics on measures of transparency for the most widely used AI models. These metrics range from how transparent companies are about where they source their data to how clear they are on the potential risks of their models.

Ten AI models were examined based on criteria of how transparent the company that operates them is about its labor practices. The index shows that tech companies have much work to do to improve transparency.

AI is becoming a growing part of our increasingly digital lives. That is why we must remain critical of a set of technologies that, unchecked and unexamined, may cause more problems than they solve and deepen divides in the world rather than eliminate them.

This article is republished from [The Conversation](#) under a Creative Commons license. Read the [original article](#).

Provided by The Conversation

Citation: Long hours and low wages: The human labor powering AI's development (2023,

November 16) retrieved 9 May 2024 from <https://techxplore.com/news/2023-11-hours-wages-human-labor-powering.html>

This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.