



Data Enrichment Sourcing Guidelines

The choices made by AI practitioners have an important influence on the working conditions of data enrichment professionals who perform tasks such as data annotation, data cleaning, and human review of algorithmic outputs.

This document lists five key, worker-centric guidelines that AI practitioners should follow when setting up a project involving data enrichment. They are based on the input from a multi-stakeholder community of data enrichment companies, labor organizers, AI practitioners, and leading researchers in the space.

To learn more about how this resource was developed, visit PAI's [responsible data enrichment sourcing library](#).

1

Ensure all data enrichment workers are paid above the local living wage

AI Practitioners must pay workers/participants at least the living wage for their location. Therefore, please take into consideration the following when setting payment terms: the worker's location, estimated time needed to complete a task and other associated activities, payment structure (e.g. per task, per hour, etc), and the difficulty of the task.

For more information on Global Living Wage, please see [Annex 1 of PAI's white paper](#). These are calculated at a country-level basis; if more exact locations are provided (e.g. State/City), please use the corresponding living wage for that location, as these may differ to the national standard.

If the data enrichment project pays per task, use the pilots (see Guideline 2) to establish a baseline estimate of how long it takes to complete a task, including time spent reading instructions, going through any training and reviewing work before submission. The per task rate should be based on the hourly living wage, divided by the adjusted number of tasks that can be completed in an hour. Furthermore, researchers should track completion times throughout the data enrichment project and adjust targets, and compensation, accordingly.

AI Practitioners should pay for any work completed by workers to compensate for their time, being mindful that mass rejections of work completed without opportunities for redress impact the workers' livelihood. The only exceptions should be in the cases of obvious abuse or fraud.

2

Design and run a pilot before launching a data enrichment project

AI Practitioners should always run a pilot before launching a new or substantially modified data enrichment project; this helps establish reasonable baselines for timeframes and payments. Pilots are smaller versions of data enrichment projects done before the main project with the goal of testing the project design so AI practitioners can make adjustments before the full data enrichment project is done under the same conditions. Pilots should also be used to test the clarity of the instructions and gather feedback on worker experience, looking at task design and tool usability. Testing should occur with a representative group of workers/participants to ensure that researchers get feedback from the types of people who will be completing the task later.

Note: Pilots can recruit a distinct set of workers compared with the main data enrichment project. However, using the same filters as you would for the actual task is helpful for recruiting similar workers for both.



3

Identify appropriate data enrichment workers for the desired task

Matching data enrichment tasks to the skill set, expertise, and/or demographic category of workers can ensure your data enrichment project keeps to time, whilst also protecting workers against the economic impacts of wasted time or rejected tasks. Depending on the data enrichment project, AI practitioners may need to identify a demographically representative set of workers or identify workers with the relevant demographic background necessary to complete the task (e.g. cultural background, age range, location). AI Practitioners should be mindful when designing eligibility criteria, assuring the requirements for successful task completion are clear and only relevant to the task (e.g. language proficiency or domain knowledge) and not onerous, overly limiting or potentially identifying.

For more complex tasks, or those that require a certain level of domain knowledge, it may be beneficial to maintain a consistent workforce across the project lifetime; in this case, both data quality and worker satisfaction can be improved by investing in a workforce. In other instances a consistent workforce may not be appropriate; for example if data collection would benefit from diversity or broader coverage of individuals.

4

Provide verified instructions and/or training materials for data enrichment workers

Create clear training materials for data enrichment tasks, taking into account the existing and required domain knowledge of workers, as well as the tools or platforms being used. Instructions should undergo review by a representative group of workers during an official pilot, before becoming verified. Instructions should typically include examples of correctly and incorrectly completed tasks and, for more difficult studies, allow workers a few practice tasks before launching a new data enrichment project. Please refer to this checklist to help guide you in what to include as a part of your instructions and training materials.

Note: As discussed in 1, time spent reading instructions or in training should be compensated.

5

Establish clear and regular communication mechanisms

Clear communication is critical to ensuring workers have the necessary information to effectively and efficiently complete tasks. AI Practitioners should clarify expectations around communication cadence with workers or managers before starting the project and ensure that workers know who and how to raise questions or concerns with, including contesting any rejected work.

Irrespective of whether your data enrichment project involves one off interactions with workers or reengagement with the same pool over time, AI Practitioners must ensure that workers are provided with clear mechanisms for asking questions, reporting concerns or technical issues that may arise during the data enrichment project. Moreover, AI Practitioners should collect feedback from workers on all aspects of the project and adjust the data enrichment project accordingly.