PARTNERSHIP ON AI

CASE STUDY

Implementing Responsible Data Enrichment Practices at an AI Developer

The Example of DeepMind

Sonam Jindal

NOVEMBER 16, 2022

Contents

Executive Summary

As demand for AI services grows, so, too, does the need for the enriched data used to train and validate machine learning (ML) models. While these datasets can only be prepared by humans, the data enrichment workers* who do so (performing tasks like data annotation, data cleaning, and human review of algorithmic outputs) are an often-overlooked part of the development lifecycle, frequently working in poor conditions continents away from AI-developing companies and their customers.

* For the purposes of this white paper we refer to individuals completing data enrichment as "**workers**." In doing so, we recognize the variety of employment statuses that can exist in the data enrichment industry, including independent contractors on self-service crowdsourcing platforms, subcontractors of data enrichment providers, and full-time employees.

Last year, the Partnership on AI (PAI) published "Responsible Sourcing of Data Enrichment Services," a white paper exploring how the choices made by AI practitioners could improve the working conditions of these data enrichment professionals. This case study documents an effort to put that paper's recommendations into practice at one AI developer: DeepMind, a PAI Partner.

In addition to creating guidance for responsible AI development and deployment, PAI's Theory of Change includes collaborating with Partners and others to implement our recommendations in practice. From these collaborations, PAI collects findings which help us further develop our curriculum of responsible AI resources. This case study serves as one such resource, offering a detailed account of DeepMind's process and learnings for other organizations interested in improving their data enrichment sourcing practices.

After assessing DeepMind's existing practices and identifying what was needed to consistently source enriched data* responsibly, PAI and DeepMind worked together to prototype the necessary policies and resources. The Responsible Data Enrichment Implementation Team (which consisted of PAI and members of DeepMind's Responsible Development and Innovation team, which we will refer to as "the implementation team" in this case study) then collected multiple rounds of feedback, testing the following outputs and changes with smaller teams before they were rolled out organization-wide:

* **Sourcing data enrichment work** is a process that requires a number of steps including, but not limited to, defining the enrichment goal, choosing the enrichment provider, defining the enrichment tools, defining the technical requirements, writing instructions, ensuring that instructions make sense, setting worker hours, determining time spent on a particular task, communicating with enrichment workers, rejecting or accepting work, defining a project budget, determining workers' payment, checking work quality, and providing performance feedback.

- A two-page document offering fundamental guidelines for responsible data enrichment sourcing
- An updated ethics review process
- A checklist detailing what constitutes "good instructions" for data enrichment workers
- A table to easily compare the salient features of various data enrichment platforms and vendors
- A spreadsheet listing the living wages in areas where data enrichment workers commonly live

Versions of these resources have been added to PAI's responsible data enrichment sourcing library and are now available for any organization that wishes to improve its data enrichment sourcing practices.

Ultimately, DeepMind's multidisciplinary teams developing AI research, including applied AI researchers (or "researchers" for the purposes of this case study, though this term might be defined differently elsewhere) said that these new processes felt efficient and helped them think more deeply about the impact of their work on data enrichment* workers. They also expressed gratitude for centralized guidance that had been developed through a rigorous process, removing the burden for them to individually figure out how to set up data enrichment projects.

DeepMind's multidisciplinary teams developing AI research said that these new processes felt efficient and helped them think more deeply about the impact of their work on data enrichment workers.

While organizations hoping to adopt these resources may want to similarly engage with their teams to make sure their unique use cases are accounted for, we hope these tested resources will provide a better starting point to incorporate responsible data enrichment practices into their own workflows. Furthermore, to identify where the implemented changes fall short of ideal, we plan to continue developing this work through engagement and convenings. To stay informed, sign up for updates on PAI's Responsible Sourcing Across the Data Supply Line Workstream page.

This case study details the process by which DeepMind adopted responsible data enrichment sourcing recommendations as organization-wide practice, how challenges that arose during this process were addressed, and the impact on the organization of adopting these recommendations. By sharing this account of how DeepMind did it and why they chose to invest time to do so, we intend to inspire other organizations developing AI to undertake similar efforts. It is our hope that this case study and these resources will empower champions within AI organizations to create positive change.

* **Data enrichment** is curation of data for the purposes of machine learning model development that requires human judgment and intelligence. This can include data preparation, cleaning, labeling, and human review of algorithmic outputs, sometimes performed in real time.

Examples of data enrichment work:

Data preparation, annotation, cleaning, and validation:

- *Intent recognition*
- *Sentiment tagging*
- *Image labeling*

Human review (sometimes referred to as "human in the loop"):

- *Content moderation*
- *Validating low confidence algorithmic predictions*
- *Speech-to-text error correction*

SECTION 1

Background

Importance of Data Enrichment Workers and Pathways to Improve Working Conditions

Data enrichment workers are central to AI development and deserve fair working conditions. Increasing transparency around the AI industry's data enrichment practices and developing guidance for how to adopt more responsible practices has the potential to improve these conditions.

Though AI development relies on large, enriched datasets, there is still limited importance placed on how those datasets are constructed.[1] As AI continues to be deployed in increasingly sensitive contexts, increasing transparency around how the underlying datasets are created is a critical step toward closing the accountability gap in the AI industry.[2 3] Examining the conditions under which the datasets that enable AI models are created is important not only to ensure the efficacy and fidelity of the models, but to ensure that the data enrichment workers who make these datasets legible to algorithmic models are treated well.

Particularly concerning are the precarious conditions data enrichment workers continue to face, despite their critical role in building the datasets that enable AI models. As the demand for this labor continues to grow, it is important to acknowledge that these workers lack formal protections and there is limited guidance for how AI companies should be interacting with these workers. The relative novelty of the demand for this type of labor poses unique challenges for companies seeking to institute ethical, worker-oriented practices.

Changing how the field treats data enrichment workers requires shifting the industry's approach to data enrichment from an ad-hoc process to one that recognizes this labor as central to AI development.

While there has been more research and coverage on the conditions facing data enrichment workers in recent years,[4] there is still limited transparency from various organizations in the ecosystem on how they approach data enrichment. This is partly because there are not yet field-wide standards on data enrichment practices in general[5] — let alone on the treatment of data enrichment workers. Changing how the field treats data enrichment workers requires shifting the industry's approach to data enrichment from an ad-hoc process to one that recognizes this labor as central to AI development.

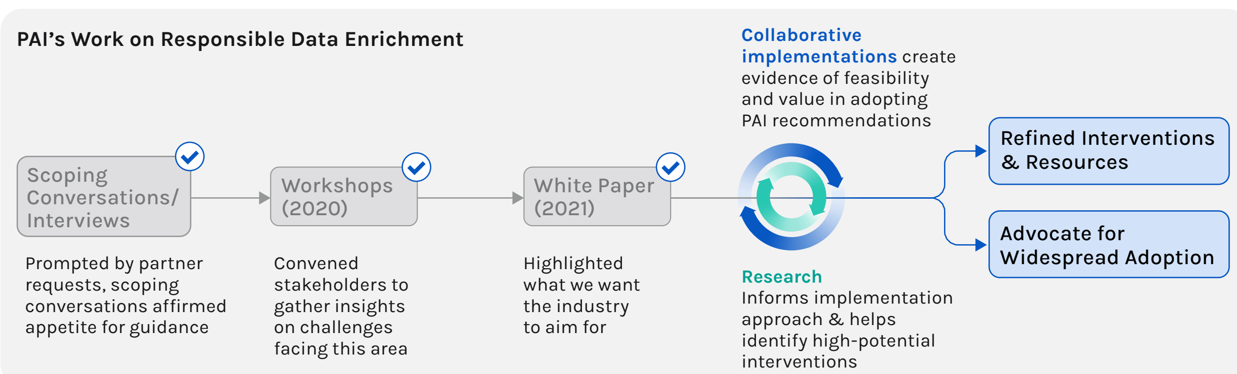
While shifting how the broader field approaches data enrichment is not a trivial task, increasing transparency regarding current practices and developing more practical guidance can move the field towards improved conditions for data enrichment workers. Greater transparency can help emphasize the central role of data enrichment workers, create the basis for a rich public dialogue of how to improve conditions for workers, and increase confidence in AI models themselves. Disseminating tested and practical guidance will help lower the barriers for AI practitioners to navigate how to adopt more ethical data enrichment practices as standard practice.

Case Study as a Method of Increasing Transparency and Sharing Actionable Guidance

This case study details how one AI company adopted responsible data enrichment sourcing practices, the challenges they faced, and how they were addressed. The accompanying resources aim to make it easier for others to adopt these practices.

Following a multistakeholder workshop series that brought together experts across industry, civil society, and academia, PAI published a white paper covering how AI practitioners' choices around data enrichment sourcing impacts workers. The paper proposed avenues for AI developers to meaningfully improve these working conditions and helped outline the goals for what the industry should strive towards. In an effort to further lower the barriers for companies to adopt these recommendations, we needed to demonstrate the feasibility of these recommendations and develop resources that would allow companies to effectively and realistically introduce these recommendations into their workflows. Motivated by this need for actionable guidance, PAI is collaborating with AI companies to implement the white paper's recommendations and document the process of implementing them.

By sharing a detailed account of how one company institutionalized standard approaches to working with data enrichment workers and a refined set of resources meant to make adoption of these recommendations more feasible, we hope to help other AI companies feel better equipped to incorporate responsible data enrichment practices into their workflows.



With this transparent account we also hope to prompt a dialogue on what else companies should be doing to improve working conditions for data enrichment workers, where the limits of their influence lie, and where additional action is needed from policymakers, labor unions or data enrichment providers.

Background on DeepMind's Motivations

As more of DeepMind's work began to involve data enrichment workers, the company recognized a need for more specific guidance and processes to uphold its commitment to building AI responsibly.

The first company we collaborated with was PAI Partner DeepMind, a British AI company with more than 1,000 employees which was founded in 2010. DeepMind is a research laboratory with an interdisciplinary team ranging from scientists and designers to engineers and ethicists. DeepMind has a Responsible Development and Innovation team with a mandate of ensuring DeepMind is upholding its commitments to responsibility by assessing the implications of their research on society, as expressed in their Operating Principles.

Given their foundation as a research-focused organization, DeepMind had an existing process to manage experiments involving human subjects or participants.* This included engagement with the company's externally chaired Human Behavioral Research Ethics Committee, which follows IRB (Institutional Review Board) protocols. This committee is tasked with reviewing projects involving research participants, meaning that their behavior is studied as part of a research project. Importantly, to fall under the jurisdiction of an IRB,* research participants have to be able to freely opt in and out of studies, and payment or employment status cannot be conditional on successful completion of a task.

* A **human subject/participant** is a living individual about whom an investigator conducting research obtains data or information. (*Adapted from the US Code of Federal Regulation.)*

* As per the US Office of Human Research Protections, an **IRB** is "a specially constituted review body established or designated by an entity to protect the welfare of human subjects recruited to participate in biomedical or behavioral research."

However, as more of DeepMind's work began involving data enrichment workers, who receive payment for tasks performed, challenges arose with assessing these projects against IRB protocols. It became clear to the Responsible Development and Innovation team that both researchers and reviewers required more specific guidance and a dedicated review process to properly address the unique ethical challenges related to data enrichment projects. Without centralized guidance on how to set up data enrichment projects, researchers would need to invest time to independently seek out best practices and use their better judgment. A centralized set of data enrichment guidelines that incorporated considerations unique to interacting with data enrichment workers and a dedicated review process would ultimately save researchers' time. DeepMind hypothesized that this would lead not only to a more rigorous and efficient review process but also an increase in data quality.

Additionally, DeepMind wanted a rigorous set of practices and processes to adhere to when constructing and publishing enriched datasets. In the absence of an existing set of standard practices (like IRB protocols), they turned to PAI's white paper on responsible sourcing of data enrichment. Given PAI and DeepMind's shared goal of improving conditions for data enrichment workers, we agreed to work together to build a set of resources that would help both DeepMind and the broader AI community to implement responsible sourcing practices.

SECTION 2

Process and Outcomes of the DeepMind and PAI Collaboration

To identify the resources and artifacts that would enable DeepMind employees to consistently adhere to responsible data enrichment practices, the implementation team (made up of PAI and members of DeepMind's Responsible Development and Innovation team) began our collaboration by examining DeepMind's existing approach and processes. Next, the implementation team developed prototypes of those resources and got feedback on them from key internal stakeholders (including researchers, engineers, program managers, lawyers, and security and privacy experts) who would either be using the resources directly or helping ensure that teams followed best practices during the data enrichment review process. The implementation team then created the internal resources, testing them with two research teams to gather further feedback and make user-informed adjustments before finalization. Once finalized, the new review process for data enrichment sourcing projects and accompanying resources were rolled out to the rest of the organization.

Changes and Resources Introduced to Support Adoption of Recommendations

The following resources and changes were introduced to help teams source enriched data responsibly: a two-page guidelines document, an updated ethics review process, a "good instructions" checklist, a vendor comparison table, and a living wages spreadsheet.

After initial conversations with the Responsible Development and Innovation team, we intended our collaboration to result in two primary outputs: a brief document summarizing the recommendations from PAI's white paper and an adapted ethics review process for projects involving data enrichment workers. However, the feedback collected by the implementation team from stakeholders across the organization helped identify additional resources that would help teams more effectively and consistently adopt the recommendations. (This approach to analyzing organizational needs and collecting feedback is described in Appendix A.) Below is the full list of outputs and changes resulting from this collaboration.

Two-Page Guidelines Document

The two-page guidelines document serves as the primary reference document for all teams seeking to set up a data enrichment project and for the Human Data Review Group to know what standards to apply to any proposed data enrichment project. Based on PAI's white paper on responsible sourcing practices, the document covers five primary guidelines and links to additional documents that might further assist teams in meeting these five guidelines. "Data Enrichment Sourcing Guidelines," a copy of the guidelines document (with DeepMind-specific provisions removed), is available for input and use by the broader AI community on PAI's responsible data enrichment sourcing library.

In summary, the guidelines are:

1. Select an appropriate payment model and ensure all workers are paid above the local living wage.
2. Design and run a pilot* before launching a data enrichment project.
3. Identify appropriate workers for the desired task.
4. Provide verified instructions and/or training materials for workers.
5. Establish clear and regular communication mechanisms with workers.

In addition to these guidelines (which are explained in the Guidelines Document in more detail), it is important to mention two notable policies the company has put in place for research teams procuring enriched data. First, DeepMind has enacted a policy to prohibit mass rejections (the rejection of a large number of tasks simultaneously, often resulting in wages being withheld and workers' ratings being lowered on platforms) without reason and always paying workers for their time unless there is clear evidence of fraud. Second, DeepMind will only use vendors in regions where workers will be paid in cash (as opposed to gift cards or vouchers).

DOWNLOAD

Data Enrichment Sourcing Guidelines Document

* In this context, **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. During pilots, AI practitioners can test the clarity of the task instructions, incorporate feedback from a group of people representative of who will be performing the data enrichment, establish a baseline of how long different elements of the task may take to set realistic expectations and appropriate payment rates, and more.

Adapted Review Process

DeepMind had an existing review process which any project involving human data was subject to. To ensure that study reviews remained streamlined, DeepMind adapted the existing process to include a triage stage, which determines whether projects involve research participants or data enrichment workers. All studies are required to fill out the same application forms and are assessed against the best practices guidelines, with the Responsible Development and Innovation team on hand to guide researchers through the process. Data enrichment studies are reviewed by representatives from the following DeepMind teams: Responsible Development and Innovation, Ethics Research, Legal, Security, and Data Solutions, whilst research participant studies are flagged for IRB review.

DeepMind's Adapted Review Process

If anything on the application diverges from what is expected based on the guidelines, the team that filled out the application must provide a strong justification for requesting an exception or will be asked to make the necessary adjustments to make sure they meet the guidelines. That being said, no exceptions can be made for paying a living wage to workers and providing them with a clear recourse mechanism to get in touch with researchers in case they have questions, concerns, or technical issues.

Good Instructions Checklist

DOWNLOAD

Good Instructions Checklist for Data Enrichment Projects

The good instructions checklist is a resource detailing what should be included in a set of task instructions to make sure they are as clear as possible for workers. While this was not originally scoped as a part of the resources the implementation team would be creating, this emerged as a need during the interviews with researchers as they continued to ask what would make instructions "clear" and "good" enough for data enrichment workers to use. The first iteration of this document had a single checklist where the implementation team asked researchers to always include all criteria. However, after receiving feedback from the teams, we separated the various items into the following categories: "should always include," "for applicable studies, include," and "depending on the task, you may also need."

Given the range of studies across the company, there were studies where some checklist items would not have been desirable. For instance, in many cases, including examples of common mistakes would provide workers with a concrete set of examples to avoid and thus minimize the risk of having work rejected. However, for some studies, teams might want to capture the instinct of workers and there may not be any "mistakes" for the given task. Cases like these underscore two broader challenges in the field. First, it is challenging to create guidance that is applicable across varied use cases and specific enough to actually help guide researchers. Second, cases like this highlight the blurry line (and in some cases, overlap) between workers who are building training data for AI models and human subjects whose ways of thinking about the world are being captured in training data, both of whom are contributing to building AI models. While we tried to balance between building useful tools while making the guidelines general enough to apply to different use cases organization-wide, these challenges should be kept in mind as the field attempts to develop additional guidance. "Good Instructions Checklist for Data Enrichment Projects," PAI's version of this checklist, can be found on PAI's responsible data enrichment sourcing library.

Vendor and Platform Feature Comparison Table

DOWNLOAD

Data Enrichment Vendor and Platform Feature Comparison Table

This table lists different considerations for the various vendors that have been onboarded and approved by DeepMind. This table allows researchers to see any worker-oriented guidance and recommendations that are platform-specific. In speaking to researchers, it became clear that in addition to helping keep the primary guidelines document to the most important, general principles, a vendor comparison table that incorporated additional worker-centric considerations specific to a particular vendor would make it easier for researchers to find guidance that was only relevant for a given platform and help them choose the right platform for their needs. The "Data Enrichment Vendor Comparison Template," a template for creating a similar vendor comparison table, is available as part of PAI's responsible data enrichment sourcing library.

Living Wages Spreadsheet

DOWNLOAD

Local Living Wages Template

This spreadsheet is a centralized resource with a list of living wages for the locales most commonly used by DeepMind researchers. The living wages listed in this spreadsheet are expected to be the minimum set by DeepMind researchers. This dynamic document will be updated as living wages change and as researchers request additional locales to be added by the Responsible Development and Innovation team. The main columns in this spreadsheet are: country, city, living wage in the workers' local currency, living wage in the researchers' currency, and source. For additional guidance on sources for living wages, see Annex 1 of PAI's "Responsible Sourcing of Data Enrichment Services" white paper. The "Local Living Wages Template," a template for creating a similar living wages spreadsheet, is available as part of PAI's responsible data enrichment sourcing library.

Addressing Practical Complexities That Arose While Finalizing Changes

Putting PAI's recommendations to the test in an applied setting allowed us to understand the complexities of adopting these guidelines in practice. Challenges to implementing PAI's recommendations in practice included: shaping the guidelines to be usable, making the guidelines apply to diverse use cases, and setting up a payment methodology in a complicated environment.

Existing research and the multistakeholder workshop series hosted by PAI helped shape the initial recommendations for how industry can improve conditions for data enrichment workers. During this applied collaboration with an industry partner, we had the chance to see what it would take to implement those recommendations, better understand their feasibility, and identify barriers to implementing some of those recommendations. In sharing how the implementation team worked through some of the complexities that arose while implementing the recommendations, we hope to make it easier for other companies to similarly incorporate these recommendations into their data enrichment practices.

Guideline Usability

One of the primary goals of this collaboration was to make it feasible to consistently uphold responsible data enrichment practices. Critical to that goal was making sure that any proposed changes would be usable by the researchers that would be setting up data enrichment projects. The feedback collected throughout this collaboration helped the implementation team create resources that were concise enough to be usable in practice and detailed enough to serve as meaningful guidance. As a result, the primary guidelines document is relatively short at two pages to maximize readability and ease of navigation. At the same time, the document includes hyperlinks to other resources that can provide researchers with more context on how best to uphold the guidelines.

During the feedback process, researchers expressed the need for additional guidance (on topics which fell outside the scope of worker-oriented considerations) on data enrichment projects. PAI and our DeepMind collaborators weighed whether we should introduce more comprehensive guidance incorporating these additional considerations. Ultimately, the implementation team decided it was important to first align on a standardized set of responsible data enrichment guidelines as initially planned, building a shared understanding across the organization of ethical standards that needed to be met. (Since then, DeepMind has independently pursued an effort to build more comprehensive guidance.) In addition to serving as a guide for researchers, this document was also intended to serve as the primary resource for the Human Data Review Group to consult when assessing whether teams' projects met the necessary standards.

Addressing Diverse Use Cases

Our intention was to design artifacts and processes in a way that would make it easy for researchers to consistently incorporate responsible sourcing practices into their data enrichment projects. The high level of variance in DeepMind's data enrichment use cases made it difficult to phrase the guidance in a way that would make them applicable across all use cases. This was partially addressed by differentiating between policies that need to be followed for all studies and those that may depend on the use case. The goal was to make sure that guidelines and accompanying resources supported DeepMind researchers in making ethical decisions for their projects, even in situations the guidelines don't explicitly cover. The review process supplements researchers' efforts by providing an additional check on their data enrichment projects, as the review committee can provide feedback and ensure that any exceptions to the guidelines are reasonable given the context.

Our collective understanding of how to improve conditions for data enrichment workers will continue to evolve as we learn more.

Additionally, it is important to recognize that our collective understanding of how to improve conditions for data enrichment workers will continue to evolve as we learn more and new data enrichment use cases emerge (both within DeepMind and across the industry generally). At this stage, it is not possible to capture all of the worker-oriented considerations that could apply to these future use cases. This being the case, we tried to share how the guidelines would impact workers in our presentations to the teams and craft the guidelines in a way that would get researchers to think about the impact on workers more broadly when they are designing data enrichment projects.

Establishing a Payment Methodology

DeepMind has formalized a commitment to paying data enrichment workers at least a local living wage based on the best available information. However, the most specific geographic location information provided by some data enrichment platforms is at the country level. This poses a challenge because living wages can vary quite drastically within some countries. Furthermore, some studies require workers from a variety of different locales but platforms don't always provide an easy option to set multiple wages for a single project based on the workers' locations. These two aspects make it difficult to consistently ensure workers are being compensated fairly. While these are important follow-ups to pursue with platforms directly, DeepMind will, in the interim, still be maintaining a policy of paying workers a living wage based on their location. When setting wages, researchers are expected to use the most specific location data they have. If the exact location of the worker is not known, researchers can use the country level average for now. And, as mentioned above, a spreadsheet containing living wages for the most common locations was created to aid researchers.

Assessing Clarity of Guidelines and Rolling Out Changes Organization-Wide

To ensure the usability and fidelity of these resources and changes, we gathered feedback from research teams through group discussions, direct comment periods, interviews, and surveys.

After incorporating the feedback received from stakeholders across the organization, the implementation team wanted to roll out the guidelines, supplemental resources, and review process across DeepMind. Prior to a full roll-out, it was important to have these new artifacts tested by a few research teams to make sure the guidelines were clear and usable.

The implementation team interviewed and surveyed the initial research teams testing the new process at two different points. The first time was after the team had reviewed the guidelines and gone through the application and review process but before they had started their labeling project. The goal was to assess perceived benefits and pitfalls of the guidelines, identify areas of improvement, and evaluate if the guidelines helped change how teams approached the data enrichment project.

The second time the implementation team interviewed research teams was once their study was underway or completed. The goals here were to measure impact of implementing the guidelines, assess if the experience of implementing the guidelines differed from what they had thought when filling out the application form, assess where the guidelines/process fell short from researchers' actual experiences, and see if researchers identified any topics where they might have needed additional guidance.

Once all the outputs had been finalized, DeepMind published all the relevant resources on an internal site accessible to all research teams and rolled this out to the broader organization through a company-wide announcement.

DeepMind will continue to collect feedback from teams going through this process and make necessary changes that will continue to make it easier for researchers to understand and follow the data enrichment guidance.

SECTION 3

Reactions, Impact, and Next Steps

Response from Research and Development Teams

Establishing clear processes with supporting resources empowered researchers and reduced the burden on them to determine best practices on their own.

In an effort to improve conditions for workers, our intention from the start of this collaboration was to build and test resources that could be feasibly adopted by the people setting up data enrichment projects.

From early conversations, the implementation team heard from many researchers that they were actively seeking more guidance on how to set up data enrichment projects. Some of this was motivated by concerns over data quality. Researchers recognized that some tasks could only be completed by human annotators and were eager for guidance on how to set up projects in a way that would result in high-quality data sets. Researchers with experience in setting up data enrichment projects were exploring how to share best practices across the company and appreciated that there was an effort to align on worker-oriented considerations, given that those challenges required buy-in and input from a more diverse set of internal stakeholders. This underscored the desire for guidance, affirmed researchers' appetite to engage on questions related to data enrichment, and presented an opportunity to shape best practices that incorporated worker-oriented considerations from the start.

By standardizing the best practices for data enrichment projects and providing actionable guidance to researchers, DeepMind was able to lessen the burden on individual researchers who previously had to grapple with questions they didn't necessarily have the background to address on their own.

While there was an appetite for guidance, some researchers raised concerns over the cost and additional time needed to implement the guidelines. DeepMind's Responsible Development and Innovation team and leadership concluded that introducing a set of clear guidelines and a dedicated review process for any new data enrichment projects would actually save researchers' time as well as reduce long term costs due to the creation of higher quality data sets. By standardizing the best practices for data enrichment projects and providing actionable guidance to researchers, DeepMind was able to lessen the burden on individual researchers who previously had to grapple with questions they didn't necessarily have the expertise or experience to address on their own.

Early feedback from teams using the new guidelines was generally positive. They were appreciative of having a clear set of best practices to follow, and shared that having an open and collaborative review process provided them with a useful level of reassurance. They highlighted that the process felt efficient and helped prompt them to consider the impact of their project design on the user experience for data enrichment workers. After going through the new process and completing a data enrichment project, one researcher shared that the experience of setting up the data enrichment project made them realize how useful the guidelines and resources on the internal site were, reflecting, "Everything you have on the microsite is exactly what we needed." Through conversation with researchers using these guidelines, it also became clear that researchers recognized and appreciated the attentiveness that was required to work with data enrichment workers in order to build higher-quality datasets. This reaffirmed that designing data enrichment projects and specifically the work that data enrichment workers do are central to building high quality AI models.

The implementation team also received constructive feedback that helped us make a few additional adjustments to the guidance itself and gave us a deeper understanding of the level of guidance that researchers sought. For instance, teams brought up that it would be helpful to have a set of examples to see how other teams had set up similar projects. This is something that the Responsible Development and Innovation team will be collecting from internal teams as more projects go through the process and will be made available to prospective teams (pending consent). To complement this, researchers also thought it would be helpful to have additional workshops or office hours where they could learn best practices and pose questions.

Both wanting more examples from teams who had already set up data enrichment projects and wanting more opportunities to engage directly with people who were knowledgeable about data enrichment demonstrated that researchers appreciated guidance. Acting on this feedback, the Responsible Development and Innovation team will be hosting regular office hours to create opportunities for DeepMind researchers to get acquainted with the changes and serve as an additional resource for teams setting up data enrichment projects.

Key Stakeholders/Leadership Reflections and Motivations

Seeing data enrichment as critical to dataset quality and recognizing the importance of ethical data supply chains, leaders at DeepMind appreciated the standardization of responsible data enrichment practices.

As the need for enriched data and, therefore, data enrichment workers has grown across DeepMind, more individuals across the organization were thinking through how to design data enrichment projects. However, much of this was being thought through in silos. By initiating this effort to centralize an ethics review process, the Responsible Development and Innovation team brought together various stakeholders from across the organization to provide feedback on the proposed ethical data enrichment guidelines and review process. This effort tapped into a large appetite to engage on these topics and helped get people to think more critically about how they were constructing these data enrichment projects. Beyond wanting to know how to design experiments well, there was also a recognition that this needed to be done in an ethical and safe way. This was driven by both an existing company culture of building AI ethically and a desire to have a similar level of guidance and review as they maintained for projects requiring IRB approval.

While facing a few questions from researchers about the potential costs mentioned above, the leadership at DeepMind supported this initiative led by the Responsible Development and Innovation team. Leadership agreed that it was important to fill the gap of not having any explicit ethical guidance for projects requiring data enrichment workers, particularly given the increasing number of projects requiring enriched data. Additionally, as mentioned above, DeepMind's Responsible Development and Innovation team previously concluded that (in the absence of existing guidance) introducing the guidelines would actually save researchers' time and lead to higher quality data. This was underscored in a conversation the implementation team had with one senior researcher. They reflected that rather than having to individually advise research teams trying to set up data enrichment projects, they could now simply point them to the internal site housing standardized resources that had been created through a rigorous process. Additionally, while many of the guidelines are focused on the impact on data enrichment workers, many are also about making sure that researchers and these workers are aligned on how tasks need to be done. In helping researchers design data enrichment projects that lead to greater alignment with workers, researchers are less likely to need to repeat data enrichment projects.

Beyond wanting to know how to design experiments well, there was also a recognition that this needed to be done in an ethical and safe way.

Continued Work for DeepMind

This collaboration acted as an important launch point for DeepMind to further invest in internal infrastructure to scale their data enrichment operations responsibly.

As the industry builds more complex AI models requiring enriched data sets and begins to scale up its reliance on data enrichment workers, data enrichment workflows and the nature of this work will continue to change. As a result, it is important to recognize that our understanding of data enrichment work might evolve and we will need to consistently analyze the impact of changing data supply chains on workers. This DeepMind and PAI collaboration represents DeepMind's starting point to formalize and consciously incorporate worker-oriented considerations into the company's data enrichment practices. Given the lack of regulation or industry-wide standards guiding how these workers need to be treated, this is an important step. However, DeepMind acknowledges that additional work is needed to continuously improve conditions for data enrichment workers. While PAI will explore ecosystem-wide changes that would help workers with a future Data Supply Lines Roadmap, there are also impactful ways for DeepMind to build on the guidelines and review process they have introduced.

First, as new data enrichment use cases emerge, the resources developed during this collaboration should be adapted to make sure they provide additional guidance to researchers as needed. While the guidance was designed to be general enough that it would broadly apply to the various research teams across the organization, new use cases may require the guidelines to be adjusted in the future.

Second, the process of finalizing the data enrichment guidelines for DeepMind revealed a few obstacles that make it difficult to fully adhere to the guidelines in some contexts. Internally, DeepMind will continue to invest in infrastructure and resources that will make it easier to adhere to these guidelines and close some of the gaps identified during this collaboration. Outlining these guidelines provided the necessary impetus for organization-level investment in creating this infrastructure. This may also need to be supplemented by working directly with vendors and platforms to make it easier for researchers to consistently uphold the guidelines.

New use cases may require the guidelines to be adjusted in the future.

For example, one recommendation that we pursued during this engagement was creating regular communication channels between workers and researchers, as well as re-engaging with the same set of workers for similar projects. However, we found that following this recommendation was not as straightforward as anticipated due to limited functionality permitted by some platforms for communicating with workers and privacy restrictions, making it difficult to re-engage with the same workers. Effectively adhering to this guideline would, among other changes, require building tools and infrastructure that allow researchers to easily communicate and re-engage with workers.

This collaboration has helped DeepMind build internal support to invest in infrastructure that would make regular communication with platform workers more feasible. That being said, though platforms are the best fit for some use cases, the use of managed services for situations that require more regular communication or re-engagement with workers can be appropriate.

Limitations of Case Study Applicability

While some AI organizations may have less infrastructure in place to make similar changes, the resources shared in this case study are designed to make it easier for organizations of any size to responsibly create enriched datasets.

Due to DeepMind having a strong research practice and adhering to IRB for studies involving humans, it may have been less of a lift for them to implement a parallel review process for projects involving data enrichment workers. Less research-focused organizations may have less infrastructure in place to support this kind of review. Despite having different organizational structures, more commercially oriented organizations are procuring data and involving data enrichment workers in similar ways and should be thinking about the impact on workers. One of the key motivating factors for DeepMind to invest in this process was that this saved time for researchers and lessened the burden on individual researchers to have to deal with these questions on their own.

The other major limitation is that it is difficult to immediately assess the ultimate impact on workers. In the absence of this information, all of the recommendations are backed by research and multistakeholder input. The intention of this effort was to implement those recommendations in practice to evaluate their feasibility. Additionally, as stated earlier, we recognize that we are still at an early stage in our collective understanding of how to transform data supply chains in the AI industry so that they work better for workers. We hope this will help us have further conversations about the additional work and guidance that needs to exist to improve conditions for these workers. Being able to put theoretical recommendations to the test has helped us identify additional levers of change that we plan on exploring to continue to strive towards improved conditions for workers. At the same time, we hope to get feedback from workers on this effort and PAI's future Data Supply Lines Roadmap.

Conclusion

One of the primary motivations for undertaking this collaboration and sharing these findings was to lower the barriers for companies to adopt responsible sourcing practices outlined in PAI's responsible sourcing white paper. Though there is more work to be done, we believe that putting the recommendations to the test was an important first step towards developing a deeper understanding of how companies can incorporate ethical sourcing practices.

Prior to the changes introduced during this collaboration, research teams at DeepMind would need to think through project design individually and seek out information on their own from the various sources. Introducing a centralized set of guidelines and a review process saves them time and allows them to benefit from the shared learnings of their colleagues. By asking teams to submit an application that documents their approach to setting up data enrichment projects based on a centralized guidance document, the Human Data Review Group and review process served as a starting point to centralize the gathering of best practices and learnings from across the organization so teams could learn from each other.

We are sharing this case study and accompanying resources in the hopes that this will serve as a guide for other AI practitioners to adopt similar types of guidelines and so that we can push the industry towards better practices. While PAI is not positioned to audit AI practitioners, we hope that sharing resources and our documentation of how these resources were developed will provide practitioners with confidence in making similar changes to their data enrichment practices. Recognizing that different organizations may have different resources and constraints, the resources developed over the course of this collaboration are meant to make it easier for organizations with less infrastructure to incorporate these guidelines into their own workflows without having to replicate the rigorous process we have undertaken here.

We are sharing this case study and accompanying resources in the hopes that this will serve as a guide for other AI practitioners to adopt similar types of guidelines.

In addition to helping us identify what AI companies are positioned to do to positively impact worker experience, this has also helped us understand the limits of what individual companies can do to impact worker experience and what action is needed from platforms/vendors and policymakers. These insights will help shape our future work in this area as we continue to push for more ethical data supply chains.

We hope that this level of transparency creates an opportunity to discuss additional avenues to improve conditions for data enrichment workers and helps recenter considerations of labor at the heart of the industry's data enrichment decisions.

Acknowledgements

We would like to thank all of the original workshop and interview participants who provided feedback or otherwise informed the recommendations proposed in PAI's "Responsible Sourcing of Data Enrichment Services" white paper. This case study builds on that white paper by putting its recommendations to the test and developing a set of resources that can be used by AI practitioners to adopt those recommendations. Many PAI staff members contributed to this work. In particular, Katya Klinova helped champion this work, was a part of the implementation team, and provided thoughtful feedback and ideas throughout the process; Hudson Hongo edited and refined the case study's narrative; and Neil Uhl provided visual design for the case study.

We would like to thank DeepMind for their transparency and for graciously allowing us to publicly share the important lessons learned as they undertook this effort to adopt responsible data enrichment practices organization-wide. We are grateful to all of the DeepMinders who shared their time, experiences, and feedback so that we could co-develop a tested set of practical resources for the broader AI field to use. We would like to especially thank Antonia Paterson and Will Hawkins for their tireless leadership in championing responsible AI practices within DeepMind and being gracious, thoughtful, and amazing partners over the course of this collaboration.

APPENDIX A

Initial Discovery Process and Getting Reactions to PAI Responsible Sourcing Recommendations

To understand the types of resources that would realistically help DeepMind's research and development teams implement the recommendations from PAI's white paper, we wanted to understand how the company and individual teams within it were approaching data enrichment and, more specifically, how they were working with data enrichment workers.

We began by having conversations with the Responsible Development and Innovation team, a team with a mandate to ensure research is done responsibly. As the team that oversees the review process for internal teams setting up projects involving human subjects, they were also driving the company's initiative to build a parallel set of guidelines for projects requiring enriched data. Given this team's mandate and knowledge of DeepMind's internal research teams, they were able to share deep insights into how teams currently approached data enrichment and how they may respond to any proposed changes. Being able to incorporate their feedback into the guidelines early on allowed us to make early adjustments and present more refined resources to researchers. Doing this upfront saved us time because it allowed us to get more substantive and targeted feedback from the researchers who would be using the guidelines. During this initial review process, the implementation team re-ordered the guidelines to make them more user-friendly, adjusted the language on some of the guidelines to make it clearer to the target audience, learned more about how the teams were currently approaching data enrichment, identified potential areas where additional guidance may be needed to effectively implement the guidelines, and identified specific follow-ups where the implementation team needed direct feedback from researchers.

After incorporating this initial feedback into the guidelines, the implementation team began engaging with a broader group of stakeholders who would be able to provide different perspectives on what organizational changes might be required to operationalize these recommendations. Additionally, involving a broad range of stakeholders from across the company helped get people acquainted with the guidelines prior to the official roll out, allowed the implementation team to address concerns up front, and helped us get early buy-in from the teams who would be using the new guidelines. We sought feedback by presenting the guidelines to the internal group who would make up the Human Data Review Group for any project involving enriched data, presenting the guidelines to a broader group of researchers working with human data, making the guidelines available for people to leave comments and questions, and conducting one-on-one interviews with various stakeholders to have more targeted conversations. Along the way, the implementation team continued to incorporate feedback, resolve questions, and reach out to relevant stakeholders to resolve uncertainties as they came up.

One of the primary reasons we wanted to collect feedback from researchers and other stakeholders was to identify how to bridge the gap between current practice and recommended practice at an organizational level. By talking to those involved with setting up data enrichment projects at DeepMind, the implementation team was able to develop a deeper understanding of the types of resources that would enable researchers to consistently meet the recommended guidance. This feedback informed the content created, either as a part of the guidance itself or in the form of additional resources linked out of the guidance. In some cases, the feedback led us to identify external (to the company) barriers that would make it difficult for researchers to adopt aspects of our recommendations.

Endnotes

1 Geiger, R. Stuart, et al. "Garbage in, garbage out? Do machine learning application papers in social computing report where human-labeled training data comes from?." *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency*. 2020

2 Denton, Emily, et al. "On the genealogy of machine learning datasets: A critical history of ImageNet." *Big Data & Society* 8.2 (2021): 20539517211035955.

3 Hutchinson, Ben, et al. "Towards accountability for machine learning datasets: Practices from software engineering and infrastructure." *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*. 2021

4 Gray, Mary L., and Siddharth Suri. *Ghost work: How to stop Silicon Valley from building a new global underclass*. Eamon Dolan Books, 2019

5 Geiger, R. Stuart, et al. "Garbage in, garbage out? Do machine learning application papers in social computing report where human-labeled training data comes from?." *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency*. 2020