ABOUT ML in Practice

An Example from the Humanitarian Sector

JANUARY 18, 2023

PARTNERSHIP ON AI

Introduction

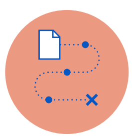
Transparency has been widely recognized as a key principle of AI ethics, one that can also enable other AI ethics goals. The Partnership on AI's (PAI) ongoing ABOUT ML* initiative has focused on one particularly promising approach to operationalizing transparency: the documentation of machine learning (ML) systems.

* Annotation and Benchmarking on Understanding and Transparency of Machine-learning Lifecycles

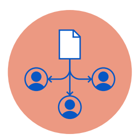
Through ABOUT ML, PAI has been bringing together a diverse range of perspectives to develop, test, and implement ML system documentation practices at scale. This Pilot Summary serves as description of how the ABOUT ML team sought to improve documentation practice at one organization and the lessons we learned along the way.

This document summarizes activities and discussions from the Pilot study where we explored opportunities to improve ML documentation practices within the Peer Review Framework developed by the Centre for Humanitarian Data from the United Nations Office for the Coordination of Humanitarian Affairs (UN OCHA). After having made incremental changes to the framework, the team thought it was the right time to take a long, hard look at the current framework and make some structural changes as needed. In subsequent workshops, we built off the discussion to find opportunities to improve the framework. The goals for the ABOUT ML Pilot are to:

1. Improve current documentation practices at UN OCHA through reviewing, workshopping, and learning about best practices



2. Understand how ABOUT ML recommendations are applied in real-world settings



3. Share extrapolated and generalized learnings with the greater responsible AI community

Background

ABOUT ML is a multi-year, multi-stakeholder initiative led by PAI that aims to promote transparent machine learning systems by identifying best practices for machine learning documentation. We have published and iterated on a Reference Document with implementation instructions as well as other tools in our Resource Library. PAI is now putting these guidelines and resources into practice by conducting Pilots with a select group of organizations who are looking to improve their documentation approaches across the ML lifecycle.

Across the four workshops, a number of key themes emerged, some of which are highlighted here:

- A need for an **earlier review** stage. Earlier access will be more impactful for the clients so that they can actively help clients while they are developing models and thinking about ethical considerations.
- **Branding** the Peer Review Framework to attract and incentivize clients to submit their models by communicating values provided by the Peer Review Framework.
- Strengthening the **visibility** of the work in various ways, such as through social media, an archive of past reviews, testimonies from previous clients, etc.
- Building **empathy** for the stakeholders of the framework by understanding their feelings, motivations, and barriers as they interact with the framework.
- Bridging the **knowledge chasm** between the reviewers and the clients, whether it's technical information or humanitarian context.

Organization at a quick glance

NAME	United Nations Office for the Coordination of Humanitarian Affairs (UN OCHA) Centre for Humanitarian Data
WEBSITE	www.unocha.org
GEOGRAPHIC LOCATION	International
COMPANY SIZE	1000+
YEAR FOUNDED	1991
ORGANIZATION MATURITY	Mature
VALUE	Connecting people and data to improve lives (Centre for Humanitarian Data)
DOMAIN	Predictive analytics in humanitarian contexts
DOMAIN REGULATION	No established standards for documenting or assessing predictive models and their intended use
FRAMEWORK	Peer Review Framework for Predictive Analytics in Humanitarian Response
DOCUMENTATION INVOLVED	Model Card Implementation Plan Model Evaluation Matrix Ethical Matrix Model Report
INTENDED AUDIENCE FOR DOCUMENTATION	Impacted populations, actors seeking to use predictive analytics models to anticipate and respond to humanitarian crises, other humanitarians wishing to improve their development process
DOCUMENTATION MATURITY	Mature
EXAMPLE CLIENTS	The 510 Global Team The Danish Refugee Council

Challenges

We mapped out the user journey along the three phases of engaging with the Peer Review Framework: when introduced to the process, while executing the process, and when evolving the process. The feelings, motivations, and barriers are identified in each of these phases.

Reviewers (Technical and Ethical Reviewers)

Reviewers are a group of experts who assess the technical rigor or ethical concerns of models during the peer review process. Reviewers are assigned to review models by the Moderator based on availability and a match of skills for the model.*

* Source: Peer Review Framework for Predictive Analytics in Humanitarian Response

STAKEHOLDER

Technical & Ethical Reviewers

PEER REVIEW FRAMEWORK	GOALS
INTRO TO PROCESS	Trying to get up to speed quickly Trying to understand UN OCHA’s rubric
EXECUTING PROCESS	Accessing relevant templates and explainers Accessing detail from client and asking questions
EVOLVING PROCESS	Adding to the Centre’s Catalogue of Predictive Models Making recommendations to improve the framework Recording lessons learned for the Centre and for the client

FEELINGS

As Reviewers get onboarded with the process, they might be **curious** about when and which model they will get invited to review as well as how much time is expected. They might possibly be **overwhelmed** by the process explanation and **worry** about deadlines (which are usually lightly suggested).

As Reviewers review the models, they might be **concerned** about meeting clients face-to-face and presenting their reviews since the reviews are not anonymous.

Finally, after the review is complete, Reviewers might have **questions** about whether the recommendations are being used or implemented by clients.

MOTIVATIONS

Reviewers might have several motivations. Reviewers might be intrinsically motivated to improve the use of models in the humanitarian sector. Another motivation might be a sense of responsibility to safeguard the interests of some priority stakeholders, typically affected populations.

In addition, doing this work can provide the Reviewers with exposure to the humanitarian sector and expand their network. Reviewers might be particularly interested in presenting the reviews in a way that will show their expertise and understanding of the subject matter. In addition, they might want visibility of the review results for their peers and recognition in the sector.

BARRIERS

Reviewers might have challenges finding the time to do this work since they are reviewing on a volunteer basis in their spare time. The expected scope might be unclear, for instance, whether they are expected to run the code or which supplemental documentation they are expected to read. In addition, the lack of visibility beyond what the Client submits, which is not a neutral description of the model and implementation, can be a challenge.

For Reviewers who are used to reading papers, the model card format can be challenging to read. Finally, the lack of visibility on the actual impact of the review is a challenge. This was also mentioned in surveys completed by Reviewers.

DISCUSSION

Due to the lack of visibility on the impact despite time and effort put in, Reviewers want evidence that points to their impact. *Visibility and communication of the rigor and impact of this work are key.*

Moderators

Moderators are members of the Centre that lead and coordinate the peer review process. Moderators will onboard the Client and provide the Model Card template for the formal submission of the model.*

* Source: Peer Review Framework for Predictive Analytics in Humanitarian Response

STAKEHOLDER

Moderators

PEER REVIEW FRAMEWORK	GOALS
INTRO TO PROCESS	Trying to match reviewers with prospects Trying to understand UN OCHA's rubric
EXECUTING PROCESS	Answering questions from reviewers and clients Coordinating timing and submission Organizing and storing inputs
EVOLVING PROCESS	Preparing and delivering the final report with reviewers Collecting learnings for clients

FEELINGS

Moderators might **question**, “If we asked the client to do the review, do they really want to be doing it?” In addition, they might feel **uncertain** when matching the reviewer and client as well as the model. They might **lack the understanding** of the approach to both technical and ethical reviews, given the depth and breadth of knowledge required.

As the review process wraps up, Moderators might feel **hopeful** that the client is willing to make the review public, and might **wonder** if the process has been useful to the client.

MOTIVATIONS

In a broad sense, Moderators might be motivated to ensure the quality of models in the field. Within the context of the framework, Moderators want to make sure that the Reviewers have what they need and also that Clients and Reviewers understand each others’ viewpoints.

Moderators might also want to provide feedback to improve the overall process and get as much useful information out as possible and tell the story of the framework in a meaningful manner.

BARRIERS

Relying on the client to provide the necessary information can lead to a risk of information being kept from the model card because it might reflect negatively on the Client.

Moderators might find it challenging to make concrete asks to volunteers (Reviewers), such as timelines. In addition, it is difficult to coordinate so many actors, especially for client consultation, and the overall process can take a very long time.

DISCUSSION

- Clients not sharing all the information on model cards can also be a problem in the field.
- Part of the job of the Moderator is to get that information and pry if needed.
- It is important to get to the impact of peer review for each actor and measure of success.

Clients

The Clients submit models that are either fully developed or exploratory for peer review.*

* Source: Peer Review Framework for Predictive Analytics in Humanitarian Response

STAKEHOLDER

Clients

PEER REVIEW FRAMEWORK	GOALS
INTRO TO PROCESS	Learning about the Centre Determining the ROI of participating in the Framework Getting acclimated to the actors
EXECUTING PROCESS	Filling out forms and providing documentation Answering questions and giving the Moderator, Technical Reviewer, and Ethical Reviewer access to stakeholders
EVOLVING PROCESS	Providing detail via survey about their framework process experience Citing actions to be taken based on review outcomes

FEELINGS

As Clients get introduced to the framework, they might get **daunted** by the long process. They also might feel **hesitant** to open up work to scrutiny and negative publicity. They might **worry** that the implementation plan might expose weaknesses in processes beyond just the model itself and that the model may be deemed “unethical.”

Once they receive their reviews, Clients might feel **reluctant** to accept criticism and share negative outcomes. They also might feel like sharing positive review results might seem **prideful**.

MOTIVATIONS

Clients might be motivated to increase publicity of their models and gain perception as leaders in the field. Going through the review process could be seen as a stamp of approval/ endorsement.

Through the review, Clients might better understand how technical models can be improved.

They might want to idealize the process and paint a better picture than the actual practice to lead to positive outcomes and avoid negative publicity and attract more funding.

BARRIERS

Completing the model card can be tedious and take a long time. A single “Client” may be more than one person or organization; for instance, model and implementation can be developed by different actors. Thus, gathering information from the different actors might require additional time and effort. Unlike academic peer review in journals, implementing changes is not required, and a lot of work may be needed. Finally, it would be hard to make the case for publication given the high risk and low perceived reward.

Our Approach

At the time of our engagement, the organization had six team members who worked remotely in two different time zones. The team members brought a wide range of expertise in AI, system validation, and quality assurance. As a startup preparing a product launch and juggling multiple priorities, the team did not have an opportunity to engage in in-depth discussions about their documentation practices.

Our hypothesis was that by bringing the team together to talk about and think about the bigger picture of documentation, we could make meaningful changes in the way the team thought about documentation and how they might implement it in their day-to-day work.

During the Pilot, the PAI team and the organization convened remotely in a series of biweekly workshops, which involved:

- In-depth reviews of current documentation artifacts and processes
- Facilitated discussions amongst cross-functional team members
- Human-centered design exercises such as persona- and user-journey-mapping to better understand stakeholder needs and experiences

Asynchronously, participants completed an online diary study each week to share updates and capture insights. Participants also completed an online survey at three points over the course of the Pilot to reflect on the value of documentation and assess their own internal documentation processes and artifacts.

Findings

Here are some key takeaways from these workshops:

Initial barriers for clients. Clients have to read a long document about getting involved in the peer review framework, which might be a barrier for clients initially. Building off the prototypes from this workshop, a one-pager that summarizes the process and focuses on why this process might be beneficial for the clients can be developed.

Incentivizing the clients. Another way to lay out the incentives to get involved in the peer review is to emphasize the stick rather than the carrots, e.g., how it is irresponsible not to do modeling without review. In academia, no one will accept your results if they are not reviewed. We need to hold models in humanitarian work to the same standards.

Knowledge gaps between clients and reviewers. First, there might be a knowledge gap within the cross-functional client team, for instance, between technical teams and decision-makers in terms of technical details. The reviewers have technical and ethical expertise but might lack the clients' humanitarian context. Thus, improving communication and education on these topics to reduce the gaps would be helpful (e.g., explaining jargon, creating a glossary of terms, writing in a way non-experts could understand, etc.)

Promoting the framework. The team shared several ideas for promoting the framework. For example, asking the clients and reviewers to share about the framework on different platforms, publishing the reviews on ReliefWeb, creating a public archive of previous reviews, etc. If these ideas are implemented, we believe there is great potential for the peer review framework to be utilized and well-known in the humanitarian community.

Insights

Through user-journey-mapping exercises, we explored the feelings, motivations, and barriers that the three main roles (reviewers, moderators, clients) might experience through the Peer Review Process. This provided an opportunity to empathize with the stakeholders and walk in their shoes.

We explored the feelings, motivations, and barriers that the three main roles might experience.

We gained the following insights for each stakeholder:

- **Reviewers:** Main areas of improvement are around the visibility of their impact and the clarity on the scope of the review.
- **Moderators:** Main concerns revolve around coordinating the review processes and getting information from clients who may not feel comfortable sharing.
- **Clients:** Main concerns are hesitation to open up their models for review as well as the effort it takes to complete the model card.

We discussed how the current framework could be reimagined as a lighter peer review process to lessen the burden for both the Client and the UN team. In addition, buy-in from the community could be increased through rebranding and storytelling strategies.

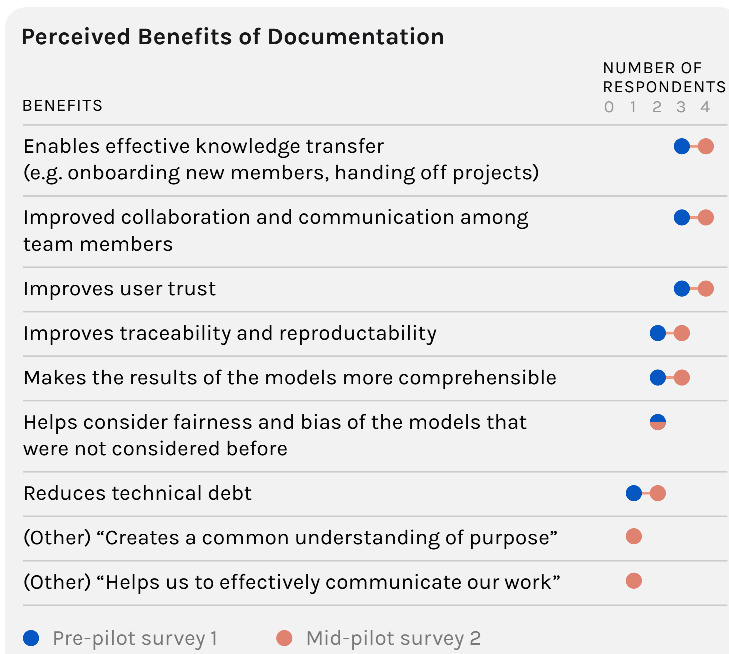
Perceived Value of Documentation (Benefits, Costs, Challenges)

1. PERCEIVED BENEFITS OF DOCUMENTATION

Best documentation practices lead to an overall greater sum of transparency. What does greater transparency entail when we really look into it? Some examples include effective knowledge transfers, improving collaboration and communication between team members, helping consider fairness and bias of the models that were not assessed before, reducing technical debt, improving traceability and reproducibility, making model results more comprehensible, and improving user trust. Of course, each case differs in the sense of perceived benefits. Survey participants had the opportunity to think critically about perceived benefits at both the pre-pilot and mid-pilot stages.

Pre-Pilot

In the pre-pilot survey, all three participants agreed that their current documentation work allowed them to enable effective knowledge transfers like onboarding new staff or handing off projects, improve collaboration and communication amongst team members, and improve user trust. One participant marked that it does also reduce technical debt, and two participants agreed that it improves traceability and reproducibility, makes model results more comprehensible, and finds gaps in seeking fairness and bias in models.



Respondents were asked to select all benefits applied at a given time.

Mid-Pilot

In the mid-pilot survey, all four participants kept the same top three responses (enable effective knowledge transfers, improve collaboration and communication amongst team members, and improve user trust). The lowest responses remain with reducing technical debt, and finding gaps in seeking fairness and bias in models, which, interestingly enough, remains the same as the pre-pilot survey responses. Optional benefits were added as “other” by two individuals citing that their current work helps them “create a common understanding of purpose and findings in a review” and it “helps us communicate the work effectively.”

The data tells us that half of the participants believe that the current work is helping them consider the fairness and bias of the models that were not thought about before and that the work is reducing technical debt.

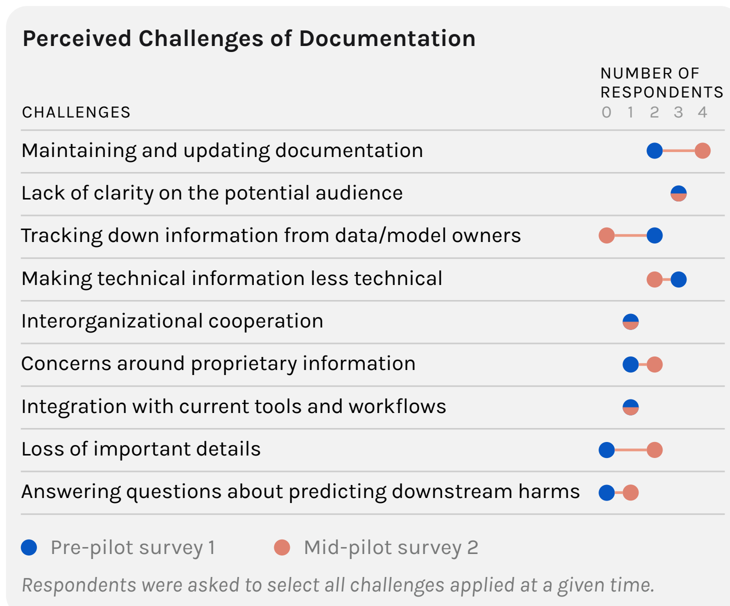
2. PERCEIVED COSTS OF DOCUMENTATION (AVERAGE RATING)

The need to incorporate documentation can be costly, as seen in the graph below. The highest cost of both surveys is time, and the lowest cost is the financial cost. The second and third highest are effort and human resources, respectively. It is not surprising that time is the highest cost since documentation efforts take time. Best documentation practices also take intentional effort and human resources, which is reflected by the team's responses.

3. CURRENT PERCEIVED CHALLENGES OF DOCUMENTATION

Pre-Pilot

In the graph at right, we can see that the major challenge is actually tied between "making technical information less technical" and the "lack of clarity on the potential audience." As a result of this data, we centered our second workshop session on targeting the various audiences and stakeholders that might be involved and what level of understanding each key stakeholder needs to know and understand in the third workshop session.



Mid-Pilot

In the mid-pilot graph, the major challenge is no longer tied between "making technical information less technical" and the "lack of clarity on the potential audience." Still, it is rather clearly "maintaining and updating documentation" with "lack of clarity on the potential audience" as the second major challenge. One of the benefits of documentation we've mentioned before is the ability to reduce technical debt, and we've seen in those surveys that it ranks as the lowest perceived benefit by the team. This is also reflected in the comments about wanting the model card to be simplified and easier to understand for stakeholders without a great degree of technical knowledge.

Anonymous elaborations from the team
(Comments from the surveys)

> “I have seen that many times the documentation is either very technical (accessible only to ML specialists) or too high-level with little technical information. What is missing is a way to explain in a way that is accessible to non-technical users the assumptions, the inputs and the limitations of the models (including potential use-cases). That’s the information we have tried to include in the model card of the peer review framework.”

> “One of the main challenges we are facing in the Model Card is how to make the documentation succinct and comprehensive. The target audience of the Model Card are non-technical partners who may get lost in long technical documentation.”

> “It is tough to find the balance of what to document. In the end I think we often have two kinds of documentation, one more technical and the other for a non-technical audience.”

> “The team I’m on has little structure in place for documentation strategies. We try our best but the information tends to be highly scattered (we mainly use Google Drive, but also have important information on Sharepoint, Slack, email…). Furthermore, we use Google Drive as our wiki which has several drawbacks.”

> “Some information is too sensitive to track in widely accessible sheets, which can create parallel workflows, costing more time. Together with regular documentation and tracking this can create resistance against formalizing processes and tracking.”

Pilot Feedback

In the last impact survey, participants were asked questions about their overall experience with the Pilot—what worked well, what could be improved, and what the insights gained were. We compiled participants' responses to each question.

Overall, did you gain any new insights from the Pilot? If so, what were some valuable things you learned?

- Insights on how to position the work of the team.
- A need to figure out how to influence the system at a higher level before developing the Peer Review Process.
- A better understanding of the audiences/ stakeholders part of the process.
- What the position and the voice of our team might look like in the humanitarian space.
- New viewpoints and perspectives of the Peer Review Process

Is there anything you would change or implement in the Peer Review Framework based on the insights gained from the Pilot?

- How we frame and present the added value it could bring our clients.
- Reconsidering the Peer Review Process's goals and how we advocate for the framework.
- Improve and clarify the outcomes for the clients.

How can we make asynchronous work easier for participants in different time zones if you found it challenging to do so?

- "I realized that it has not been easy, mainly given conflicting priorities and busy agendas on our ends. Overall, I think the asynchronous work was relevant and not too much time demanding."
- "Yes, I did find it difficult. But a solution is hard. Maybe that participants in the same time zone come together to do the work, and then during the synchronous times, the two (or more) groups update each other with their insights?"

How has the Pilot process helped you and your organization achieve your initial goals? (E.g. Become an authoritative voice, documenting earlier in the process, incentivizing stakeholders to participate in the Peer Review Process, etc.)

- "Not yet, but I think it provides a good starting point for reflecting on the next strategy on the peer review framework."
- "We still have to implement the new insights we got from the pilot. But as said above, it definitely helped to understand better what we mean by those goals and start thinking of ways how we should concretize them."
- "Made us take a step back and evaluate the framework from different perspectives."

If you could change something about the Pilot, what would you change? Or was there an opportunity you hoped to address or see that the Pilot did not cover?

- Making the objectives and final outputs of the pilot clear to everyone from the beginning.
- Wanting more tangible action into actually changing the Peer Review Process and perhaps having this mindset from the start. "Use momentum to implement new ideas."
- Being present for the entire process.

With the completion of this ABOUT ML Pilot, we are continuing to incorporate our lessons learned to solidify and standardize our approach for future Pilot cohorts. Our ongoing learnings will also help identify what practices work best for different organization types and iterate on guidance as additional Pilots are launched.

We invite you to stay up-to-date with the ABOUT ML work by subscribing to our mailing list. To learn more about joining future Pilots, please reach out to Jiyoo Chang (jiyoo@partnershiponai.org).

Acknowledgments

Jiyoo Chang, Inclusive Research and Design Program Lead and former Research Fellow for the ABOUT ML Program, led the synthesis of the Pilot engagement outputs and the development of this summary document. She had extensive support in the Pilot execution and document creation from Albert Tanjaya, consultant for the ABOUT ML Program. For their tireless efforts, we are very appreciative.

We are grateful to Michael Hind, Meg Mitchell, Kush Varshney, Hanna Wallach, and Jennifer Wortman Vaughan for contributing their expertise and valuable feedback to this work. We would like to thank Kasia Chmielinski (visiting consultant) and Dr. Tina Park (Head of Inclusive Research and Design at PAI) for providing insights, and Dr. Christine Custis for overseeing this project as the Director of Programs and Research at PAI.