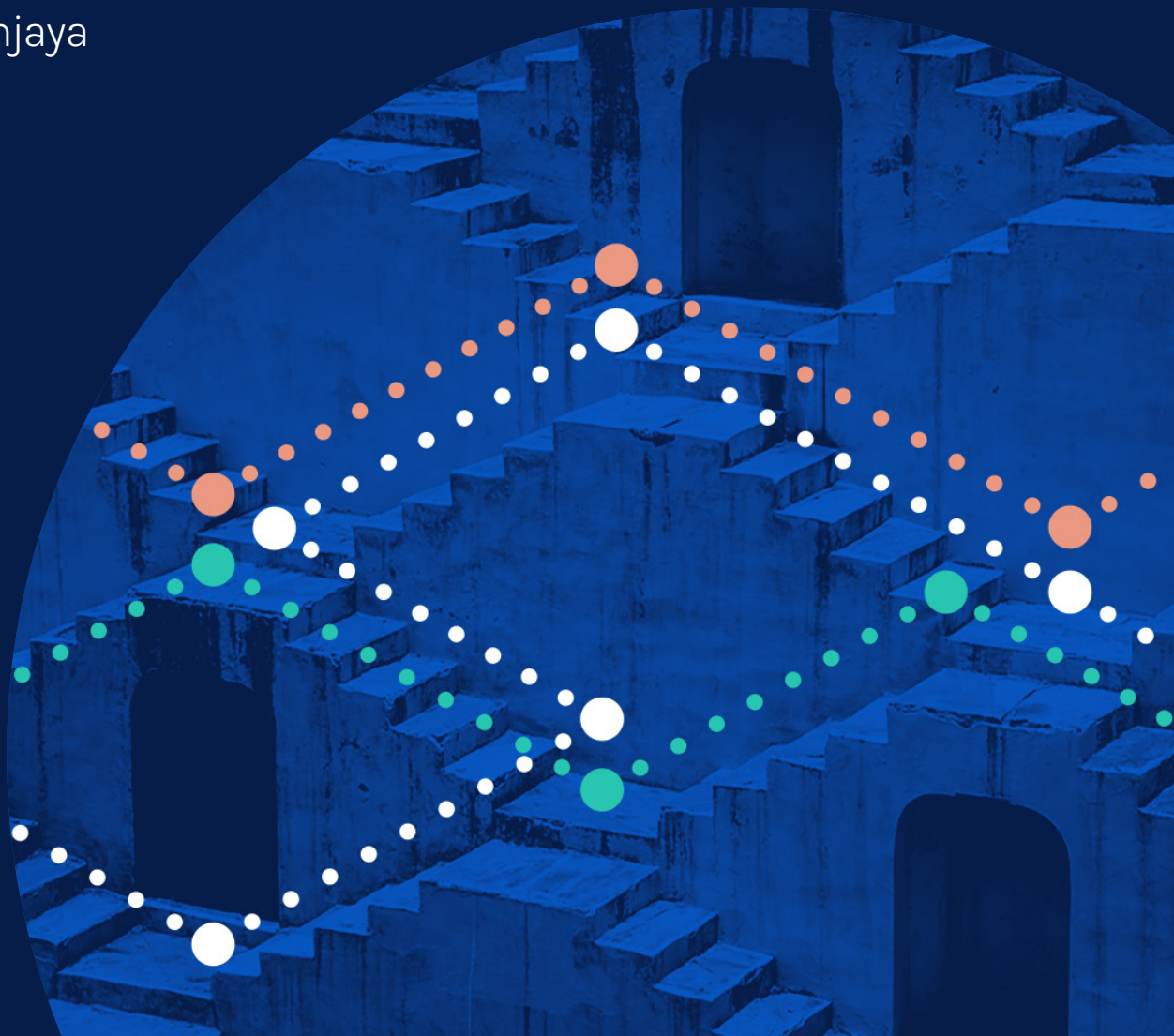
PARTNERSHIP ON AI

Documentation for Better Transparency

Lessons from Three ABOUT ML Pilots on the Practice of AI/ML Documentation

Albert Tanjaya



As artificial intelligence systems become more integrated into daily life, the need for transparency in how those systems work and operate grows. A better understanding of the AI/ML development, deployment, and decision-making processes can support user trust in AI/ML systems. Users need assurance that these systems will reliably offer accurate and informed outputs, safeguard against failures, and protect and uphold privacy. Transparency involves making a system's properties, purpose, and origins clear and explicit to users, practitioners, and other impacted stakeholders. It is a foundation for both internal accountability among the developers, deployers, and API users of an AI/ML system and external accountability to customers, impacted non-users, civil society organizations, and policymakers.

Governments and international government institutions are pressing for stronger transparency of AI systems to ensure safe and reliable use. The recently passed EU AI Act, the White House Executive Order on the Safe, Secure, and Trustworthy Development and Use of AI, and the Hiroshima Process International Code of Conduct for Advanced AI Systems are just some examples of emerging policy calls emphasizing the importance of transparency. However, these examples differ in their specific requirements and requests. Standardizing clear and rigorous approaches to the documentation of models, training data, and impacts on society would advance these shared goals of transparency and the societal accountability it enables. Standardization would also reduce unhelpful compliance burdens created by overlapping but different requirements for transparency issued from different jurisdictions and authorities.

PAI's ABOUT ML* initiative aims to increase the standardization and improve the rigor of AI/ML documentation. "Documentation" includes two key components: "artifacts," which provide details of ML systems, and "processes," which describe the steps followed to create artifacts.

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

ML Documentation in Practice

PAI's initial contribution to standardization was a detailed reference document of best practices in AI documentation, cited in the White House Blueprint for an AI Bill of Rights as an important effort to move from principles to practice. After identifying that many organizations run into implementation challenges when attempting to institute or improve AI documentation practices, PAI conducted pilots to test and improve the real-world usability of the ABOUT ML recommendations. Regardless of an organization's size, specific domain, or product, the lessons learned in these pilots can help all organizations understand the challenges of implementing documentation best practices.

RESOURCES

ABOUT ML Reference Document

ABOUT ML Resource Library

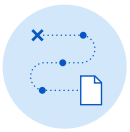
The most recent ABOUT ML pilot was conducted in collaboration with Intuit, a large

consumer-facing financial software and services company. PAI and members of the Intuit pilot explored how their responsible AI intake review processes could be further improved to streamline artifact and process creation. Pilots were also conducted with the United Nations Office for the Coordination of Humanitarian Affairs (UNOCHA) and Biologit, a start-up in the pharmacovigilance field.

Over the course of the three pilots, we saw the following themes emerge in the pursuit of AI transparency and enabling better AI/ML documentation:



THEME 1
Identifying common challenges to achieving transparency through documentation practices



THEME 2
Coordinating internal procedures to develop better documentation practices



THEME 3
Tailoring documentation for the needs of different stakeholders

RESOURCES

PAI has completed three ABOUT ML pilots. Read the summary PDFs:

biologit

INTUIT

OCHA

Themes Across ABOUT ML Pilots



THEME ONE

Identifying common challenges to achieving transparency through documentation practices

Transparency is a widely recognized core value of many organizations' responsible AI principles. It is foundational to other responsible AI goals like accountability and safety. Emerging regulatory policies, attention from civil society, and the development of field-wide norms from the AI industry's evolving practices of self-governance are increasing the pressure on organizations to move from stated principles to actual practices. To respond, organizations must actively define what transparency looks like for their work and how to achieve it, partly through the artifacts and processes of documentation.

The organizations that participated in the pilots had a range of transparency goals shaped by the maturity of their own practices and the relative levels of practice in their respective sectors. **Biologit** aimed to establish itself as a leader in AI transparency within the pharmacovigilance space by implementing robust documentation practices throughout its AI/ML development process. Similarly, **UNOCHA** wanted to advance new documentation standards and norms for models used for predictive analytics in the humanitarian sector. As a part of a broader goal to automate its documentation efforts, **Intuit** sought

to understand the current state of its documentation practices and the gaps to be filled. In pursuing their goals during the pilots, each organization encountered challenges and resource prioritization questions. To better understand the challenges faced by pilot teams, team members completed short check-in surveys throughout the pilots.

One question asked participants to select all the challenges they faced implementing documentation from a list of nine common issues on this topic (or to offer their own answers). Three challenges rose to the top for each pilot, several of which were shared by more than one organization. Across all three pilots, "Maintaining and updating documentation" was frequently selected by participants, indicating a shared challenge despite organizational differences in size, funding, and maturity of documentation. One **UNOCHA** pilot team member stated,

> "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."

Another member from **Biologit**'s pilot said,

> "It seems we [have not yet] clearly documented in our processes what will trigger the production and refresh of documentation. Most are created on an ad-hoc basis or often long after a release, creating the chance of doc production being missed and getting stale."

Difficulties with updates and maintenance to processes and practices are common for organizations well beyond the realm of AI/ML documentation. Maintenance and updates require valuing incremental improvements to what already exists, not just celebrating new efforts. Deliberate attention and resourcing are needed to ensure artifacts do not become out of date.

Two other frequently cited challenges were "Lack of clarity on the target audience of documentation" and "Making technical information less technical," which are among the top three most frequently selected challenges for the **Biologit** and **UNOCHA** pilots. It is harder to define audiences for documentation when there are potentially many users interacting with their documentation efforts and knowing how to tailor the content to a given user's background. Since **Biologit** was an early stage start-up at the time of the pilot, they needed to spend more time on what artifacts and processes they wanted to create and who those artifacts would serve. The **Biologit** pilot team sought clarity on how to frame their artifacts around the highly regulatory pharmacovigilance space. For **UNOCHA**, the pilot team wanted to motivate humanitarian actors to deploy new technologies and tools to aid in their respective missions, all while ensuring that new technologies did not create

negative societal impacts. Their unique position as a player in the global humanitarian space gave them an opportunity to help advance documentation standards and norms of models specializing in predictive analytics. The **UNOCHA** pilot team wanted to dive deeper into how their documentation process could bring value to both non-technical and technical humanitarian actors and have them engage with the organization voluntarily since **UNOCHA** is a not-for-profit organization.

The **Intuit** pilot team had two other challenges in their top three most selected. "Tracking down information from data/model owners" and "Integration with current tools and workflows" were tied as the second most frequent set of challenges. Since **Intuit** is a large organization with a global reach and a mixture of staff with various skill backgrounds, they were interested in automating much of their documentation to improve effectiveness and efficiency. **Intuit**'s pilot team hoped to combat these challenges by centralizing its documentation artifacts through an in-house platform. To tackle this goal, internal coordination would be needed to ensure information was accessible and intelligible.

Understanding the resourcing challenges in documentation practices is crucial as it involves investment in time, effort, and funding.

Understanding the resourcing challenges in documentation practices is crucial as it involves investment in time, effort, and funding. These resources, whether for new or updated documentation, are often traded off against other organizational needs. To gain a deeper insight into these challenges, participants were asked to rank the given costs from 1 (very low) to 5 (very high). This question aimed to identify the most significant resourcing challenges and advocate for strategic allocation of resources in documentation efforts.

- Time
- Effort
- Financial cost (e.g., purchase of new software)
- Human resources (e.g., team members involved in documentation)

Across all pilots, team members rated "time" as high, which appeared to be genuine resource constraint. The **Biologit** and **UNOCHA** pilot teams rated "effort" as high. This is likely due to their organizations' nature as a start-up and a nonprofit, respectively, since start-ups tend to need more initial work to get the organization running, and non-profit organizations tend to need to market its mission well. Compare this to **Intuit**, which is already a well-established organization. For each pilot team, members consistently rated "Financial costs'' as low or very low, indicating that documentation efforts could likely be supported through existing systems used for other purposes.

Organizations that want to support documentation practices need to be mindful of the work that documentation takes and ensure that staff have the necessary time, resources, and structures to complete documentation tasks in addition to their other responsibilities.

Key Takeaway

Organizations should support a culture that values transparency by identifying and addressing documentation challenges and barriers. They should anticipate and plan for documentation challenges. Chances are that the primary challenge found across these three pilot teams — maintaining and updating documentation — may not be unique to just the three pilot organizations but other organizations as well. Addressing these challenges will require providing the necessary resources to empower teams to prioritize and contribute to documentation practices; employee time and effort deserve particular attention here. Even though "financial cost" was the lowest ranked cost from an individual perspective, there are clear indirect financial costs for the organization as a whole associated with the time and effort required to support documentation practices—but the alternative of poor or absent documentation can create its own costs through inefficient internal coordination or external reputational costs.



THEME TWO

Coordinating internal procedures to develop better documentation practices

While the pilots had different team compositions, they were intentionally selected to include team members representing different roles and experiences. Pilot teams included a range of representatives from model developers, data scientists, and product managers to CTOs and directors. Diversity in the pilots is integral because each pilot team member holds key knowledge of how their organizational team contributes to their current documentation practices. Effective documentation requires collaboration and coordination between cross-functional teams. This collaboration allows for better visibility into existing artifacts and processes, and helps identify opportunities for standardization and performing documentation maintenance checks. Additionally, it promotes careful consideration of jargon and person or team-specific content.

To explore what internal documentation practices currently exist, we hosted workshops for the pilot teams to indicate:

1. What internal organizational teams were involved in creating an artifact
2. What and when artifacts were created
3. Which part of the lifecycle was being documented
4. How differently each internal organizational team engaged with both artifacts and processes.

For the **UNOCHA** pilot, the team wanted to assess their documentation process, also known as the Peer Review Framework. It is a single unified process that involves a moderator who onboards clients and guides them through a series of peer reviews before publishing a model report. "Clients" are teams of humanitarian actors who voluntarily

submit models for review through **UNOCHA**'s documentation process. These teams can be a mix of technical and non-technical experts who know their models and have to engage in a technical and ethical review to evaluate a given model's information via an in-house model card.* Moderators are guides who help walk clients through the process, being the ones who are in constant contact with clients.

* Model card: A type of documentation artifact accompanying a given model that details the model itself, its intended uses, potential limitations, training parameters, datasets used, experimental information, and model evaluation results.

(Definition from Hugging Face. See the ABOUT ML Gloassary for this and other terms.)

Their pilot focused on one collaborative documentation process between three different stakeholders (clients, technical and ethical reviewers, and moderators) with a defined set of artifacts (e.g., model cards, the model report, ethical matrix, etc). The workshops highlighted needs in the interactions between clients, reviewers, and moderators, which included:

1. How to successfully generate stronger buy-in at the start of the framework
2. How to effectively facilitate better communication in meetings
3. How to show the impact of work that relies on voluntary commitment from both clients and reviewers.

The pilot team, as architects of the framework, found that they needed to proactively engage with their clients earlier on in the documentation process. This is essential to encouraging buy-in and relaying the importance, value, and impact of the work for the humanitarian field since documentation norms in this space are not well established. The pilot team also recognized that although clients understood how their models worked and how to translate the code and algorithms into technical write-ups, it was not always clear to clients what the ethical implications of their models were (e.g., limitations, constraints, bias, risk, etc). The pilot team realized that moderators and reviewers needed to be better equipped to lead communication and facilitation challenges. Such examples include:

1. Carefully considering the language used in the review of their feedback so everyone could understand
2. Conveying the value and impact of the process
3. Explaining the importance of ethical considerations and how that might impact the models themselves and the communities they serve.

Similarly, since submitted models exist in the context of specific humanitarian work, clients need opportunities to explain and further elaborate that background to reviewers, which might aid in how ethical considerations are made. The pilot team determined that more interactions between clients and reviewers would be beneficial to help surface questions and comments earlier on in the process but due to the voluntary nature of the work, people's time constraints should be taken into consideration.

Unlike **UNOCHA**'s single documentation framework, the **Biologit** and **Intuit** teams each had multiple documentation processes that included a range of internal teams for different artifacts. Each internal stakeholder created documentation artifacts differently depending on the team they were part of. The minimal standardization of the documentation creation process among cross-functional teams led to varying qualities and types of artifacts. In both pilots, there were instances of limited knowledge sharing between and across

teams. Due to the organization's size, the **Intuit** pilot team found it difficult to track down documentation from different model and data teams, and relied on engaging with individual model and data team leads. They further elaborated that some standardization might be useful, especially if all models had the same information within their artifacts. The **Biologit** team recognized a need for more coordination between team members regarding the documentation artifacts they were creating, leading to a more structured review of their practices.

Key Takeaway

For any organization looking to build out their internal best practices, it's important to start by establishing internal procedures across all relevant teams within an organization.* Such procedures may include:

* Different artifacts may benefit from centralized or decentralized approaches in the creation process.

- Developing or instilling a shared understanding of the "why" or purpose of documentation for internal teams interacting with documentation artifacts and processes.
- Building better communication between internal teams responsible for creating documentation artifacts.
- Ensuring coordination between relevant teams in the artifact creation process (e.g., coordination between engineering, marketing, leadership, etc.) to allow for better visibility and involvement in current practices.
- Making documentation artifacts more accessible across the organization, including access points and knowledge transfers of artifacts located within an organization and careful consideration of person-or-team-specific content that goes into the development of the artifacts.

Coordination of internal practice can be helpful to achieve the desired quality and standard of these artifacts as well as building opportunities to engage teams in different modes of knowledge sharing.



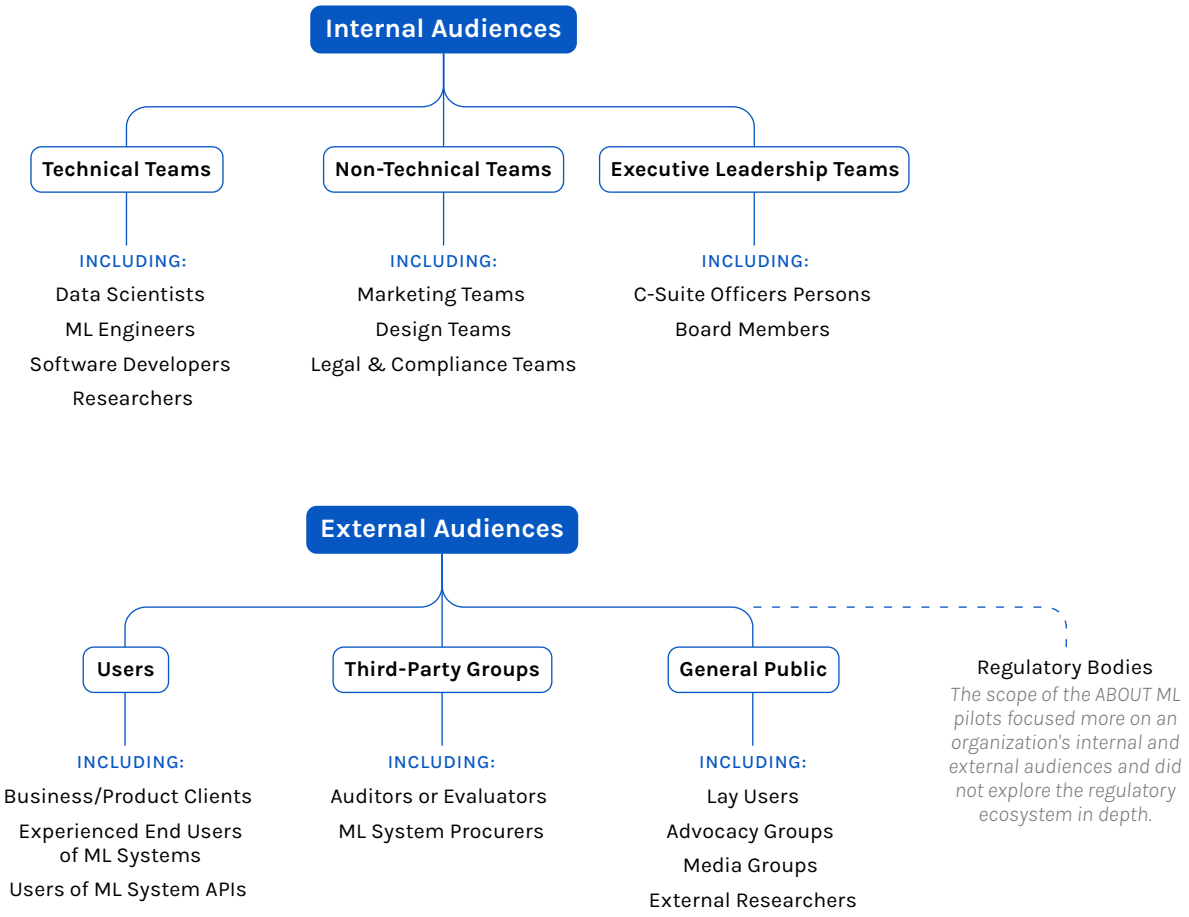
THEME THREE

Tailoring documentation for the needs of different stakeholders

So, who does documentation exist for? The pilots explored how different audiences have different goals and uses for documentation. By listing out the different audiences and their documentation needs, pilot teams could understand what information served only one audience compared to what information served multiple audiences. The development and deployment of ML systems impact not only organizations' internal relationships (e.g., between different role hierarchies) but also their relationships with different external and regulatory stakeholders (e.g., product users, regulatory bodies, etc). Each group has different levels of technical understanding, specific information uses, and unique

expectations from documentation. Transparent documentation means more than just providing a one-size-fits-all artifact. It ensures each group receives information in a format they can understand, information that can be utilized for their own end goals, and information that supports organizational accountability. Organizations can help serve the goals of their various audiences and stakeholders through effective documentation by providing relevant and accessible information, thus promoting transparency and accountability to build trust in a complex relational ecosystem.

A useful place to start would be to classify these audiences into two broad groups, internal audiences and external audiences. A non-exhaustive list of these groups includes:



In all three pilot processes, participants identified their audiences and focused on internal technical and non-technical teams and product users navigating their documentation. The user journey workshops gave insight into how stakeholders might interact with their documentation. These insights surfaced key considerations like user sentiments, incentives and motivations, and barriers to accessing and understanding documentation artifacts and processes.

The exercise helped the **UNOCHA** pilot take a deeper look into how their clients felt

incentivized and motivated to not only continue through the framework process but to promote it as well. For **Biologit** and **Intuit**, the exercise helped them consider how their current artifacts (e.g., factsheets and model cards) met the needs of the various stakeholders who would utilize them. The **Biologit** pilot determined that their user-facing documentation could be improved by making the language and content more user-friendly and tailoring it for different audiences. The **Intuit** pilot elaborated on evaluating the ideal state of documentation for their Responsible AI & Legal team and end-user clients. For their Responsible AI & Legal team, they pointed out the need for an aggregated documentation view for all their models and the need for an easy distillation of important information for other stakeholders that interact with this team. This team also faced some challenges in how documentation was collected, as discussed in Challenge 2. For their end-user clients, the pilot team also noted that it might be important to understand where and when AI models are used in interactions with **Intuit** products and what information might be useful to share on how their data is used. Since **Intuit**'s pilot team was already in the process of exploring the use of a comprehensive version of a model card, this pilot was able to procure insights on what current information they already have, missing information they don't have, and where they might procure the missing information.

Key Takeaway

Organizations that want to achieve clarity on what to document and who to document for should consider mapping out the user journey of various intended audiences. By focusing on the sentiments, incentives and motivations, and barriers for intended audiences any team can determine a good starting point for specifying documentation goals. If possible, it is highly recommended that all stakeholders be actively involved in this exercise. It is important to note that this exercise is useful even if a given organization's target audience cannot be involved, for example, if processes are proprietary knowledge.

What's Next

For any questions, concerns, or to learn more about the ABOUT ML work, please email Albert Tanjaya.

The next exciting phase of work involves advancing the PAI Guidance for Safe Model Deployment ("PAI Guidance"). This workstream aims to assess the progress of the field of documentation and understand how different stakeholders comprehend and utilize various documentation best practices from the PAI guidance to operationalize transparency. PAI will engage with model providers, model deployers, policymakers, and civil society organizations to achieve these goals. We plan to gain insights into the progress made in the field of documentation for foundation models, develop a template for a documentation artifact, and identify and highlight important shared documentation practices. As a result, we will define more detailed guidance to bridge the gaps between emerging regulatory

frameworks like the US Executive Order on AI, the EU AI Act, etc, and industry practices.

Acknowledgments

Albert Tanjaya, Program and Research Associate of the ABOUT ML Program, led the synthesis and development of this summary document with support from PAI staff. Stephanie Bell, Aimee Bataclan and Thalia Khan contributed editorial support, and Neil Uhl contributed design support.

Thanks to Dr. Tina Park, Jiyoo Chang, Kasia Chmielinski, and Dr. Christine Custis for building the design and the execution of the initial pilot processes. Thanks to Biologit, UNOCHA, Intuit, and the pilot participants for their time and resources in completing their pilots.

We are grateful to Michael Hind, Meg Mitchell, Kush R. Varshney, Hanna Wallach, and Jennifer Wortman Vaughan for contributing their expertise and valuable feedback as advisory committee members for the ABOUT ML work at Partnership on AI.