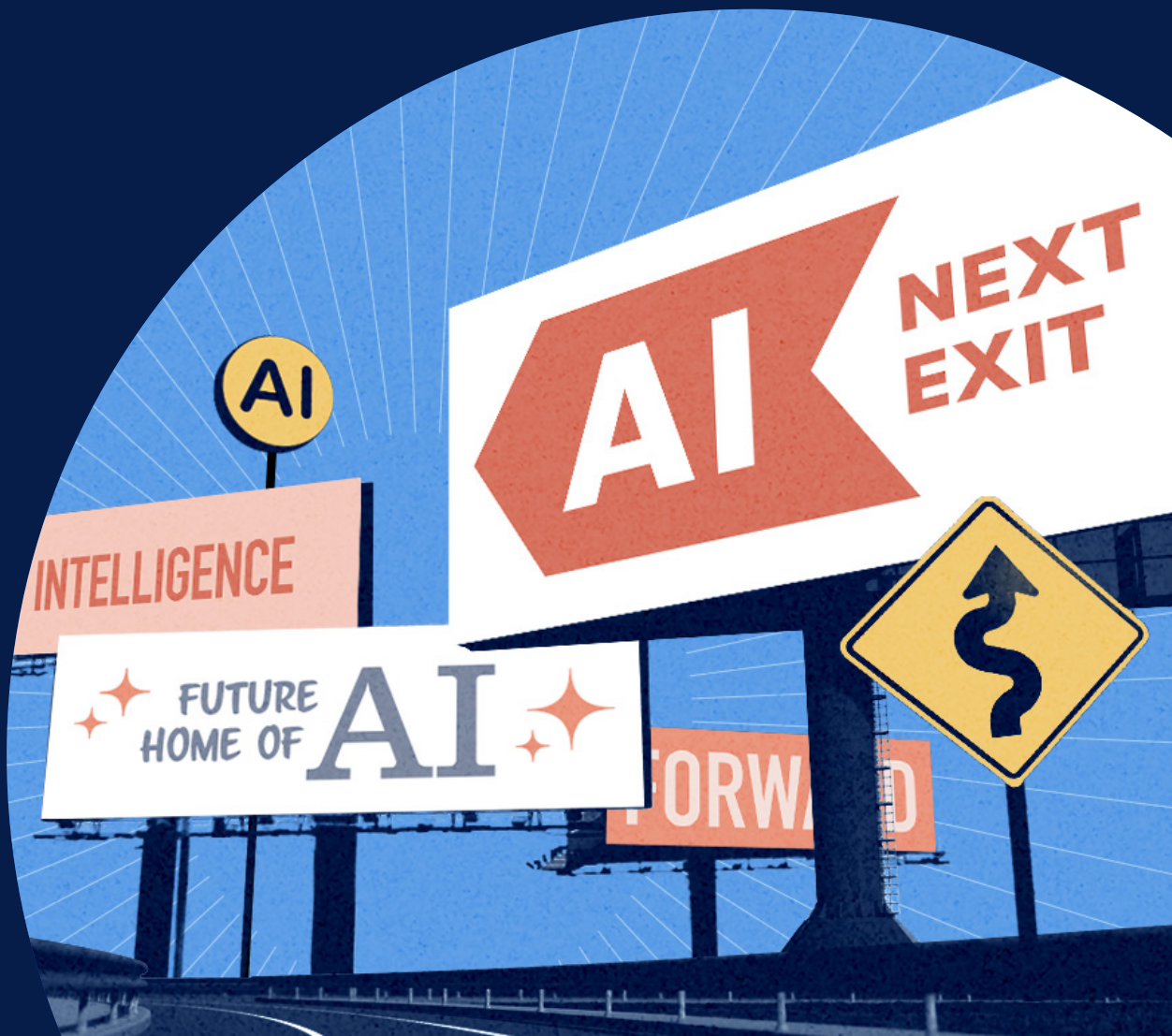
PARTNERSHIP ON AI

Responsibly Navigating the Enterprise AI Landscape

Promises, Challenges, and Opportunities

Sarah Villeneuve
Albert Tanjaya



Contents

Introduction

Since the release of ChatGPT in 2022, AI adoption by businesses and other organizations around the world has accelerated. According to McKinsey, organizations reporting use of generative AI in at least one business function jumped from 33 percent in 2023 to 71 percent in mid-2024. Decreased costs and greater accessibility through cloud-based infrastructure, pre-trained models, and user-friendly interfaces are transforming the way businesses and other organizations do their work.

The next generation of AI, agentic AI, is also starting to see uptake by companies. Agentic AI refers to AI systems that act on behalf of users, acting with some degree of autonomy to achieve goals without human intervention or guidance. In contrast to other forms of AI, which require human input and oversight, agents are designed to understand a user's general goals and utilize context to solve specific problems without explicit instructions.

This increased enterprise AI use is taking place against a backdrop in which responsible AI guidance is mostly focused on design and development, less so on use. While development and design decisions shape the possibilities of an AI system's impact through use, the actual use decisions made by businesses and other organizations ultimately define its real-world effects. The same AI model or system can often be used in many ways; its harms and benefits are frequently determined by its adopters.

The same AI model or system can often be used in many ways; its harms and benefits are frequently determined by its adopters.

Given the importance of businesses and other organizations' role in guiding AI's impacts, it's not surprising that research conducted by IBM found many leaders are concerned about how to responsibly adopt AI:

- 80 percent of business leaders view AI ethics, explainability, trust, and bias as major hurdles to the adoption of generative AI
- 50 percent of business leaders see a lack of infrastructure and governance required to manage generative AI as a barrier to adoption
- Many executives want more clarity on AI standards and regulations before making major investments in generative AI

Responsible adoption brings many returns on investment, ranging from positive reputational standing to maintaining a competitive market advantage. Importantly, responsible use also benefits society through ensuring widespread technologies are used safely, transparently, equitably, and in ways that build trust.

For businesses and other organizations that are looking to adopt AI models directly from foundation model providers, there is a clear need for guidance to ensure responsible procurement, adoption, and use. No single organization has all the answers, and it will take a comprehensive multistakeholder effort to get it right.

While recommendations for businesses and other organizations adopting enterprise AI are beginning to emerge, a full description of the challenges these organizations face when navigating AI adoption remains unclear. To begin building a better picture of the kind of guidance required for responsible enterprise AI adoption, Partnership on AI (PAI) hosted two workshops, in December 2024 and February 2025, with the support of Salesforce. At the workshops, PAI brought together more than 20 organizations, including private businesses, foundation model providers, civil society organizations, and academic institutions. The workshops were designed to identify key challenges organizations face in relation to responsible AI adoption and explore paths forward for future collaboration and research.

What Do We Mean by Businesses and Other Organizations That Use Enterprise AI?

This report focuses on businesses and other organizations that use enterprise AI. This includes companies, non-profits, foundations, and other organizations whose main purpose is not to develop or design AI systems or products, and which seek to adopt and utilize AI internally. In this capacity, organizations may:

1. **Acquire AI:** Procure AI solutions directly from foundation model providers[1] for internal use. AI solutions may include restricted API models and open models that may have differing release capabilities (narrow purpose or general purpose).[2]
2. **Customize AI:** Adapt, fine-tune, or build upon AI models using proprietary data and in-house expertise to address specific business or organizational needs.
3. **Use AI:** Implement AI solutions internally.

1 See PAI's Deployment Guidance for Foundation Model Safety

2 See PAI's Definitions on the different models and their release types

PAI brought together more than 20 organizations, including private businesses, foundation model providers, civil society organizations, and academic institutions.

Key Challenges to Responsible AI Adoption for Businesses and Other Organizations

In the workshops, participants identified three critical challenge areas that businesses and other organizations face throughout the various phases of AI adoption:

1. Responsible AI adoption readiness
2. Evaluation, monitoring, and compliance
3. Trust and collaboration across the AI value chain

Responsible AI Adoption Readiness

ESTABLISHING ORG-WIDE AI GOVERNANCE INFRASTRUCTURE	Prior to the adoption of AI, organizations should establish formal structures and principles to guide responsible AI across the organization. Many participants in our virtual convening recognized the importance of institutional oversight (such as through the creation of an AI Council or Committee), aligning responsible AI use with existing organizational standards, and establishing responsible AI principles. However, organizations can lack a clear picture in terms of what an ideal org-wide AI governance infrastructure should involve.
ORGANIZATIONAL AI INVENTORY & MANAGEMENT	With the rise of widely accessible generative AI tools, organizations must work to understand both "official" AI use (AI that was procured through a formal organizational process), and "unofficial" AI use (AI that has been adopted informally by employees), what roles these AI tools play internally, and whether they align with the business or organization's standards and values.
AI ETHICS EDUCATION IN THE WORKPLACE	Many organizations lack guidance, standards, or best practices when it comes to teaching employees how to use AI tools responsibly. Additionally, staff may lack needed knowledge to make informed choices about AI's ethical implications and potential impacts. Building organizational awareness and understanding of AI is critical to ensuring AI is adopted and used responsibly.
BALANCING MARKET PRESSURE WITH RESPONSIBLE AI ADOPTION AND USE	Many organizations are facing increased pressure to adopt AI to gain advantages over competitors and remain relevant in an evolving marketplace. However, this raises a significant challenge of balancing innovation goals and speed with ethical dimensions such as safety. Additionally, a combination of market pressure, competitive needs, and the fear of missing out can overshadow the practical need for an AI solution, leading to adoption without a clear problem definition. This can lead to companies attempting to address real issues their employees are facing with AI products, and ignoring or neglecting alternative, more effective solutions that are not AI-based. Organizations should not only have clear criteria to assess the AI solution's fit to the organizational problems or needs, but also an understanding of what infrastructure and resources are required to responsibly adopt and maintain a certain AI solution.
HUMAN-AI INTEGRATION	Organizations need to assess the impact that an AI solution would have on their workers, and consider how to best balance human and AI capabilities within the workplace.

Evaluation, Monitoring, and Compliance

EVALUATION & MONITORING CHALLENGES OF AGENTIC AI SYSTEMS	The advent of agentic AI has been met with uncertainty regarding how traditional frameworks for evaluation and monitoring might apply. Compared to generative and predictive AI, agentic AI carries significant unpredictability in terms of the actions it may take, and the outcomes it may produce. Not only do AI adopting organizations (along with the broader AI field) currently lack benchmarks to evaluate new capabilities against, but they also lack clarity into how to balance real-time vs. retrospective monitoring, and what constitutes as "human-in-the-loop".
STANDARDIZATION OF EVALUATION AND MEASUREMENT FRAMEWORKS	Across the broader AI ecosystem and AI value chain, there is a lack of consistent evaluation and measurement frameworks to assess the quality and performance of AI systems. This creates uncertainty for business and enterprise organizations in the procurement and deployment process. This includes inconsistent evaluation metrics and benchmarks (e.g. different companies deploy different methodologies to measure performance; different companies have different risk appetites affecting which evaluation benchmarks they use; and sector-specific differences impact what constitutes acceptable performance) as well as uncertainty in determining monitoring depth (e.g. determining what amount of information about a models performance is sufficient). Lack of standardized metrics for measuring performance increases the risk that models may be used in ways that could be unsafe, unreliable, and untrustworthy, and ultimately result in real downstream harms.
NAVIGATING COMPLEX AND EVOLVING REGULATORY REQUIREMENTS	Organizations need to navigate an increasingly complex regulatory landscape. For organizations that operate across different jurisdictions, this often includes navigating conflicting regulatory requirements which can lead to adopting fragmented AI solutions across different regions posing integration and implementation challenges. Regardless of where an organization operates, they are faced with the challenges of keeping pace with the rapidly evolving AI regulation landscape, balancing existing and anticipated regulatory requirements, and the practical challenges of ensuring their AI systems meet these requirements in practice.
ENVIRONMENTAL IMPACTS OF AI MODELS	Organizations must consider the environmental impacts of AI adoption to ensure they are meeting their organizational sustainability objectives. This involves understanding the full environmental footprint of AI models that include both upstream development and downstream use. However, much of this information is either poorly captured or unaccessible to adopting organizations.
CONTINUOUS AI SYSTEM MAINTENANCE	Businesses and other organizations face the challenge of navigating ongoing updates and maintenance of AI systems they acquire. AI system maintenance can be both resource- and time-intensive (such as conducting continuous security quality assurance), temporarily disrupt operations (due to incoming system updates), or introduce security vulnerabilities for the business or organization.

Trust and Collaboration Across the AI Value Chain

INFORMATION TRANSPARENCY AND TRUST CHALLENGES ACROSS THE AI VALUE CHAIN	One of the most pressing challenges organizations face relates to the lack of information shared by foundation model providers. Limited transparency into model development, capabilities, limitations, and evaluation (such as version tracking and performance across different contexts) and data usage and governance policies (e.g. what data is being collected by model developers when a system is used by another entity) pose significant barriers to responsible AI adoption by businesses and other organizations. This lack of transparency leaves businesses and other organizations unable to inform their own employees, investors, and other internal stakeholders about the limitations of their AI usage, and exposes their organization and society more broadly to the attendant risks.
DATA GOVERNANCE AND ACCOUNTABILITY FOR ENTERPRISE AI	Businesses and other organizations face the unique challenge of working efficiently with foundation model developers to establish a shared understanding of accountability when it comes to the adoption and use of enterprise AI. This may include identifying and agreeing on who is responsible for conducting and providing information on data quality and utility, model accuracy and performance, model outcomes, as well as data collection and management. Without appropriate accountability shared between organizations and foundation model providers, preventable downstream risks can go unseen and unaddressed resulting in social harms.
STAKEHOLDER ENGAGEMENT IN AI ADOPTION AND USE	Identifying stakeholders who might be affected by the adoption and use of AI and ensuring inclusive feedback to inform an organizational approach and address concerns or harms posed by an AI model is a critical but difficult step.

Areas for Future Research and Collaboration

Reflecting on the challenges listed above, participants surfaced four areas in which future research and collaboration would support responsible adoption of enterprise AI. These are:

1. Knowledge and terminology alignment
2. Governance structures
3. Adoption and implementation guidelines
4. Measurement and monitoring frameworks

Each of these areas are outlined below in more detail.

Knowledge and Terminology Alignment

Improving knowledge sharing and developing a shared terminology for enterprise AI can help to overcome the challenge of building organizational awareness and capacity for thinking critically about responsible AI as well as being able to effectively communicate both internally (among teams and staff members) and externally (with foundation model developers). This can, in turn, increase AI literacy among employees to help reduce misuse of AI and increase the understanding of possible risks and harms associated with AI systems in the workplace.

Developing a Common Taxonomy of Enterprise AI Terms

Businesses and other organizations often find themselves speaking different languages when interacting with foundation model developers. This creates difficulty when requesting specific information from, or asking questions directly to, model developers. Without a shared lexicon to base in-depth discussions on, details can get lost in translation that can be critical to how an organization assesses an AI solution from a model developer, or manages downstream risks and issues that arise internal to an organization. More work is needed to develop a common language for enterprise AI between adopting organizations and foundation model developers.

Developing an AI Literacy Guide for Enterprise Workers

In addition to developing for a common taxonomy, participants highlighted the importance of educating their workforce from basic awareness for all employees to advanced technical understanding for specialized teams. Utilizing such a guide can help employees take part in meaningful participatory engagements to help align workers with training, policies, and guidelines that match an organization's responsible AI values. Many organizations have already begun to develop their own internal guidelines for staff when it comes to using AI. The ACLU, for example, has developed an internal use policy for generative AI. Likewise, the Future of Privacy Forum released their own Generative AI Internal Policy Checklist. Since differing organizations have different values, internal policies, and internal teams, more work is needed to identify the common materials teams with similar functions across organizations might need to get started.

Governance Structures

Setting up appropriate infrastructure for governance and accountability is not only critical to supporting responsible AI adoption within businesses and other organizations but can also provide critical information to foundation model providers to improve safety and security, inform data collection methods, and support version control.

Define the Anatomy of a Responsible AI Council

Participants of both workshops recognized the importance of establishing a responsible AI council internal to the business or organization to oversee the development of organization-wide responsible AI principles in alignment with other organizational standards and policies, and provide guidance on the adoption and use of models. While many large organizations have responsible AI councils or teams, for example, Microsoft's Office of Responsible AI, more collaboration is needed to develop a shared understanding of best practices when it comes to the ideal anatomy of a responsible AI council for businesses and

other organizations. Recognizing that while specific structure may differ from organization to organization, defining the core components of a responsible AI council, such as objectives, function, composition of members and their roles, would provide essential guidance for organizations seeking to adopt AI responsibly.

Define Roles and Responsibilities Across the AI Value Chain

Defining the roles and responsibilities of each actor is a crucial step in establishing accountability for assessing and managing risks at each stage of the enterprise AI model lifecycle. Roles and responsibilities should be determined based on who is best positioned to understand, monitor, and act upon a certain set of risks. This requires further collaboration and information sharing between foundation model developers and the business or organization seeking to adopt their models for internal use, and will likely differ sector to sector.

Adoption and Implementation Guidelines

How businesses and other organizations choose to evaluate and procure AI can have significant upstream and downstream impacts. The information businesses and other organizations require during the procurement process can significantly influence a foundation model developer's documentation processes, such as information about the model's capabilities, limitations, performance, safety, and security. Acquiring more information during the procurement process can lead to benefits downstream, as employees within the business or organization adopting AI will be more informed about the model, possibly resulting in less risk and/or harm at both the organizational and societal level.

How businesses and other organizations choose to evaluate and procure AI can have significant upstream and downstream impacts.

Establish Standards for Responsibly Assessing and Procuring AI

Businesses and organizations need clear information to make informed decisions when procuring AI solutions. However, the lack of a standardized procurement template, which may cover information like evaluation metrics and benchmarks, makes the procurement process difficult. It is made especially challenging because the evaluations and benchmarks vary not just between foundation model developers and businesses or organizations, but between foundation model developers themselves and across sectors. Moreover, businesses and organizations may lack access to documentation from model providers on information such as training data, latency, security controls, performance evaluations, versioning, limitations, capabilities, and model weights. Even when businesses and organizations can access artifacts such as model cards or system cards, they still lack the critical information required to make responsible decisions. Past PAI research has also highlighted the importance of worker-centric considerations during AI procurement, such as understanding the working conditions of data enrichment professionals. Developing standards for assessment and

procurement criteria, as well as expectations for documentation by model providers through more detailed model card templates, can serve as a basis to ensure that downstream actors, such as businesses and other organizations, are being provided with the type of information they need to make responsible procurement decisions.

Develop Risk Guidance for Adopting Agentic AI

Given the rapid development of AI capabilities, it's no surprise that businesses and other organizations are in need of guidance for assessing the risks of agentic AI. Since these technological capabilities are so new, there are, understandably, a lot of unknowns. Unlike traditional and generative AI systems, agentic AI systems can independently execute tasks, access multiple tools, and make decisions with limited human oversight. However, many businesses and organizations lack specific in-house approaches or frameworks for evaluating the distinct risks and harms these technologies present when used, including concerns about control mechanisms, appropriate levels of autonomy, evaluation strategies, and monitoring requirements. Developing risk guidance for adopting agentic AI systems can help businesses and other organizations establish new AI governance mechanisms, an understanding of "AI risk" specific to their sector, and create new mitigation strategies for when the risks become harms actualized within an organization.

Develop Responsible Enterprise AI Implementation Guidelines

Businesses and other organizations may be left wondering how to translate their responsible AI values and principles into practical implementation guidelines across different departmental teams. Without clear and cohesive implementation guidelines, teams can be left applying inconsistent standards and practices when interacting with deployed systems. It can lead to misalignment in how risks are managed, how systems are monitored and evaluated, and overall documentation efforts across the organization. AI adopting businesses and organizations can benefit from these guidelines to ensure that their internal policies and safety guardrails are consistent, align with their responsible AI values, and can reliably inform evidence-based decision making across the organization. Mulitstakeholder and multi-sector collaboration is required to develop a standard template to be used and iterated upon by different organizations.

Mulitstakeholder and multi-sector collaboration is required to develop a standard template to be used and iterated upon by different organizations.

Measurement and Monitoring Frameworks

In order to effectively monitor AI models once implemented, businesses and other organizations will need to choose appropriate measurement and monitoring frameworks. Tracking performance, documenting emergent risks, and documenting post-monitoring analysis can not only help ensure an AI model is operating safely and securely and mitigate downstream risks, but can also provide critical information to foundation model developers about performance issues, and safety and/or security concerns. Such metrics and monitoring efforts are not limited to just a technical analysis and understanding of a deployed system, but an understanding of how those deployed systems are behaving in real-world contexts and their effects on stakeholders and impacted communities. For this reason, businesses and other organizations can benefit from having both a technical and sociotechnical approach not only to improve their deployed systems but also to generate positive impacts on their end users and workers.

Develop and Implement Post-Deployment Monitoring Framework for Enterprise AI

Post-deployment monitoring presents significant challenges for businesses and other organizations implementing models internally, as models often behave differently in real-world environments than in testing phases or sandbox environments. A monitoring framework would establish standardized protocols for observing multiple aspects of AI models in use, including but not limited to policy compliance, user interactions, security operations, and system performances. Agentic AI specifically might pose additional challenges than traditional and generative AI, since they interact autonomously with other systems which can lead to incremental risks. Therefore, additional monitoring mechanisms and further human-in-the-loop mechanisms become necessary. In addition to an internal monitoring framework, businesses and other organizations should consider implementing incident response and disclosure mechanisms to help create feedback loops and foster trust with employees and other affected stakeholders as means for them to report issues upstream to both the internal AI monitoring body, as well as externally to foundation model developers.

Effective frameworks must be adaptable to different operational contexts, as monitoring requirements vary substantially across industries and use cases.

Effective frameworks must be adaptable to different operational contexts, as monitoring requirements vary substantially across industries and use cases. By implementing consistent post-deployment monitoring, organizations can identify potential issues before they escalate, document incidents methodically, and gather empirical data to inform future governance decisions internally and system improvements that can be relayed to the foundation model provider. However, many businesses and other organizations currently lack guidance on how to effectively conduct post-deployment monitoring within their own organizational environments.

Incorporate Metrics for Broader Societal Impacts

A sociotechnical understanding of an AI system's impact in a real-world setting is critical for businesses and organizations deploying these technologies. AI systems impact various communities, the environment, and social systems. Businesses and other organizations may find it difficult to identify the right societal impact metrics or indicators that are measurable and relevant to the AI models being used internally, since this is a challenge for even foundation model providers to get right. By expanding post-deployment monitoring beyond technical performance and metrics, businesses and other organizations can better identify unintended consequences, demonstrate responsible stewardship, and make more informed decisions about when to modify a system or decommission a model when necessary. It will require a larger cross-disciplinary multistakeholder effort to determine appropriate and meaningful metrics for organizations and context to assess for broader societal impacts.

Next Steps and How to Get Involved

PAI will be continuing this work through the Enterprise AI Steering Committee. If you are interested in participating in future work on this topic, please contact Sarah Villeneuve.

Acknowledgments

We appreciate the invaluable input provided by experts who participated in our workshops, including:

James Alexander, Salesforce

William Batholomew, Microsoft

Kathy Baxter, Salesforce

Daniel Berrick, Future of Privacy Forum

Dawn Bloxwich, Deepmind

Ruby Bolaria Shifrin, Chan Zuckerberg Foundation*

Trey Causey, Indeed

Hank Dempsey, Anthropic

Ian Eisenberg, Credo AI

Elisa Espinoza, Chan Zuckerberg Foundation

Rachel Gillum, Salesforce

Sam Gong, LinkedIn

Paula Goldman, Salesforce

Brian Green, Santa Clara University

John Hearty, Mastercard

Brittany Hinchliff, Salesforce

Chris Hopfensperger, Google

Sakshi Jain, LinkedIn

Ruchika Joshi, Center for Democracy and Technology

Derek Kennedy, BCG

Marques Matthews, Uber

Emily McReynolds, Adobe

Madison Mohns, Indeed

Lama Nachman, Intel

Chigozie Okorie, The Annie E. Casey Foundation

Nicole Ozer, ACLU

Alok Panigrahy, Credo AI

Bonnie Parisi, Salesforce

Nikki Pope, Nvidia

Irina Raicu, Santa Clara University

Eric Schroeder, ServiceNow

Nivedha Sivakumar, Apple

Amy Smith, Intuit

Saishruthi Swaminathan, IBM

Sarah Tan, Salesforce

Reggie Townsend, SAS Institute

Kush Varshney, IBM

Kip Wainscott, JP Morgan Chase

Eric Yu, SAS Institute

** This person has since moved on from their organization at the time of the report release.*

Special thanks to PAI staff who have been integral to the development of this report:
Aimee Bataclan, Neil Uhl, Rebecca Finlay, Stephanie Bell, Madhu Srikumar, Eliza McCollough, Jessica Gratton, and Tina Park.