From:	Evangelos Razis
To:	aiframework
Subject:	Workday"s Response to NIST"s AI RMF RFI
Date:	Tuesday, August 17, 2021 10:40:17 AM
Attachments:	image003.png Workday Comments_RFI_NIST AI RMF_8.17.21 (Final).pdf Workday Whitepaper_Building Trust in AI and ML.pdf

---

Dear Colleagues:

On behalf of Workday, I am pleased to submit: 1) our response to NIST's Request for Information on an AI Risk Management Framework; 2) and our recent whitepaper, *Building Trust in AI and ML*.

Workday firmly supports NIST's efforts to advance trustworthiness in AI, including through an AI Risk Management Framework. We thank NIST for this opportunity and its commitment to engage stakeholders.

Please do not hesitate to reach out if we can be of any assistance or if you have any questions.

Kind Regards,

Evangelos Razis

Evangelos Razis | Senior Manager, Public Policy |

workday.

Workday's Response to the National Institute of Standards & Technology's Request for Information on an *Artificial Intelligence Risk Management Framework*

August 2021

Workday is pleased to respond to the National Institute of Standards and Technology's (NIST) request for information on an *Artificial Intelligence* (AI) *Risk Management Framework* (RMF).

Workday is a leading provider of enterprise cloud applications for finance and human resources, helping customers adapt and thrive in a changing world. Workday applications for financial management, human resources, planning, spend management, and analytics have been adopted by thousands of organizations in the U.S. and around the world and across industries—from medium-sized businesses to more than 45 percent of the Fortune 500. Workday incorporates machine learning (ML) technology within our applications that enable customers to make more informed decisions and accelerate operations, as well as assist workers with data-driven predictions that lead to better outcomes. Workday believes ML technology has the potential to impact enterprises in the near-term by making operations more efficient. In the longer term, organizations will be able to reorganize their operations around machine learning and AI's unique possibilities.

AI is becoming an ever-increasing and transformative presence in our lives, driving human progress in countless ways. To achieve AI's full potential, however, there must be broad confidence that it is being developed ethically and used responsibly. With this in mind, Workday welcomes NIST's efforts to develop "forward-thinking approaches that support innovation and confidence in AI systems" and is pleased to offer the following comments.

I. Support for NIST's AI Risk Management Framework

Workday is a firm supporter of NIST's efforts to advance trustworthiness in AI, including through the development of a risk management framework. The issue of AI trustworthiness is ready-made for NIST, as it has a well-developed track record in convening government, industry, and other stakeholders to cooperatively develop cutting-edge voluntary frameworks. Given the success of the *Framework for Improving Critical Infrastructure Cybersecurity* and the *Privacy Framework – An Enterprise Risk Management Tool*, NIST has best-in-class experience with open, transparent, and collaborative processes.

Workday has played an active role in supporting NIST's leadership in this area through extensive legislative efforts and coalition building. We were pleased to see AI framework provisions receive bipartisan support and inclusion in the *Fiscal Year 2021*

National Defense Authorization Act. We are keen to build on these efforts by constructively engaging in NIST's AI workstreams.

As NIST, together with stakeholders, embarks on the path of developing an AI RMF, timeliness is essential. The European Union is moving forward with the *Artificial Intelligence Act* (AI Act), which would require organizations developing and deploying covered AI applications to implement risk-management systems. Workday recently provided comments to the European Commission on this consequential legislation. It is essential that the U.S., EU, and other like-minded governments harmonize their approaches to AI risk management and regulation so as to prevent unneeded barriers to transatlantic trade, investment, and innovation from developing. Indeed, NIST's RMF may facilitate enterprises' compliance with their future legal obligations under the AI Act. With Europe's proposal certain to become law, such harmonization is impossible if the U.S. lacks a consensus approach to managing AI risks, voluntary or otherwise. NIST's work on an AI RMF is very likely the most expeditious vehicle for establishing a consensus U.S. approach to building trustworthy AI.

If done in a timely, collaborative, and iterative manner, NIST's work will serve as an important U.S. contribution to global AI policy and to international regulatory cooperation, including to discussions under the auspices of the U.S.-EU Trade and Technology Council. These efforts would build on U.S. leadership at the Organization for Economic Cooperation and Development (OECD) and the Global Partnership on Artificial Intelligence, which bolster U.S. economic competitiveness and values.

II. Workday's Trustworthy AI Framework

In addition to NIST's work on an AI RMF, Workday supports AI regulation that is risk-based, enables innovation, and meaningfully addresses issues of trustworthiness. Earlier this year, we released a white paper, *Building Trust in Artificial Intelligence and Machine Learning*, outlining a framework for promoting trust, accountability, and transparency, while also giving organizations broad flexibility to innovate. This paper offers an informed perspective based on our experience in offering ML-based services to our business customers as NIST begins the framework development process.

Specific provisions we call for in our "Trustworthy by Design" regulatory framework would require organizations to:

- *Adopt Principles & Publish a Trustworthy AI Policy.*
  The paper calls for organizations to adopt principles setting out their trustworthy AI commitments and publish a public trustworthy AI policy addressing identified core elements of ethical artificial intelligence. Organizations would be expected to provide only a summary of their governance framework in the trustworthy AI policy itself. It would serve as a reference point for regulators and allow an appropriate agency to impose sanctions for misrepresentations.

- *Adopt an AI Governance Framework.*
  The governance framework should include:
  - Designation of a senior leader and establishing a trustworthy AI compliance team;
  - An approach to AI Impact Assessments and data documentation; and
  - Commitments for personnel training and providing cross-company compliance resources.

- *Implement Procedures to Identify and Mitigate Harmful Bias.*
  The paper calls on AI developers to implement and describe procedures to identify and mitigate sources of potentially harmful bias in their AI systems, both in the AI models they develop and in the data they use (including data used to train their systems and the data analyzed by those systems in real-world settings). It also calls on AI developers to document the procedures they use to test for, identify, and mitigate the effects of potentially harmful bias, as well as establish diverse teams to design and develop AI systems.

- *Conduct Impact Assessments.*
  The paper recommends a separate impact assessment for each AI system that developers create. That impact assessment should identify potential risks to individuals or society, reasonably quantifying the amount of risk and degree of potential harm and listing safeguards adopted to mitigate these risks to an acceptable level.

- *Maintain Data Documentation.*
  The paper calls for AI developers to document the provenance of AI training data and take reasonable steps to test whether the use of these datasets may lead to unfair or discriminatory outcomes.

- *Provide User Transparency & Recourse.*
  The paper calls for AI deployers to provide users appropriate transparency about the individual AI systems they interact with, the safeguards implemented against untrustworthy uses of the system, and the recourse available to them, such as the ability to appeal decisions to a person. In the majority of instances, it is the AI deployer who has the most direct relationship with affected individuals and is therefore the actor best suited communicate with them.

- *Supply Information for Deployers.*
  The paper calls for AI developers to provide AI deployers with:
  - The intended purpose and the acceptable use of the AI system;
  - Steps on how the system can be properly deployed; and
  - Any known limitations in the system, model notices, and any unintended or unacceptable uses, as well as the level of human oversight, if any, that deployers should provide.

- *Provide Deployer Support for Individuals.*
  The paper calls for AI developers to provide AI deployers with sample notices and explanations, which deployers would use to communicate key aspects of the system to affected individuals. These explanations also might include:
    - High-level description of the internal workings of the AI system; and
    - The logic of the AI system and/or information about the accuracy, reliability, safety, or other features of the system.

Workday's regulatory framework would require companies to publish their trustworthy AI policies and enable customers and users to compare policies between enterprises. This in turn would create market-based incentives for companies to adopt robust, meaningful policies. The framework gives AI producers flexibility to adopt principles and practices that are most appropriate for their businesses, tailored to the type and degree of risk their AI systems present. As previously mentioned, it also envisions that future standards and industry best practices would be developed that will play a key role in enabling organizations to demonstrate trustworthiness.

There is an emerging and broad consensus throughout industry and stakeholders in the U.S. and abroad around the baseline principles that should guide trustworthy AI. These baseline principles, which NIST should account for, include fairness, transparency, accountability, and respect for fundamental human rights. For example, similar to Workday's AI regulatory framework, BSA | The Software Alliance recently published their *Confronting Bias: BSA's Framework to Build Trust in AI,* a detailed AI bias risk management framework that organizations can use to perform impact assessments to identify and mitigate risks of bias that may emerge throughout an AI system's lifecycle.

III. Response to Specific Requests for Feedback

A. Challenges to Managing AI Risk

AI risk management faces a number of challenges that reflect both the breadth of AI use cases and the emerging nature of the field. The AI ecosystem is home to a multitude of diverse stakeholders, including developers, users, deployers, and, of course, consumers. For enterprises, identifying, assessing, prioritizing, responding to, and communicating risks across complicated business relationships at scale is no small task. These challenges are amplified by the heterogeneity of AI risks, whether in degree or in kind, which include harmful bias, health and safety, privacy, and consumer protection, among others.

While the field of AI risk management is still maturing, NIST's cybersecurity and privacy frameworks serve as useful examples for how to address these challenges. An AI RMF can serve as a common grammar that organizations can use to communicate their risk management practices. With both Workday and BSA emphasizing the need for flexibility, we urge NIST to recognize that any proposed path forward that seeks to

promote AI ethics and trust across millions of scenarios and use cases in a prescriptive, one-size-fits-all manner will be unworkable.

B. Relevant Frameworks, Principles, & Policies

It is worth recognizing that, in contrast to prior NIST frameworks, the area of AI is comparatively less mature in terms of policy, regulation, standards, and best practices. Where past NIST frameworks on privacy and cybersecurity drew on more mature bodies of work, the AI RMF will be developed in an actively emerging field, without well-established approaches for systematically governing AI risks and its uses. In particular, no existing U.S. regulation holistically addresses the AI issues posed by the AI RMF, and many technical standards and best practices are still in early development. Yet this reality, in fact, makes the AI RMF exercise even more critical, because it is likely to serve as the groundwork for future approaches for governing AI in the U.S., including eventual regulation.

While AI regulation remains nascent in the U.S., the AI RMF should account for AI policy initiatives emerging elsewhere in the world. The EU's proposed AI Act, for example, moves beyond voluntary approaches to directly regulate AI applications deemed as high-risk. As Europe is the U.S.'s largest bilateral trade and investment partner, we recommend NIST consider how to leverage the AI RMF to support transatlantic regulatory cooperation, an imperative supported by political leaders at June's U.S.-EU Summit. Additionally, the Government of Singapore's *Model AI Governance Framework* is relevant, both for its consideration of governance programs and as a risk-based framework developed in collaboration with stakeholders through an iterative process.

Against this backdrop, NIST should ensure the AI RMF is iterative, scalable, and able to effectively incorporate and build on new regulations, best practices, and technical standards as they are developed.

C. Inclusion of Governance Issues

Workday strongly endorses the inclusion of governance issues in NIST's forthcoming AI risk-management framework. Simply put, organizations put in place governance programs to make tangible the goals, principles, and values underpinning trustworthiness. Absent such programs, AI risk-management is a constellation of tools and practices implemented unevenly, without transparency and accountability. Recognizing that the specifics of a governance program will necessarily vary according to the size and capacity of organizations, NIST should consider their basic elements. These include the involvement of senior management, such as appropriate C-Suite executives, to oversee the company's AI product development lifecycle, and a trustworthy AI compliance team responsible for carrying out impact assessments, documentation, training, and serving as a cross-company resource. In doing so, NIST's

AI RMF can assist enterprises and regulators alike by providing a standardized approach for ensuring accountability.

IV. Conclusion

Thank you for the opportunity to respond to NIST's request for information on an AI risk management framework. We congratulate NIST on the work put into AI thus far, including stakeholder involvement. Workday welcomes opportunities to support NIST in its efforts to develop a workable AI framework that is timely, impactful, and promotes trusted AI innovation and U.S. leadership on global AI regulatory cooperation.

We stand ready to provide further information and to answer any additional questions. Please do not hesitate to reach out to Evangelos Razis at evangelos.razis@workday.com for assistance.

workday.

Whitepaper

Building Trust in AI and ML

Through Principles, Practice, and Policy

Executive Summary

Artificial intelligence (AI) is quickly transforming modern life. Fueled by exponential growth in data, computing power, and network capacity, AI is improving healthcare, optimizing commerce, improving energy resilience, enhancing employees' careers, and driving human progress in countless other ways. Businesses use applications incorporating AI technologies across their operations to support better business decisions, accelerate operations, and deliver data-driven predictions to inform better human decisions.

To achieve AI's massive potential, there must be broad confidence that it has been developed ethically and is being used responsibly. In other words, it must be trusted.

As with any emerging technology, concerns about AI's potential risks must be addressed. These include questions about AI's accuracy and safety, its impact on human autonomy and privacy, and whether AI will treat people fairly. Unless these questions are answered directly, people may lack faith that AI systems will treat them fairly, safely, and with dignity.

Although companies, policymakers, and public- and private-sector organizations are working to build trust, current efforts are likely to leave gaps in coverage—gaps that could on the one hand give rise to irresponsible or unintended uses of AI, and on the other result in underutilization of safe and productive AI solutions.

Workday provides financial, human capital management, planning, and analytics applications to large organizations globally. Its applications are delivered through the cloud and are highly trusted by thousands of customers and tens of millions of their employees. Many Workday applications are enriched by machine learning (ML) technologies. Based on our experience developing trusted applications and our extensive public policy engagement on AI issues, we propose an AI regulatory framework to promote an ecosystem of trust. Our program would build a regulatory foundation for private-sector efforts to promote ethical AI through a set of core obligations based on widely shared goals and values.

Our proposal is based on institutionalizing a pro-innovation "Trustworthy by Design" regulatory framework. Drawing from risk-based models in the fields of cybersecurity and privacy, the framework would promote trust, accountability, and transparency while also giving organizations broad flexibility to innovate. The framework would be supported by a series of enabling measures. Chief among these is a call for policymakers to work toward greater harmonization and interoperability of AI regulatory regimes across jurisdictions. In addition, policymakers should support global standards and best practices in trustworthy AI, promote access to government-held data that may be useful for AI training or analysis, and monitor the application of existing liability rules to AI before adopting new ones. Taken together, Workday believes these proposals offer the best hope for promoting the public trust that is so vital to unlocking AI innovation.

I. Introduction

The transformations brought by AI and ML will soon affect nearly every aspect of our lives. Though AI systems have been used for years—in web searches, music streaming recommendations, trip routing recommendations, and a myriad of other applications—the prevalence of these technologies is increasing rapidly. Fueled by technology advances and a wide range of use cases, this trend is expected to accelerate.

The potential benefits to society are enormous. Two U.S. companies, for instance, recently announced the use of AI to detect early signs of Alzheimer's by analyzing patients' word usage.[1] In Germany, researchers at the University of Bonn are using AI to help detect the presence of rare diseases by analyzing a combination of patient portrait photos and genetic and other health data.[2] Others are using AI to detect and prevent fraud in real-time commercial transactions,[3] to identify new catalysts that could vastly improve renewable energy storage,[4] and to enable driverless cars to more effectively navigate routes in complex environments.[5] AI is also helping governments grapple with the COVID-19 pandemic. As just one example, the U.S. Department of Veterans Affairs is piloting an AI tool to help doctors make better treatment recommendations for patients with the virus.[6] There are countless other examples like this, with more emerging every day.

At Workday, we are fully engaged in ML-driven innovation. We are harnessing the power of ML to help our customers make more informed decisions and accelerate operations, as well as assist workers with data-driven predictions that lead to better outcomes.[7] We believe that the most transformative uses of AI are those that leverage the insights and predictive power of AI to enhance human judgment and decision-making, rather than seeking to replace it.[8]

Machine Learning Innovation at Workday

- **ML for learners.** Workday creates personal and engaging learning experiences by using ML to curate content for learners based on their skills, job role, experience, and development interests. Workday Learning recommendations based on skills rely on the foundation of the skills cloud universal skills ontology, which uses ML to maintain an evolving system of standardized, interrelated, and canonical skills.
- **ML for accountants.** The journal insights feature in Workday leverages ML to analyze and surface anomalies in journal lines automatically and continuously, which enables greater efficiency and confidence in closing accounting periods. As the journal insights feature discovers and flags anomalies, it gives accountants the ability to provide feedback to ensure only the most relevant results are surfaced.

[1] Jeremy Hsu, "AI Assesses Alzheimer's Risk by Analyzing Word Usage"; *Scientific American*, October 22, 2020.

[2] University of Bonn, "How Artificial Intelligence Can Help Detect Rare Diseases"; *Science Daily*, June 6, 2019.

[3] Louis Columbus, "Top 9 Ways Artificial Intelligence Prevents Fraud"; *Forbes*, July 9, 2019.

[4] Sam Shead, "Facebook to Use Artificial Intelligence in Bid to Improve Renewable Energy Storage"; CNBC, October 14, 2020.

[5] Rob Matheson, "Bringing Human-Like Reasoning to Driverless Car Navigation"; MIT News Office, May 22, 2019.

[6] Aaron Boyd, "VA Piloting AI to Predict Mortality Rates of COVID-19 Patients"; NextGov, October 23, 2020.

[7] Workday is a leading provider of enterprise cloud applications for finance and human resources. Founded in 2005, Workday delivers financial management, human capital management, planning, and analytics applications designed for the world's largest companies, educational institutions, and government agencies. For more information, visit workday.com.

[8] Workday, "The AI Imperative"; 2019.

AI's massive potential to help humankind, however, also comes with risks. Research published in *Science* last year, for instance, revealed that an algorithm used with good intentions to target medical interventions for the sickest patients ended up funneling resources to healthier white patients to the detriment of less-healthy minority patients.[9] In another example, in the summer of 2020 the UK government halted use of an algorithmic system to grade students for university entrance following claims that the system perpetuated educational inequity and was biased against students from poorer backgrounds.[10]

AI's risks have not escaped the attention of governments. In the United States, several federal entities, including the Federal Trade Commission (FTC) and the Office of Management and Budget, have recognized the potential of harmful bias and other risks in AI and related big data analysis.[11] The European Commission is in the process of proposing legislation to address the risks of AI,[12] and the Council of Europe is likewise exploring these issues.[13]

Given these and other examples, people's concerns about AI are understandable. How will we know, for instance, if decisions about us, or that affect our lives, are being made on the basis of AI? How can we ensure that those decisions are accurate and that AI-powered products are safe? Who's responsible if something goes wrong? Are AI systems using our personal information in ways that violate rights to privacy or personal autonomy? More fundamentally, how can we ensure that AI systems treat people fairly and are used responsibly?

These are important and difficult questions. Left unanswered, they could undermine people's confidence in AI. Unless society addresses these concerns directly and transparently, there is a real risk that AI will suffer a "trust gap"—one where people instinctively distrust AI systems because they lack faith that these systems will treat them fairly, safely, and with dignity. Once people's basic trust in AI is lost, it could be very difficult to regain.

Failure to address this trust gap early could significantly hinder the growth of AI. Companies will be less willing to invest in developing innovative AI solutions and bringing them to market if they fear that customers will not embrace them. Those who could usefully deploy innovative AI for good—from hospitals and first responders to farmers and schoolteachers—may instead fall back on less-effective options. Bad actors could exploit the situation by hiding their use of AI, or using AI in irresponsible and unlawful ways, thus exacerbating this trust deficit even further.

[9] Ziad Obermeyer, et al., "Dissecting Racial Bias in an Algorithm Used to Manage the Health of Populations"; *Science*, October 25, 2019.

[10] Sean Coughlan, "Why Did the A-Level Algorithm Say No?"; BBC, August 14, 2020.

[11] U.S. Federal Trade Commission, "Big Data: A Tool for Inclusion or Exclusion?"; January 2016. U.S. Office of Management and Budget, "Guidance for Regulation of Artificial Intelligence Applications"; January 7, 2019.

[12] European Commission, "On Artificial Intelligence–A European Approach to Excellence and Trust," pp. 10–13; February 19, 2020.

[13] Council of Europe, "Algorithms and Human Rights"; March 2018.

Leaders in both the public and private sectors already recognize the need to address this trust gap, and several have taken steps to help close it. The European Commission's High-Level Expert Group on AI (AI HLEG), comprising 52 experts from academia, civil society, and industry, released its "Ethics Guidelines for Trustworthy AI" in April 2019,[14] followed by its "Assessment List for Trustworthy AI" in July 2020.[15] These documents provide both principles and practical guidance for organizations wishing to develop or deploy AI solutions in a safe, ethical, and trustworthy manner.[16] In the United States, the Office of Management and Budget released its "Guidance for Regulation of Artificial Intelligence Applications" in January 2020 to inform federal agency development of regulatory and non-regulatory approaches to AI.[17] The Organisation for Economic Co-operation and Development (OECD), the Global Partnership on AI, and other multilateral bodies have likewise released principles or guidance to help promote responsible and ethical uses of AI.[18] Academic researchers and others are also making important contributions to this effort, including the Partnership on AI,[19] the Frauenhofer Institute in Germany,[20] and others.[21]

Importantly, responsible industry leaders are also taking steps to build trust in AI—and Workday is proud to be among them. In May 2019, we published our "Commitments to Ethical AI," which set out the six key principles that serve as a North Star for our thinking about responsible development of machine learning applications in the enterprise space.[22] We have built these principles into the fabric of our product development—including through internal processes, such as a set of ethics-by-design controls for machine learning, and robust review and approval mechanisms for the release of new technologies.[23] And this whitepaper builds on an earlier paper released in 2018, in which we raised a number of key technological and policy issues still at play today.[24] Workday, of course, is not alone in these efforts. Other leading technology providers engaged in developing and deploying AI are making similar commitments to promote trustworthy AI.[25] And we have worked with organizations such as the World Economic Forum to share best practices and lessons learned, helping to enable ethical AI development and use at other companies around the world.[26]

The Workday Commitment to Ethical AI

1. **We put people first.** Workday always respects fundamental human rights. We apply ML to deliver better business outcomes and help improve decision-making. Our solutions give customers control over how recommendations are used.
2. **We care about our society.** We believe that humans will always be at the center of work. We focus on how ML can align opportunity with talent, and on ways to develop an ML-ready workforce.
3. **We act fairly and respect the law.** Workday acts responsibly in our design and delivery of ML products and services, and strives to identify, address, and mitigate bias in our ML technologies. We aim to ensure that ML recommendations are equitable. We develop and design our products and services to enable compliance, and are engaged in the policy dialogue around regulation of new technologies.
4. **We are transparent and accountable.** We explain to customers how our ML technologies work and their benefits, and describe the data needed to power any ML solutions we offer. We demonstrate accountability in ML solutions to customers and give them a wide range of deployment options.
5. **We protect data.** The Workday Privacy Principles apply to all of our products and services, including our ML efforts. We minimize the data used and embrace good data stewardship and governance processes.
6. **We deliver enterprise-ready ML technologies.** We apply our leading quality processes—with input from customers—when developing and releasing ML technologies. We deliver meaningful ML-powered solutions that help our customers tackle real-world challenges.

[14] AI HLEG, "Ethics Guidelines for Trustworthy AI"; April 8, 2019.

[15] AI HLEG, "Assessment List for Trustworthy AI (ALTAI) for Self-Assessment"; July 17, 2020.

[16] Workday provided input to the AI HLEG on its draft assessment list: "Consultation Feedback on the Draft AI Ethics Guidelines Published by the High-Level Expert Group on Artificial Intelligence"; January 31, 2019.

[17] U.S. Office of Management and Budget, "Guidance for Regulation of Artificial Intelligence Applications"; January 13, 2020.

[18] See Organisation for Economic Co-operation and Development, "OECD Principles on AI"; accessed December 2020; Yoshua Bengio and Raja Chatila, "An Introduction to the Global Partnership on AI's Work on Responsible AI"; (September 1, 2020).

[19] See "Partnership on AI."

[20] See "Fraunhofer Institute for Intelligent Analysis and Information Systems."

[21] See "OpenAI"; "The Future of Life Institute"; "Data & Society"; "AI Now Institute"; "The Alan Turing Institute"; "Ethics of AI Lab."

[22] Barbara Cosgrove, "Workday's Commitments to Ethical AI"; May 8, 2019.

[23] Ibid.

[24] Workday, "Enterprise Intelligence: A New Frontier for Innovation"; 2018.

[25] See, e.g., IBM, "AI Ethics"; Microsoft, "Responsible AI"; Salesforce, "AI Ethics."

[26] Barbara Cosgrove, "8 Ways to Ensure Your Company's AI Is Ethical"; World Economic Forum, January 16, 2020.

The work described above is important and necessary, yet more can be done. To truly guarantee that AI technologies are trusted, and to spur the innovation that flows from that public trust, policy and regulatory efforts will need to build on the foundation of principles and ethical frameworks that exist today.

As a provider of ML-powered enterprise applications for many of the world's largest organizations, Workday is actively engaged in public policy conversations to ensure AI's trustworthy development and use and to foster AI-based innovation.[27] We believe policymakers should adopt regulatory frameworks for AI that help promote an ecosystem of trust. These frameworks would build on and reinforce many valuable private-sector efforts now underway and would ensure companies have incentives to adopt new ethical AI initiatives in the future. Frameworks would provide a regulatory foundation for these efforts, one that builds trust in AI by articulating a set of core obligations based on widely shared goals and values, support for future standards that can help companies measure and document their compliance, and targeted enforcement measures to promote compliance.

Our proposal is based on institutionalizing a pro-innovation "Trustworthy by Design" regulatory framework. Drawing from risk-based models in the fields of cybersecurity and privacy, the framework would promote trust, accountability, and transparency while also giving organizations broad flexibility to innovate. The framework would be supported by a series of enabling measures. Chief among these is a call for policymakers to work toward greater harmonization and interoperability of AI regulatory regimes across jurisdictions. In addition, policymakers should support global standards and best practices in trustworthy AI to promote access to government-held data that may be useful for AI training or analysis, and to monitor the application of existing liability rules to AI before adopting new ones. Taken together, Workday believes these proposals offer the best hope for promoting the public trust that is so vital to unlocking AI innovation.

At Workday, we understand the massive potential of AI depends on its trustworthy development and use. Smart public policy, based on well-established and widely accepted principles and compatible with future standards and industry best practices, offers the best hope for promoting the public trust necessary to spur AI innovation.

Our AI Policy Advocacy

In the United States, Workday provided input to the Office of Management and Budget on its draft AI guidance[28] and to the National Institute of Standards and Technology (NIST) on AI initiatives on standards development[29] and explainability.[30] Workday has encouraged bodies such as NIST to support development of a federal approach embracing trustworthy AI and drive consensus-based best practices to underpin future standards and regulatory requirements.

Workday led a multistakeholder effort with the Senate Committee on Commerce, Science, and Transportation; the House of Representatives Committee on Science, Space, and Technology; and the House Committee on Appropriations, resulting in the inclusion of language calling for such a risk management framework in several legislative vehicles, including the National Defense Authorization Act (NDAA) for Fiscal Year 2021.[31] Workday also collaborated with the Bipartisan Policy Center and expert stakeholders on a national AI strategy, resulting in a bipartisan House resolution (H.Con. Res.116) by Representatives Robin Kelly (D-IL) and Will Hurd (R-TX) that includes a similar call for a voluntary framework.[32] Workday endorsed this resolution, along with Senator Maria Cantwell's FUTURE of Artificial Intelligence Act of 2020 (S. 3771)[33] and Representatives Eddie Bernice Johnson (D-TX) and Frank Lucas's (R-OK) National Artificial Intelligence Initiative Act of 2020 (H.R. 6216).[34]

In the European Union, Workday actively participated in the work led by the High-Level Expert Group on AI, delivering input on the "Assessment List for Trustworthy AI"[35] and participating in a smaller in-depth pilot, which allowed us to provide granular feedback in the context of real-world features and controls. Workday commented on the European Commission's February 2020 whitepaper and subsequent "Inception Impact Assessment" for AI legislation[36] and, through engagement with the European Parliament, offered input on the Parliament's various AI reports.

[27] Workday, "Public Policy: Advancing Our Values Through Policy"; accessed December 2020.

[28] In its final guidance released in November, 2020, the Office of Management and Budget incorporated a number of Workday suggestions, including a voluntary risk management framework as a highlighted potential non-regulatory approach: "Comments on the Office of Management and Budget's Guidance for Regulation of Artificial Intelligence Applications"; March 13, 2020.

[29] Workday, "Response to the National Institute of Standards and Technology's Request for Information on Artificial Intelligence Standards"; June 10, 2019.

[30] Workday, "Response to the National Institute of Standards and Technology Call for Comments on Four Principles of Explainable Artificial Intelligence"; October 15, 2020.

[31] 116th Congress, H.R. 6395, National Defense Authorization Act for Fiscal Year 2021; introduced March 26, 2020.

[32] 116th Congress, H.Con.Res. 116, Expressing the Sense of Congress with Respect to the Principles That Should Guide the National Artificial Intelligence Strategy of the United States; introduced September 16, 2020.

[33] 116th Congress, S. 3771, FUTURE of Artificial Intelligence Act of 2020; introduced May 20, 2020.

[34] U.S. House of Representatives Committee on Science, Space, and Technology, H.R. 6216, Endorsements, National Artificial Intelligence Initiative Act of 2020; March 12, 2020.

[35] Workday, "Comments on the European Commission High-Level Expert Group on AI Draft Ethics Guidelines for Trustworthy AI"; January 31, 2019.

[36] Workday, "Response to European Commission Consultation on the White Paper on Artificial Intelligence"; June 14, 2020; and Workday, "Comments on the European Commission Inception Impact Assessment for a Proposal for a Legal Act of the European Parliament and the Council Laying Down Requirements for Artificial Intelligence"; September 10, 2020.

II. "Trustworthy by Design" Regulatory Framework

A common challenge policymakers face in considering whether and how to regulate AI is that AI can be used in an almost infinite variety of scenarios, each of which may raise vastly different types and degrees of risk.[37] An AI system used to navigate traffic in an autonomous vehicle, for instance, raises very different types of risk than an AI system that informs credit scores, and very different degrees of risk than one that makes restaurant recommendations. Any regulation that seeks to promote AI ethics and trust across these and millions of other scenarios in a prescriptive, one-size-fits-all manner will be unworkable.

To avoid this, the "Trustworthy by Design" regulatory framework set out below is risk based. It gives AI producers flexibility to adopt principles and policies that are most appropriate for their businesses, tailored to the type and degree of risk their AI systems present.[38] This ensures that low-risk AI systems and applications are not saddled with obligations that are only appropriate for higher-risk scenarios.[39] At the same time, the proposed framework also requires all AI producers to ensure—and be able to demonstrate—that they are living up to their commitments. It also envisions that future standards and industry best practices will play a key role in enabling organizations to make this showing.

The overall goal of the framework is to require AI producers to commit to trustworthy AI, to be transparent with users about how they are meeting those commitments, and to be accountable for the impacts of AI systems—while at the same time fully accommodating the many different types and uses of AI. The framework will be effective in promoting an ethical and human-centric approach to AI development that instills trust, while also giving companies flexibility and strong incentives to innovate and commercialize AI technologies in responsible ways.

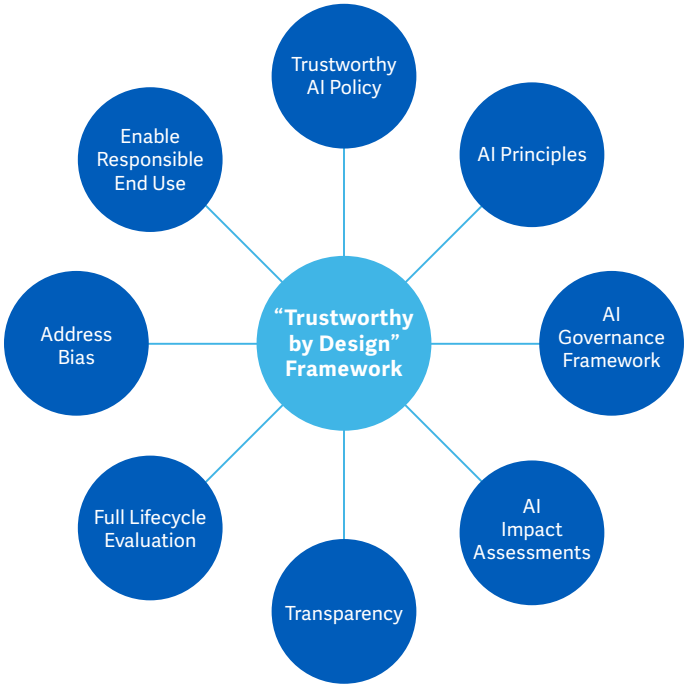
The policy proposals do not seek to answer every question or fill in every detail. That work is vital but must be done in an open dialogue with interested stakeholders, including civil society organizations, researchers, academia, and industry. For example, as discussed in detail in section III, the ongoing conversations on standards development will mature and feed into and guide regulatory requirements.

[37] At the outset, it might be asked what is this AI that we propose be regulated? Generally, "artificial intelligence" is an umbrella term that refers to technologies that tackle problems that humans have typically been good at solving and computers have traditionally not been. This includes ML, a subdiscipline of AI, which applies algorithms to massive amounts of data to recognize patterns and predict or infer insights or answers and improve automatically through experience. The Organisation for Economic Co-operation and Development (OECD) defines an AI system as a "machine-based system that can, for a given set of human-defined objectives, make predictions, recommendations, or decisions influencing real or virtual environments," and that is "designed to operate with varying levels of autonomy" (OECD, "Recommendation of the Council on Artificial Intelligence"; May 21, 2019). This definition appropriately captures the technologies that would fall within the scope of our policy proposals.

[38] AI producers are organizations that develop AI applications for use by other organizations; deployers maintain a direct relationship with individual end users and often determine software configuration and end use.

[39] AI producers would assess risk at first, though we expect systematic risk assessment frameworks to be developed over time.

The key elements of the proposed trustworthy AI regulatory framework are depicted in the graphic below and described in more detail in the sections that follow.



A. Adopt and Publish a Trustworthy AI Policy

The cornerstone of the "Trustworthy by Design" regulatory framework is the obligation for AI producers to adopt and publish a public trustworthy AI policy addressing identified core elements of ethical AI. The core elements would form a part of the producer's policy and all would need to be addressed.

Apart from the mandatory core elements, producers would have significant flexibility in deciding the details of this policy. This flexibility is essential to ensuring that companies can tailor their obligations to the unique aspects of their business, the degree of risks that their AI systems present, and the expectations of their customers and other stakeholders.[40] The rigor of this obligation would flow from a requirement for companies to publish their policies to the world and to demonstrate compliance on an ongoing basis. Similar to the requirement in many jurisdictions for companies to publish their privacy policies,[41] the obligation to adopt and publish a trustworthy AI policy would enable customers and users to compare policies between companies, which in turn would create market-based incentives for companies to adopt robust, meaningful policies.

[40] Although the proposed regulatory framework focuses primarily on organizations that develop AI systems for use by enterprise customers and other third parties, it would also apply to organizations that develop AI systems solely for internal use or only for use in their own products or services. This requirement would allow businesses in those categories to tailor their trustworthy AI policies accordingly.

[41] See, e.g., GDPR, Articles 13–14; California Business and Professions Code, section 22575(a); Electronic Code of Federal Regulations, Children's Online Privacy Protection Rule, 16 CFR 312.4(d), 2020; Notice of Privacy Practices for Protected Health Information, 45 CFR 164.520, 2020; Privacy of Consumer Financial Information, Regulation P, 12 CFR 1016, 2020.

In addition, the trustworthy AI policy would serve as a reference point for enforcement. Similar to many data protection regimes, the framework would require AI producers to comply with their published trustworthy AI policy and would allow an appropriate agency to impose sanctions for misrepresentations.

B. Adopt AI Principles

As a first step in populating a trustworthy AI policy, AI producers should publish principles setting out their trustworthy AI commitments. These guiding principles will frame and in some respects inform the remaining elements in the policy. They should also serve as committed corporate norms for all people within the organization who are engaged in designing, developing, deploying, marketing, or selling the producer's AI systems.

Organizations would have broad flexibility to tailor their principles to the unique aspects of their business and AI offerings, the scenarios in which they expect their AI to be used, and the needs of their customers. By way of example, the Workday trustworthy AI principles focus on respecting fundamental human rights, placing humans and their interests at the center of work, acting fairly and respecting the law, being transparent and accountable, protecting personal data, and delivering enterprise-ready technologies that help Workday customers tackle real-world challenges.[42] We have tailored these principles to align both with Workday core values and the sector in which we operate—finance and human resource management.[43] Organizations that operate in other sectors have customers with different needs, or that face different risk profiles and therefore might prioritize different principles.

That said, there is broad international consensus around the baseline principles that should guide trustworthy AI.[44] These include fairness, transparency, accountability, and respect for fundamental human rights. Indeed, it is difficult to imagine how an AI system could be considered ethical and trustworthy if it failed any of these requirements. Accordingly, even while granting AI producers broad leeway to articulate their own trustworthy AI principles, it would be appropriate, in our view, for the regulatory framework to require all AI producers to include at least these four principles in their own trustworthy AI policies.

[42] Barbara Cosgrove, "Workday's Commitments to Ethical AI"; May 8, 2019.

[43] Workday, "Our Values"; accessed December 2020.

[44] See, e.g., "Principles on Artificial Intelligence" adopted by the OECD, the G20, and the Global Partnership on AI (GPAI); AI HLEG, "Ethics Guidelines for Trustworthy AI"; April 8, 2019; UK Government Digital Service, "Data Ethics Framework"; September 16, 2020.

C. Adopt an AI Governance Framework

A constructive regulatory foundation should include a set of core obligations based on widely shared goals and values such as the principles described above. In order to meet these obligations, AI producers should be required to adopt and implement an internal governance framework. That framework would enable everyone within the organization who interacts in any way with the company's AI development—including senior management, product management, quality assurance, data scientists, engineers, sales and marketing, compliance, and in-house counsel—to understand how the company's AI principles are to be implemented in practice. Because the governance framework would likely constitute several different policies, operating procedures, and practical guidance, organizations would only be expected to provide a summary of their governance framework in the trustworthy AI policy itself.

Like governance frameworks in other contexts, a key purpose of the trustworthy AI governance framework would be to help drive accountability and transparency within the organization. It also would provide a basis for organizations to document their compliance to regulators and to the outside world.

Recognizing that these requirements must allow for flexibility and scale to the size and capacity of the organization, including the possibility of combining some governance roles, organizations should nonetheless include certain basic elements, in particular:

- The organization should designate a team of senior-level (including appropriate C-suite) personnel from across the company ("Governance Team") to oversee the company's AI product development lifecycle. This team would have overall responsibility and accountability for the company's compliance with its trustworthy AI principles, policies, and practice, including by designating a senior decision-maker ultimately responsible.[45]

- The organization also should designate a trustworthy AI compliance team ("Compliance Team") that would coordinate compliance and would report compliance metrics directly to the Governance Team. The Compliance Team would be responsible for developing the company's policies and practices on trustworthy AI and ensuring that these are implemented in a consistent and systematic way across the organization. This team also would be responsible for ensuring that the company's AI development policies and practices, and the AI systems themselves, comply with applicable laws or enable compliance with such laws.[46] In particular:

[45] In smaller organizations, such as start-ups, these teams might out of necessity consist of only a few people. Even in small organizations, however, the Governance Team should include at least one person with the authority to require compliance with trustworthy AI policies and practices across the organization. This requirement could be modeled on similar accountability structures in existing law. See 107th Congress, Public Law 107–204, Sarbanes-Oxley Act of 2002; July 30, 2002.

[46] For each element that follows in this list, some organizations might wish to delegate certain of these responsibilities to other teams within the organization, or to allocate these responsibilities between different teams.

- **AI impact assessments.** The Compliance Team would be responsible for ensuring that AI design and engineering teams conduct robust AI impact assessments for each AI product they develop. For AI systems that could have material detrimental impacts on individuals or their rights, the Compliance Team would be responsible for reviewing the impact assessment and for approving all proposed safeguards (discussed further in the section that follows).
- **Documentation.** The Compliance Team would be responsible for ensuring that all relevant aspects of the company's trustworthy AI compliance are properly documented, which is an effective way to promote accountability and transparency in organizational settings, and also provides AI producers with a trail should compliance questions arise. For instance, the Compliance Team should ensure that personnel document their AI impact assessments, the provenance and key characteristics of any data used to train the AI system, the results of relevant system testing including for potentially harmful bias, and how any identified issues were resolved. AI producers should also document the AI system explanations provided to customers and, where relevant, to users or other affected individuals (see below).
- **Training.** The Compliance Team would be responsible for ensuring that employees are provided with appropriate training on the company's trustworthy AI policy and all subsidiary policies and practices; for example, on how to conduct and document impact assessments. General training should be augmented by role-specific training where appropriate (for system designers, salespeople, and more).
- **Cross-company resource.** More broadly, the Compliance Team would serve as a cross-company resource when personnel, customers, or others have questions or concerns about the company's AI systems and related practices. To this end, the governance framework should establish clear procedures enabling personnel across the company to escalate concerns about the company's compliance in a safe and secure manner, without any risk of adverse consequences, and to ensure that the resolution of these issues is documented and raised to senior management as appropriate. The Compliance Team should likewise provide a mechanism for customers, partners, and other third parties to raise questions or concerns.

Although companies could develop their own governance frameworks, they would also be free to adopt appropriate third-party frameworks instead. The government of Singapore, for instance, has made significant strides in developing a model AI governance framework,[47] and elements of the European AI HLEG ethics guidelines and assessment list (mentioned earlier) could provide the foundation for such a framework as well.

[47] See Government of Singapore, "Model Artificial Intelligence Governance Framework," Second Edition; January 2020.

A regulatory framework could also encourage the development of other standardized accountability frameworks by well-suited government bodies such as NIST or by recognized standard-setting organizations such as the International Standards Organization, both of which are already actively engaged in standard-setting activities in the area of AI and ML (see section III, B).

D. Conduct AI Impact Assessments

The regulatory framework should require organizations that develop AI to conduct a separate AI impact assessment for each AI system that they plan to develop for deployment (whether through their own deployment or by others).[48] The assessment would identify potential risks to individuals or society presented by the envisioned AI system, to reasonably quantify the amount of risk and degree of potential harm, and to adopt safeguards to mitigate these risks to an acceptable level. AI producers would be required to retain the assessment for inspection by the appropriate regulator, if necessary. Again, organizations should at a minimum provide a summary description of their impact assessment methodology in their trustworthy AI policy.

Impact assessments are a well-established tool in the privacy and cybersecurity contexts. Examples include the NIST cybersecurity risk assessment framework,[49] and the data protection impact assessments required under the EU's General Protection Data Regulation (GDPR) and recommended by the U.S. FTC.[50] As with privacy, impact assessments should be required of those developing and deploying AI-based solutions. The AI impact assessment would draw relevant elements from pre-existing risk assessment frameworks, but would be tailored to the combination of safety, security, and ethical risks that AI systems may raise. While many organizations that engage in AI development are already familiar with the value of risk assessments, more research and standards work remains to be done to reach widespread consensus on how to carry out an impact assessment for AI.[51] Governments in countries including Canada have released versions of such impact assessments, as have non-governmental organizations such as the OECD.[52]

The proposed framework would permit organizations to adopt impact assessment frameworks tailored to their specific business and AI offerings, to the types of risks that are likely to arise, and to the needs of their customers. For AI systems that pose relatively few risks (for example, an AI tool that makes avatar recommendations to video game players), these assessments could be a light touch. For others (such as an AI system that helps autonomous vehicles detect obstacles), the assessment might be significantly more involved.

[48] An impact assessment would not be required for AI systems designed for research purposes only and not for use in real-world settings.

[49] National Institute of Standards and Technology, "Cybersecurity Framework"; accessed December 2020.

[50] European Union, GDPR, Article 35; Federal Trade Commission, "Privacy Impact Assessments"; accessed December 2020.

[51] Ada Lovelace Institute and DataKind UK, "Examining the Black Box"; April 29, 2020.

[52] See Government of Canada, "Algorithmic Impact Assessment (AIA)"; last modified July 28, 2020; Organization for Economic Co-operation and Development, "Algorithmic Impact Assessment"; 2019.

For instance, when undertaking AI impact assessments for higher-risk AI systems, producers should use reasonable efforts to evaluate factors such as: (1) the system's potential and reasonably foreseeable material adverse impacts, including physical safety, individual privacy, and other fundamental rights; (2) the desired level of accuracy and reliability for the system, since some AI systems designed for particularly high-risk contexts may require greater accuracy and reliability than others; and (3) the types of explanations that should be provided and to whom. When the risk assessment identifies significant and reasonably foreseeable adverse risks that cannot be reasonably mitigated through other measures, it should require some form of human oversight or involvement.

To be effective, risk assessments should be conducted early in the AI development process, ideally within the design phase, and refreshed at regular intervals. This will encourage organizations to design their systems to promote trustworthiness—which experience has shown to be more effective in mitigating risks than seeking to "bolt on" trustworthiness after the fact.[53] Because many ML and AI systems change significantly during the development process and in deployment, AI producers should have a responsibility to update their risk assessments when they know, or could reasonably foresee, that the system has changed in ways that present new or greater risks.

E. Address Potentially Harmful Bias, Including Through Diverse Teams

It is of the utmost importance that AI applications deliver business value in a fair, trustworthy manner, and any regulatory framework must proactively address potentially harmful bias in AI. One of the ways harmful bias can affect AI systems is through the categories of data that AI systems consider when making a decision—a problem that computer scientists call "feature selection."[54] Potential sources of harmful bias arising from this problem include: models using membership in a protected class directly as inputs (such as gender); considering inadequate factors to assess members of a protected class as accurately as nonmembers (for example, having few samples of a minority class compared to the samples of a majority class); or relying on factors that serve as proxies for class membership (such as zip codes as a proxy for income or race).

To address these risks, AI producers should implement procedures to identify and mitigate sources of potentially harmful bias in their AI systems, both in the AI models developed and in the data used (including to train their systems and what these systems analyze in real-world settings). Organizations should describe these procedures, at least in a summary form, in their trustworthy AI policies.

Detecting and eliminating certain forms of harmful bias will not be simple or straightforward. While some potentially harmful biases in AI systems can be fairly easily recognized and addressed, others are more insidious and harder to detect. Even reaching consensus on what constitutes harmful bias can be challenging.[55]

[53] See, e.g., David Leslie, "Understanding Artificial Intelligence Ethics and Safety: A Guide for the Responsible Design and Implementation of AI Systems in the Public Sector," pp. 44–48; The Alan Turing Institute, June 11, 2019.

[54] See, e.g., Solon Barocas and Andrew D. Selbst, "Big Data's Disparate Impact," pp. 688–691; 104 *California Law Review* 671, 2016; Feihu Yan, "How Machines Discriminate: Feature Selection"; *Medium*, May 21, 2020.

[55] There is a wealth of research on the challenges of identifying AI bias and how to address these challenges. See, e.g., Mark MacCarthy, "Fairness in Algorithmic Decision-Making"; Brookings Institution, December 6, 2019; Frederik Zuiderveen Borgesius, "Discrimination, Artificial Intelligence, and Algorithmic Decision-Making," pp. 1, 18–20; Council of Europe, 2018.

Governments will need to work with industry, standards organizations, and the research community to develop common standards for identifying, measuring, and mitigating harmful bias in a consistent way across AI systems and applications. Fortunately, the "Trustworthy by Design" regulation can help address this issue:

- First, in appropriate circumstances, a regulatory framework should require AI producers to document the provenance of AI training data and to take reasonable steps to test whether the use of these datasets may lead to unfair or discriminatory outcomes. This may include requirements to assess the extent to which such datasets are reliable and suitable for the intended purposes. For instance, training data generally should be representative of the people on whom it will be used.

- Second, organizations should document the procedures they use to test for, identify, and mitigate the effects of potentially harmful bias, both generally and with respect to each AI application they develop. This should include documentation of the mitigation steps used, including pre-processing techniques such as re-weighting, up-weighting, masking, or excluding features and their proxies.

- Finally, organizations should establish diverse teams to design and develop AI systems. It is critically important that traditionally underrepresented perspectives are included throughout the lifecycle of the AI design and development process.

F. Provide Transparency

Transparency is essential to consumer trust. For that reason, transparency obligations underpin many regulatory regimes, including data protection and consumer protection frameworks in the United States, Europe, and many other jurisdictions.[56] Transparency is also a hallmark of nearly every major proposal for trustworthy AI. People will trust AI only if they are confident that they will know when they are being affected by AI, and if they have enough information to believe that the AI system will treat them fairly. Transparency is also necessary to enable those who deploy an AI system to ensure that they are using the system as intended, and to take account of any relevant limitations in the system.

As discussed earlier, the obligation for AI producers to publish a trustworthy AI policy should help provide this transparency.[57] Beyond this, however, it is important that users are provided appropriate transparency about the *individual AI systems* they interact with. Two stakeholder groups fall at the center of transparency considerations: individuals who may interact with or otherwise be meaningfully affected by an AI system ("affected individuals") and the deployers of AI systems.

[56] See, e.g., European Union, GDPR, Article 5(1); Canadian Personal Information Protection and Electronic Documents Act, Section 5.

[57] UK Information Commissioner's Office and The Alan Turing Institute, "Explaining Decisions Made with Artificial Intelligence," p. 58; May 20, 2020.

Importantly, in the majority of instances it is the deployer who has the most direct relationship with affected individuals and is therefore the actor best suited to *provide* such transparency to affected individuals. The regulatory framework should therefore require AI producers to provide AI deployers with example notices and explanations, which deployers would communicate to affected individuals. This is particularly important because in many cases it is the deployer, not the producer, that decides how the AI system is configured and for what purpose it is used—and therefore should provide the appropriate transparency based on those decisions. In contrast, AI producers with no (or a less) direct relationship with the affected individual are ill-equipped to provide contextualized, meaningful transparency about the specific use of the AI system.

Depending on the context, producers should typically provide deployers information on: (1) the intended purpose and the acceptable use of the AI system; (2) steps on how the system can be properly deployed; (3) any known limitations in the system and any unintended or unacceptable uses; and (4) the level of human oversight, if any, that deployers should provide. For AI systems that pose potentially significant risks, these explanations also might include a high-level description of the internal workings and logic of the AI system and/or information about the accuracy, reliability, safety, or other features of the system. As is well established in other areas like privacy, this approach should give producers sufficient latitude in how they provide this information (for example, through contractual terms and more).

Regardless of how obligations are ultimately distributed, a fact- and context-intensive exercise, affected individuals should always be notified when meaningful decisions about them are being made on the basis of AI systems. This information should be provided in a clear and accessible way, in simple terms, prior to the decision occurring to ensure individuals are able to seek out alternatives if they choose.[58] It should generally include information about the factors or data that the AI considers and how it weighs these factors in reaching a prediction or decision.[59] Affected individuals could also be provided additional detail if the AI system poses relatively higher risks.

[58] The EU's GDPR, Article 12(1), requires that information provided to data subjects be provided in a "concise, transparent, intelligible and easily accessible form, using clear and plain language."

[59] Again, privacy regulation can provide a useful guide here. The EU's GDPR, for example, requires that, where automated decision-making is involved, the data controller must provide data subjects with "meaningful information about the logic involved, as well as the significance and the envisaged consequences of such processing for the data subject" [GDPR, Article 13(2)(f)].

G. Adopt Full Lifecycle Evaluation

Several of the elements set out above will require organizations that develop AI to undertake various forms of evaluation of an AI system (for example, to conduct impact assessments or testing where necessary for potentially harmful bias). One of the unique features of AI, however, is that systems may change over time.

In cases where an AI producer can reasonably foresee that the risks associated with an AI system might change over time, a regulatory framework should require organizations to monitor the operation of their system on a regular basis. This will also require occasional updates to the framework, such as refreshing prior impact assessments or adopting additional or different measures to address newly identified risks. Because the need for such evaluation will vary substantially depending on the characteristics of the AI system itself and the scenarios in which it is deployed, a constructive regulatory framework should give AI producers significant flexibility to determine the nature and frequency of such evaluations and the appropriate scenarios for tools such as statistical testing. For instance, higher-risk AI deployments might require more continuous and sophisticated evaluations than those involving lower-risk activities.

The goal should be to ensure that the AI producer's initial commitment to trustworthy AI is maintained across the full lifecycle of the AI system. The need for trust does not end when the AI-enabled product leaves the store, or when the contract with the customer is signed. People will measure the trustworthiness of AI based on their experiences, and this necessarily includes situations when the AI system is out in the marketplace.

H. Enable Responsible Use by Customers

Although the trustworthy AI elements discussed above fall first and foremost on AI producers, many of them will also necessarily involve cooperation with organizations that deploy AI systems. There are limits, of course, to how much AI producers can require of their customers, and the extent to which they can monitor or enforce their customers' compliance. Any AI regulatory framework must acknowledge these limits and should not impose obligations that would transform the commercial and collaborative relationship of AI producers and deployers today into an oversight or enforcement role better suited to regulators than to commercial partners.

In addition, many organizations that deploy AI systems today are already subject to a number of existing laws that could provide a cause of action against them for unethical or otherwise harmful deployments. Further experience is needed to determine whether additional rules for deployers are necessary, at least on a generalized, non-sector-specific basis.

That said, AI producers should be obligated to take certain steps to assist and encourage their customers to deploy AI systems safely, ethically, and responsibly, including:

- Clearly communicating the level of human oversight required and the appropriate methods for maintaining human control over the AI system.
- Disclosing key elements of their AI systems to customers, including limitations and intended uses, and providing deployers with model notices and explanations that these customers can use to provide appropriate transparency to affected individuals (such as the deployer's employees or its own customers).
- Identifying foreseeable unintended uses of AI systems that reflect materially greater risks to individuals or society and contractually restrict customers from the unintended uses.
- Making the deployer aware of the importance of safeguards where the AI system includes such measures against untrustworthy uses of the system.

Given the important role that AI deployers play in determining how AI is used, steps such as these are important to help promote accountability and trust across the full array of ways in which people interact with and may be affected by AI.

I. Enforcement

The best rules in the world will do little to promote trust in AI if they are not enforced. Although the appropriate enforcement regime for AI regulation requires further discussion, Workday believes that enforcement by an administrative agency, using administrative remedies such as injunctive relief and fines, is likely the best option. Existing enforcement models around the world offer useful places to begin. The enforcement regime under the FTC Act in the United States is one model worth exploring.[60] In the European Union, enforcement by data protection authorities is another sensible option.

Enforcing AI regulation may pose new practical challenges and require new expertise within enforcement agencies. For example, while existing anti-discrimination laws will typically apply with equal force to discrimination caused by AI systems, proving that an AI system caused unlawful discrimination poses different challenges than making the same showing with respect to human decision-makers. We encourage policymakers to ensure that their enforcement agencies are appropriately resourced to build the expertise needed to effectively investigate and enforce the "Trustworthy by Design" requirements discussed earlier.

[60] U.S. Federal Trade Commission, "A Brief Overview of the Federal Trade Commission's Investigative, Law Enforcement, and Rulemaking Authority"; October 2019.

III. Promoting an Enabling Environment for Trustworthy AI

The "Trustworthy by Design" framework seeks to encourage practices by AI producers and deployers that will promote trust in AI. Although adopting this approach is the cornerstone of a framework to build trust and avoid the emergence of an AI "trust gap," policymakers should take additional steps to create a truly enabling environment for innovation and market growth in trustworthy AI.

A. Promote Cross-Border Harmonization Between Regulatory Regimes

As policymakers promote trustworthy AI with local legislation, they should work to ensure that the laws they adopt are interoperable with corresponding rules in other major jurisdictions. Today, a lack of harmonization on trustworthy AI requirements risks creating a global patchwork of inconsistent or even contradictory rules. Conflicting compliance obligations in different jurisdictions could block cross-border research and innovation in AI and impede global trade in AI products and services. This risk is particularly great given that AI systems are likely to be more robust, safe, and reliable if they can be trained on or analyze data from multiple jurisdictions. It would be deeply regrettable if jurisdictional conflicts in the rules meant to promote trustworthy AI had the effect of making it more difficult for organizations to develop and market AI systems that are truly trustworthy.

Fortunately, we are starting from a foundation of widely shared values and even broad consensus on the core elements of trustworthy AI. As mentioned above, leaders from the United States, Europe, and other nations recently announced the Global Partnership on AI (GPAI), with the goal of advancing "the responsible and human-centric development and use of AI in a manner consistent with human rights, fundamental freedoms, and our shared democratic values, as elaborated in the OECD Recommendation on AI."[61] Many of the core principles endorsed by the GPAI (through its support of the OECD AI principles[62]) are echoed in similar statements from other organizations.[63] Moreover, two jurisdictions that are among those leading the charge on AI development and deployment—Europe and the United States—have separately published proposals that reveal a significant degree of consensus on the goals of trustworthy AI regulation.[64]

Despite this consensus on principles, however, there remains a substantial divide on approach. In Europe, for instance, the European Commission is expected to issue a proposed regulation on AI in 2021 that would impose detailed regulatory requirements on AI systems and applications deemed to be "high risk."[65]

[61] GPAI, "Joint Statement from Founding Members of the Global Partnership on Artificial Intelligence"; June 15, 2020.

[62] Organisation for Economic Co-operation and Development, "OECD Principles on AI"; accessed December 2020.

[63] See Partnership on AI, "Tenets"; accessed December 2020.

[64] European Commission, "On Artificial Intelligence—A European Approach to Excellence and Trust," pp. 10–13; February 19, 2020; U.S. Office of Management and Budget, "Guidance for Regulation of Artificial Intelligence Applications"; January 7, 2019.

[65] European Commission, "Inception Impact Assessment: Proposal for a Legal Act of the European Parliament and the Council Laying Down Requirements for Artificial Intelligence"; July 23, 2020.

Among the proposals under consideration are requirements that all high-risk AI applications be subject to both a mandatory pre-marketing conformity assessment within the EU and post-marketing surveillance by EU member state authorities.[66] All non-high-risk AI systems would be eligible to participate in a voluntary labeling scheme reflecting their adherence to EU rules. The Commission is also considering changes to EU product liability rules to make it easier for plaintiffs to sue AI suppliers.[67]

The United States so far has taken a slightly different approach. In January 2020, the White House Office of Management and Budget released draft guidance, finalized in November, that set forth 10 principles for U.S. federal agency approaches to AI. The guidance calls on agencies to focus on "narrowly tailored, evidence-based regulations that address specific and identifiable risks" that are not addressed by existing law or regulation.[68] In addition, some congressional measures have been introduced to address AI, including the National Artificial Intelligence Initiative Act of 2020, The FUTURE of AI Act, and the Hurd-Kelly resolution to create a national artificial intelligence strategy, among others. While these are all important steps, AI policy in the United States has further to go if it is to be truly interoperable with more advanced emerging regulatory approaches around the world.

To avoid a world in which innovators are subject to overlapping but contradictory regulatory regimes, the European Union and the United States should work proactively and collaboratively to find a middle ground. The governments need not adopt identical regulations—that would ignore important political, cultural, and historical differences between them. But as both move forward on AI regulation, they should make efforts to ensure that the rules they adopt are as interoperable and aligned as possible. Closer dialogue and collaboration between the European Union and United States could lead to the development of de facto global best practices for trustworthy AI grounded in common democratic values. Like other information technology, the use of AI across international borders is inevitable; it is only a question of whether regulatory approaches will be harmonized to foster innovation or remain balkanized, blunting AI's potential.

[66] European Commission, "On Artificial Intelligence–A European Approach to Excellence and Trust," pp. 10–13; February 19, 2020.

[67] European Commission, "Commission Report on Safety and Liability Implications of AI, the Internet of Things, and Robotics"; February 19, 2020.

[68] U.S. Office of Management and Budget, "Guidance for Regulation of Artificial Intelligence Applications," p. 2; November 17, 2020.

B. Support the Development of Trustworthy AI Standards

As described throughout our proposed regulatory framework, future standards on trustworthy AI will be critical to ensuring that companies across the AI economy can meaningfully comply with their obligations. Voluntary, consensus-based standards can help AI producers operationalize trustworthy AI and assess compliance, while also forming the basis for future regulatory requirements with widespread support. Standards are especially important for smaller companies that might not have the resources or in-house expertise to develop relevant policies, methodologies, and tools themselves. This work is a priority issue that will be a cornerstone of a trustworthy environment for AI and should therefore be staffed and resourced accordingly.

Numerous global standards bodies are in the process of developing standards that are relevant to trustworthy AI. The International Organization for Standardization (ISO), for instance, recently issued a technical report that analyzes the factors that can impact the trustworthiness of systems providing or using AI and provides practical guidance to businesses.[69] The Institute for Electrical and Electronics Engineers (IEEE) has published "The Ethics Certification Program for Autonomous and Intelligent Systems" that seeks to create specifications for certification and marking processes that advance transparency, accountability, and the reduction in algorithmic bias in autonomous and intelligent systems.[70] Many other standards organizations are also at work on relevant AI standards.[71]

We urge governments to support these international standardization efforts, including through active participation in standards organizations and by enabling AI producers to demonstrate their compliance with regulatory obligations through their implementation of such standards.[72] We also encourage the United States in particular to support efforts by NIST to develop consensus-driven best management practices and voluntary standards that can form the underpinnings of specific regulatory requirements. We applaud the work NIST has done to date in this area, including with its proposed principles for explainable AI[73] and other matters. We urge NIST to move forward deliberately on its standardization work on these issues, in particular in connection with developing processes for conducting AI impact assessments, helping organizations assess when AI systems should involve humans (for example, "human in the loop"), and developing standards for assessing fairness and mitigating potentially harmful bias.[74]

[69] Elizabeth Gasiorowski-Denis, "Towards a Trustworthy AI"; International Organization for Standardization, July 7, 2020.

[70] Institute for Electrical and Electronics Engineers Standards Association, "The Ethics Certification Program for Autonomous and Intelligent Systems (ECPAIS)"; accessed December 2020.

[71] See European Committee for Standardization, "CEN and CENELEC Launched a New Focus Group on Artificial Intelligence"; May 16, 2019.

[72] We also welcome public-sector statements in support of using standards to help assess and mitigate risk in AI. See U.S. Department of Homeland Security and U.S. Office of the Director of National Intelligence, "AI: Using Standards to Mitigate Risks"; June 2019.

[73] P. Jonathon Phillips, et al., "Four Principles of Explainable Artificial Intelligence"; National Institute of Standards and Technology, August 2020.

[74] Workday has been involved in extensive legislative efforts calling on NIST leadership in this area. See, e.g., U.S Congressman Anthony Gonzales (R-OH), Letter to The Honorable Walter G. Copan, Director, National Institute of Standards and Technology (December 12, 2019); 116th Congress, H.Con.Res. 116, Expressing the Sense of Congress with Respect to the Principles That Should Guide the National Artificial Intelligence Strategy of the United States (2020); 116th Congress, S. 3771, FUTURE of Artificial Intelligence Act of 2020 (2020); 116th. Congress, H.R. 6216, National Artificial Intelligence Initiative Act of 2020 (2020).

NIST has a proven track record of developing widely accepted guidelines and standards, including its "Framework for Improving Critical Infrastructure Cybersecurity" and "The NIST Privacy Framework."[75] This experience should serve it well in developing future trustworthy AI standards. As highlighted above, Workday is actively engaged in ensuring NIST has the necessary statutory authorization and resources to tackle this important work.

C. Promote Access to Data

Given the importance of data for AI, ML, and similar technologies, policymakers should accelerate efforts to expand access to data for AI product development, training, and analysis. This serves twin goals: facilitating the evaluation of AI systems, including through the use of testing where necessary, as well as promoting AI innovation writ large. The ML product development process requires continuously identifying and testing potential applications to determine which investments will yield timely and measurable business value. To the extent policymakers promote and expand access to data, this type of data-driven innovation will only increase.

Several governments have already made important progress on this front. In the United States, the Open Government Data Act of 2019 and the Executive Order on Maintaining American Leadership in Artificial Intelligence[76] are important steps toward making federal data more easily discoverable and usable for AI purposes, though more work remains to be done.[77] Similarly, the European Commission has adopted several measures to increase access to, and the usability of, public-sector and publicly funded data, including the 2019 Open Data Directive[78] and the 2018 recommendation on access to and preservation of scientific information.[79] The Commission is expected to announce further measures in the coming months to expand access to data, including through a proposal on the governance of common European data spaces, a data act, and measures to promote access to high-value public-sector datasets.[80]

Workday urges governments to accelerate and expand efforts to make data more easily available. For instance, we encourage public-sector organizations in both the EU and United States to make datasets available in standardized formats using widely adopted APIs or similar mechanisms—steps that will make data more usable and facilitate the combining of datasets from different departments or jurisdictions. In addition, we encourage governments to responsibly promote access to data that organizations can use to test AI systems for potentially harmful bias without needing to collect or store such information themselves.

[75] National Institute of Standards and Technology, "The NIST Privacy Framework: A Tool for Improving Privacy Through Enterprise Risk Management," Version 1.0; January 16, 2020; National Institute of Standards and Technology, "Framework for Improving Critical Infrastructure Cybersecurity," Version 1.1; April 16, 2018.

[76] Executive Office of the President, Order 13859, section 5(a), 84 FR 3967, Maintaining American Leadership in Artificial Intelligence; February 11, 2019. The document directs federal agencies to "improve data and model inventory documentation to enable discovery and usability" and to "prioritize improvements to access and quality of AI data and models based on the AI research community's user feedback."

[77] United States Government Accountability Office, "Open Data: Agencies Need Guidance to Establish Comprehensive Data Inventories; Information on Their Progress Is Limited"; October 8, 2020.

[78] European Commission, "Directive (EU) 2019/1024 of the European Parliament and of the Council of 20 June 2019 on Open Data and the Re-Use of Public Sector Information"; L 172/56, 2019.

[79] European Commission, "Commission Recommendation (EU) 2018/790 of 25 April 2018 on Access to and Preservation of Scientific Information"; L 134/12, 2018.

[80] See, e.g., European Commission, "Inception Impact Assessment: Legislative Framework for the Governance of Common European Data Spaces"; July 2, 2020; European Commission, "A European Strategy for Data"; COM (2020) 66 final, February 19, 2020.

D. Monitor the Application of Existing Liability Rules to AI

The growing prevalence of AI has also raised questions on whether people will be entitled to compensation when they are harmed by AI. The European Commission in particular is examining whether the region's existing product liability rules are adequate to keep consumers safe from defective AI products and services.[81] Options under consideration by the Commission include the adoption of a strict liability regime for high-risk AI, and changing rules on the burden of proof to make it easier for consumers injured by AI to prove fault.[82]

We encourage policymakers to analyze the application of existing liability rules before adopting new rules, at least until there is clear evidence that the existing rules are not fit for purpose. Given the relatively few reports of harm from AI-powered products, and the relatively little case law on the topic in either the United States or Europe, we think it is premature to consider changes to either jurisdiction's liability rules. Moreover, there are important differences between product and software liability. Product liability has traditionally focused on health and safety risks that are not posed by standalone software.

We also agree, however, that this is an area that merits careful monitoring. Ultimately, the adoption of the "Trustworthy by Design" regulatory framework—including its proposed obligations to enhance transparency and accountability—may address many of the concerns that are motivating calls for new liability rules.

IV. Conclusion

AI is one of the most promising technologies of the future. Unlocking its full potential, however, will require that it is trusted. Although policymakers, organizations, and experts across the world are working hard to promote trustworthy AI, preventing an AI "trust gap" from emerging will require governments to act—in particular by providing a regulatory foundation for trustworthy AI.

Workday believes the "Trustworthy by Design" regulatory framework combined with enabling policy measures provide a solid way forward on these efforts. We also know, however, that progress on these issues will require an inclusive and rigorous dialogue with all interested stakeholders. We look forward to being an active participant in those discussions.

[81] European Commission, "Commission Report on Safety and Liability Implications of Artificial Intelligence, the Internet of Things, and Robotics"; COM(2020) 64 final, February 19, 2020.

[82] Ibid.

workday.

+1-925-951-9000 +1-877-WORKDAY (+1-877-967-5329) Fax: +1-925-951-9001 workday.com