Portfolio Media. Inc. | 111 West 19th Street, 5th Floor | New York, NY 10011 | www.law360.com
Phone: +1 646 783 7100 | Fax: +1 646 783 7161 | customerservice@law360.com

10 Steps To Creating Trustworthy AI Applications

By **Lee Tiedrich and Lala Qadir** (April 29, 2020, 4:31 PM EDT)

Lee Tiedrich

Lala Qadir

Artificial intelligence is burgeoning in many industries, including financial services, law, agriculture, transportation, mobility, defense, energy and health care.

As a reflection of AI's tremendous promise, the Trump administration issued a "call to action" to AI experts to develop new tools to assist with COVID-19.[1] As AI developments continue, promoting "trustworthy" AI remains a high priority for policymakers and other stakeholders around the globe.

In May 2019, the Organization for Economic Cooperation and Development adopted five AI trustworthiness principles focusing on: (1) using AI to stimulate inclusive growth, sustainable development and well-being; (2) human-centered values and fairness; (3) AI transparency and explainability; (4) making AI secure, robust and safe throughout its life cycle; and (5) accountability.[2]

The G20 later adopted these principles, and many governments and organizations have embraced the same or similar principles.[3]

Given the broad adoption of AI trustworthiness principles coupled with AI's expanding use, attention also has turned to developing ways to operationalize trustworthiness. In some cases, policymakers intend to enact new legislation to promote this objective.

For example, the European Commission solicited comments on an artificial intelligence white paper that describes a proposed AI regulatory framework.[4] The white paper embraces a risk-based and proportionate approach and acknowledges that high-risk AI applications should be regulated differently than lower-risk applications.

In the United States, the National Institute of Standards and Technology has submitted a plan for developing AI standards, including for trustworthiness.[5] This plan follows President Donald Trump's executive order, signed on Feb. 11, 2019, calling for a coordinated federal AI strategy.[6] The Trump administration's draft guidance on AI regulation supports a "risk-based" and proportionate approach, but it proposes lighter regulation than the European Commission white paper.[7]

Although this guidance is not final, the Federal Trade Commission released a blog post on April 8 that discusses existing FTC guidance for AI and algorithms and outlines five principles for AI and algorithm use.[8] Other organizations, such as the Institute of Electrical and Electronics Engineers and the International Organization for Standardization, also have launched AI standards initiatives.

Best Practices for Operationalizing Trustworthy AI

While this AI policy work continues, organizations can take steps now to operationalize trustworthy AI. These steps may vary based on the intended use and associated industry. For example, steps for health care AI may differ from those used for entertainment chatbots.

Crafting policies, practices and procedures to enhance AI trustworthiness requires careful consideration of many factors, including the associated risks and intended uses of the AI and existing and proposed laws, regulations and guidance. Establishing appropriate AI practices is important, not only to reduce the risk of legal liability, but also to prevent reputational and other harm that potentially may arise from malfunctioning AI.

The following are 10 best practices that organizations can follow when preparing practices tailored for their AI applications.

1. Understand the AI Applications and Associated Risks

At the outset, organizations should understand the features, functionalities and intended uses of their AI applications and undertake risk assessments to gauge their impact. The risk assessments should evaluate, among other things, the sectors in which the applications will operate, since some sectors such as health care and transportation, typically present higher risks than others.

The risk assessment also should consider how the AI will be used. For example, the European Commission white paper acknowledges that AI used for hospital scheduling typically will not pose significant risks, although used in a high-risk sector.[9]

2. Comply With Existing Legal Requirements

Although the law with respect to AI continues to evolve, organizations should identify and review the existing laws, regulations and governmental guidance that apply to their AI applications and comply with them. For example, the European Commission white paper confirms that AI "[d]evelopers and deployers ... are already subject to European legislation on fundamental rights (e.g., data protection, privacy, nondiscrimination), consumer protection, and product safety and liability rules."[10]

In addition, the FTC AI guidance discusses applicable prior FTC guidance, outlines five principles for AI and algorithm use, and mentions that sector-specific laws may apply, such as the Fair Credit Reporting Act and the Equal Credit Reporting Act. This guidance also cautions that the manner in which data is collected for AI use potentially could give rise to liability. Furthermore, AI utilizing facial recognition technology may be subject to laws and regulations across multiple jurisdictions.

3. Examine the Evolving AI Landscape

In addition to examining current laws, regulations and governmental guidance, organizations should review pending proposals. This is particularly crucial for those legal proposals that would materially impact their AI operations should they become effective.

For instance, the European Commission white paper discusses a potential premarket conformity assessment requirement for high risk AI.[11] Developers of high-risk AI will want to understand this proposal and take into account that it may be enacted.

Some policy guidelines describe helpful steps for operationalizing trustworthy AI. For example, the European Union High Level Expert Group on Artificial Intelligence published its "Ethical Guidelines for Trustworthy AI,"[12] which includes a Trustworthy AI Assessment List that details possible steps for operationalizing trustworthiness.[13] More than 350 organizations have tested the assessment list and submitted feedback for updating the list.[14]

The growing body of governmental and nongovernmental AI materials can help organizations craft appropriate AI practices to implement trustworthiness.

4. Adopt Trustworthiness Principles

Organizations may want to adopt their own AI trustworthiness principles to guide their operations. Several companies like Microsoft, Google and IBM, have already done so. When preparing these principles, organizations can draw upon those published by others as well as existing and proposed legal requirements and guidance documents.

5. Establish Implementation Procedures

Organizations also should consider how best to address trustworthiness throughout the AI application life cycle, including the design, development, deployment and retirement phases.[15] While the precise implementation will depend upon various factors, establishing a governance framework tailored to address the applicable factors can be very effective.

The AI governance framework can establish a committee (or working group) to manage AI trustworthiness. The committee can include appropriate legal, business, technology, security and other experts from across the organization.

It can draw upon a variety of resources, including (1) applicable AI principles, (2) existing and proposed applicable laws, regulations, and government and nongovernmental guidance, (3) practices in analogous contexts, (4) the organization's existing policies and practices, including those for diversity and inclusion, and (5) the risk assessment and intended uses of the AI.

The committee can establish appropriate decision-making, documentation, training, accountability, auditing and other procedures, including procedures for staying abreast of and optimizing the resources mentioned above.

6. Implement Fairness

Most, if not all, AI trustworthiness principles incorporate some form of fairness in order to guard against AI discriminating against certain groups or establishing or perpetuating bias. This issue has garnered attention from policymakers and other stakeholders globally.

As noted above, existing laws relating to fairness already apply to some AI, and failure to comply with these laws could result in liability as well as reputational and other harm. Organizations can implement data governance systems to help manage and track information about their training data to help address, among other things, the risk of bias.

Furthermore, they can support diversity and inclusion within their AI workforce and leverage their testing, monitoring, training, accountability, explainability, feedback and redress mechanisms to enhance fairness.[16] The resources mentioned above provide further guidance for operationalizing fairness.

7. Address Security

Appropriate security also can promote trustworthy AI. For AI applications that are trained on personal or proprietary information, or that generate AI outputs that include such information, appropriate security is needed to protect the value of such information, and in some cases, to comply with applicable legal or contractual requirements.

In addition, tampering with training data, AI outputs or algorithms can distort AI operations. Appropriate security can guard against these risks too.

When developing security plans, organizations should consider the security needs at both the central processing and data storage sites, and in the case of internet-of-things applications, at remote devices where data also may be collected or received. The plans also should include appropriate mechanisms for addressing actual and suspected security breaches.

8. Foster Safety, Reliability and Accountability

Organizations also will want their AI to be safe and reliable[17] and to establish accountability for malfunctions and for violations of AI practices or applicable law. Safety and reliability, for example, can be addressed in the AI risk assessment and product design. Testing, monitoring, feedback, incident response and redress mechanisms also have important roles in operationalizing these principles and in implementing accountability.

Organizations should evaluate the extent to which these mechanisms should be employed on a pre- and post-market basis. Providing for an appropriate level of human oversight also can help promote the organization's AI objectives, including those focused on human-centered values, fairness, safety, reliability and accountability.

The FTC AI guidance also encourages organizations to consider using third-party tools and services to test their AI applications. Finally, officers and boards of directors should consider their appropriate level of oversight, taking into account their duties and responsibilities, the nature and extent of their organization's AI operations, and the risks such operations present.

9. Make AI Explainable

Organizations should evaluate how to make their AI explainable, and can draw upon the many available resources for guidance. Certain laws already require explainability in some circumstances.[18]

When addressing explainability, organizations should tailor their communications for the intended audiences, which may be customers, end users or affected individuals. The AI explanations also may include redress mechanisms for individuals who believe they are adversely affected by the AI, and in certain circumstances, existing law already requires such mechanisms.

Furthermore, organizations should consider whether notice should be given to individuals when they interact with AI instead of humans.[19] Customer agreements also can address explainability, the scope of permitted AI uses, and other relevant terms.

10. Expand Compliance

Organizations may want to expand their compliance program to help ensure they properly implement their AI practices, adhere to applicable laws and regulations, and prepare for new laws and regulations.

A well-developed compliance program can help organizations implement AI practices by integrating effective mechanisms to promote compliance, such as document creation and retention, training, auditing, and escalation procedures in the event of actual or suspected noncompliance. These efforts can help organizations maintain and promote trustworthy business practices throughout the AI product life cycle, as contemplated by their governance frameworks, and address any allegations that may arise with respect to their AI operations.

Although there is no one-size-fits-all approach, by following the best practices outlined in this article, organizations should be able to develop and implement trustworthy AI practices customized for their specific circumstances.

Lee J. Tiedrich is a partner at Covington & Burling LLP and co-chairwoman of the firm's global interdisciplinary artificial intelligence initiative.

Lala R. Qadir is an associate at the firm.

The opinions expressed are those of the author(s) and do not necessarily reflect the views of the firm, its clients, or Portfolio Media Inc., or any of its or their respective affiliates. This article is for general information purposes and is not intended to be and should not be taken as legal advice.

[1] The White House Office of Science and Technology Policy, Call to Action to the Tech Community on New Machine Readable COVID-19 Dataset (Mar. 16, 2020), https://www.whitehouse.gov/briefings-statements/call-action-tech-community-new-machine-readable-covid-19-dataset/.

[2] OECD, OECD Principles on AI, https://www.oecd.org/going-digital/ai/principles/ (last visited Apr. 14, 2020) ("OECD AI Principles"); Michelle Choe, AI Update: OECD Adopts AI Policy Guidelines, Covington & Burling LLP: Inside Privacy (May 22, 2019), https://www.insidetechmedia.com/2019/05/22/ai-update-oecd-adopts-ai-policy-guidelines/.

[3] See Recommendation of the Council on Artificial Intelligence, OECD Legal Instruments (May 21, 2019); Angel Gurría, OECD Sec'y-General, Address Before the G7 Leaders Summit on Digital Economy and Artificial Intelligence, (Aug. 26, 2019).

[4] White Paper on Artificial Intelligence: A European Approach to Excellence and Trust, Euro. Comm'n 1 (Feb. 19, 2020), https://ec.europa.eu/info/sites/info/files/commission-white-paper-artificial-intelligence-feb2020_en.pdf ("EC White Paper").

[5] Micha Nandaraj Gallo, AI Update: NIST Releases its Plan to Develop AI Standards, Covington & Burling LLP: Inside Privacy (Aug. 14, 2019), https://www.insideprivacy.com/artificial-intelligence/ai-update-nist-releases-its-plan-to-develop-ai-standards/; National Institute of Standards and Technology, U.S. Leadership in AI: A Plan for Federal Engagement in Developing Technical Standards and Related Tools (Aug. 9, 2019), https://www.nist.gov/system/files/documents/2019/08/10/ai_standards_fedengagement_plan_9aug2019.pdf ("NIST Plan").

[6] Jadzia Pierce & B.J. Altvater, AI Update: President Trump Signs Executive Order on Artificial Intelligence, Covington & Burling LLP: Inside Tech Media (Feb. 11, 2019), https://www.insidetechmedia.com/2019/02/11/ai-update-president-trump-signs-executive-order-on-artificial-intelligence/; The White House, Executive Order on Maintaining American Leadership in Artificial Intelligence (Feb. 11, 2019), https://www.whitehouse.gov/presidential-actions/executive-order-maintaining-american-leadership-artificial-intelligence/.

[7] Lee Tiedrich, AI Update: White House Issues 10 Principles for Artificial Intelligence Regulation, Covington & Burling LLP: Inside Tech Media (Jan. 14, 2020), https://www.insidetechmedia.com/2020/01/14/ai-update-white-house-issues-10-principles-for-artificial-intelligence-regulation/; Russell T. Vought, Guidance for Regulation of Artificial Intelligence Applications (Draft) (Jan. 7, 2019), https://www.whitehouse.gov/wp-content/uploads/2020/01/Draft-OMB-Memo-on-Regulation-of-AI-1-7-19.pdf.

[8] Andrew Smith, Using Artificial Intelligence and Algorithms, Federal Trade Commission: Business Blog (Apr. 8, 2020), https://www.ftc.gov/news-events/blogs/business-blog/2020/04/using-artificial-intelligence-algorithms ("FTC AI Guidance").

[9] EC White Paper at 17.

[10] EC White Paper at 10.

[11] EC White Paper at 23.

[12] HLEG Ethics Guidelines.

[13] Id. at 26–31; Lisa Peets et al., AI Update: EU High-Level Working Group Publishes Ethics Guidelines for Trustworthy AI, Covington & Burling LLP: Inside Tech Media (Apr. 9, 2019), https://www.insidetechmedia.com/2019/04/09/ai-update-eu-high-level-working-group-publishes-ethics-guidelines-for-trustworthy-ai/.

[14] EC White Paper at 9.

[15] See, e.g., NIST Plan at 8; OECD AI Principles; EC White Paper at 14, 22; HLEG Guidelines at 15.

[16] HLEG Guidelines at 17.

[17] Reliability encompasses accuracy, robustness, and other similar features.

[18] GDPR art. 15 & 22. See also FTC AI Guidance; 15 U.S.C. §1681m; U.S. Gov't Accountability Office, GAO-19-111, Financial Technology: Agencies Should Provide Clarification on Lender's Use of Alternative Data (Dec. 2018), https://www.gao.gov/assets/700/696149.pdf.

[19] EC White Paper at 20.