COSMOS
Navigate the Compliance Universe

Compliance Today - January 2024

Shawn E. Marchese (smarchese@evergreennephrology.com, linkedin.com/in/shawn-e-marchese/) is Chief Compliance Officer at Evergreen Nephrology in Austin, TX.

Nakis Urfi (nakis.urfi@gmail.com, linkedin.com/in/nurfi/) was Compliance Officer & ESG Leader at Babylon Health in Dallas, TX.

Implementing ethical AI oversight while regulations lag behind

by Shawn E. Marchese and Nakis Urfi

Imagine receiving a call from your child's school saying your child has been caught cheating on an assignment. As you sit in the school office, your child's English teacher explains that your child cheated by using ChatGPT—a popular online artificial intelligence (AI) language generator—to write a paper. When you ask the teacher how she knows, the teacher hands you a paper filled with phony quotations from books that don't exist. "Of course," you say to yourself as you read it, "the 18th-century British novelist Jane Austen did not fight at the Battle of Pearl Harbor." But the AI said she did, so into the paper it went.

This is just one example of what those who study AI call a "hallucination"—a confident but inaccurate response to visual or other data inputs. Hallucinations have gained recognition recently in so-called "generative" AI systems like ChatGPT because of the popularity and accessibility of these models online and the odd and occasionally hilarious whoppers of misinformation they produce. But hallucinations are by no means limited to generative AI and certainly not just to text chatbots. Other types of AI are susceptible to hallucinations and equally confident in their inaccurate responses.

Imagine that instead of a text-generating AI like ChatGPT making up false facts about English novelists, you are the unfortunate victim of another AI hallucination. What if the AI in a self-driving car confidently "recognizes" a stop sign as an empty night sky and speeds through an intersection you happen to be driving through? Or what if an AI designed to assist a radiologist in identifying anomalies on a scan confidently assesses what is actually a malignant tumor as benign?

While there is no universally agreed-upon definition of AI, it can be broadly defined as technologies that "enable machines to carry out highly complex tasks effectively—tasks that would require intelligence if a person were to perform them."[1] More simply put, AI enables a machine to do something that usually requires a human brain: writing a paper, driving a car, or diagnosing a disease. Of course, as we all know, confident humans can also make mistakes when doing these tasks and sometimes those mistakes can be dangerous. But thankfully, regulations and industry standards exist to mitigate the risk of such mistakes—such as licensing and vision testing for human drivers and medical education and licensing. Compliance with these regulations and standards mitigates risks, prevents costly damage, and saves lives.

But what if there were no regulations? What if instead of regulations, there were only voluntary guidelines suggesting who can drive a car or diagnose cancer—but at the end of the day, it's up to you whether you want to follow them? Many of us would not feel very safe leaving the house in a world like this; however, this is exactly

the landscape for much AI development today.

The slow progress of regulation

And yet, despite the risks, regulation to mitigate the impacts of AI—from the inconvenient to the potentially disastrous—has been slow to emerge. Nearly 70 countries around the world have adopted some form of AI policy as of September 2023, but many of these, like the Pirate Code from Disney's *Pirates of the Caribbean* movie, are "more what you'd call 'guidelines' than actual rules." Some global economic players, such as China, Brazil, and Japan, are progressing towards robust governance and regulation of AI. The EU's proposed AI Act could have a "Brussels effect," leading developers outside the EU to comply with it to streamline business operations, but debate continues whether or not this will happen.[2] Closer to home, Canada has proposed the Artificial Intelligence and Data Act (AIDA) to significantly regulate AI systems by setting standards for responsible design, development, and deployment of AI with a particular focus on "high-impact" AI systems (such as employment screening systems, biometric identification and profiling systems, and any AI system critical to health and safety).[3]

In the U.S., there are laws in place that govern the results of AI when used for its most common applications, such as Section 5 of the Federal Trade Commission (FTC) Act[4] — which prohibits "unfair or deceptive acts or practices in or affecting commerce"—and the Equal Credit Opportunity Act,[5] which prohibits discrimination against applicants for credit. But these laws exist already and do not speak directly to AI; moreover, a lack of transparency in the way some AI models arrive at conclusions (so-called "black box" AI models) can make it difficult to identify when AI is making decisions based on criteria that would be illegal for a human brain to base a decision on. The FTC has also released helpful guidance on keeping "AI claims in check"[6] and open and fair competition concerns.[7] At the state level, multiple states included AI regulations as part of larger consumer privacy laws that were passed or are going into effect in the future. Some states have proposed similar bills, while other states have proposed task forces to investigate AI.[8]

However, despite these steps in the right direction, real AI lawmaking in the U.S. is still far from a reality while AI continues to develop at an exponential pace. But there have been numerous sets of "guidelines" that can help fill the gap until proper regulation is in place. There have been no less than two significant releases of voluntary guidelines at the federal level to establish rights for users and frameworks to mitigate risk, which can aid organizations in making the right choices now while regulation is still out of reach.

Blueprint for an AI Bill of Rights

The first is the Biden administration's Blueprint for an AI Bill of Rights**,** published in October 2022, to "guide the design, use, and deployment of automated systems to protect the American public in the age of artificial intelligence."[9] Essentially a declaration of the rights of Americans in the face of rapid developments in AI, the blueprint serves in practice as a prototype design for standards for the ethical development and use of AI.

Each of the five principles of the blueprint articulates a fundamental right of individuals in the face of AI.

1. **Safe and effective systems** speak to individuals' right to be protected from *un*safe or *in*effective systems calling for pre-deployment testing, risk identification and mitigation, and monitoring and reporting to demonstrate safety and effectiveness.

2. **Algorithmic discrimination protections** call for equitable use and design of AI algorithms and systems, to ensure that individuals do not face discrimination based on race, color, ethnicity, gender identity, sexual

orientation, religion, age, national origin, disability, veteran status, genetic information, or any other classification protected by law.

3. **Data privacy** states that individuals should be protected from abusive data practices, have agency about how their data is used, and have an opportunity to consent before data is used, accessed, transferred, or deleted.

4. **Notice and explanation** provide that users should be notified in plain language when any AI system is used, and how it may impact them to ensure transparency and accountability for AI outcomes.

5. **Human alternatives, consideration, and fallback** recommend allowing individuals to opt out of AI systems and request a human alternative, where appropriate, and escalate to a human when needed to address problems that may occur as a result of the use of AI systems.

While it may sound like the stuff of science fiction—perhaps a reimagining of Isaac Asimov's "Three Laws of Robotics" with their creed that "a robot may not injure a human being"—the Blueprint for an AI Bill of Rights is intended to guide American society through the next few critical years of AI development, protecting citizens from potential threats (the extraordinary and the mundane) and ensure that technology continues to be developed and used in ways that reinforce our values as a society.

NIST AI Risk Management Framework

Another substantial set of guidelines released recently by the U.S. government is the National Institute of Standards and Technology (NIST) AI Risk Management Framework (RMF). The RMF was developed in collaboration with the AI development community over three years through workshops, reports, requests for information, and public comment periods and finally released in January 2023. The RMF's stated goal is "to better manage risks to individuals, organizations, and society associated with artificial intelligence (AI)," and it achieves this goal by recommending specific controls for organizations to use in mitigating these risks.[10]

The substance of the RMF is principally in the NIST AI RMF Playbook, which provides a roadmap of controls and activities for organizations to consider implementing to manage AI risks effectively. The playbook—available online at the NIST website[11] —outlines four specific functions for AI risk management which, when boiled down, may look familiar to those of us in the compliance profession:

- **Govern** aids organizations in designing, developing, deploying, or acquiring AI systems to foster a culture of risk management around these systems. It suggests policies, procedures, training, and accountability structures within the organization, an organizational commitment to a culture that considers and discusses risk, and even for factors like workforce diversity, equity, inclusion (DEI), and accessibility to be considered when assessing AI risks.

- **Map** aims to enhance the organization's ability to identify risks and contributing factors. It recommends mapping the contexts in which AI is developed and used, acknowledging the business value of AI in meeting organizational strategic goals, setting risk tolerance, and identifying potential controls.

- **Measure** recommends tools and tests to analyze, assess, benchmark, and monitor AI risks and related impacts. This function suggests techniques for determining appropriate metrics to monitor AI system performance, effectiveness of controls, and document risks identified in the map function.

- **Manage** provides guidance on risk treatment to decrease likelihood and negative impacts and establish mechanisms to respond to, recover from, and communicate about incidents or events. It also suggests

options for organizations to ensure continuous improvement in effectively managing AI risks.

If all of this sounds a lot like the seven elements of an effective compliance program, that's because, essentially, it is. All the same elements are there: policies and procedures, oversight, education, risk assessment and monitoring, lines of communication, enforcement, response, and correction. And this is why compliance officers are uniquely positioned with our organizations to be the leaders in implementing a framework like this with the help of risk professionals, technology, and information security leaders.

The challenge, of course, is that both frameworks just discussed—the NIST AI RMF and the Blueprint for an AI Bill of Rights—are purely voluntary. Unlike implementing an effective compliance program, there is no force of law compelling organizations to implement AI risk management, even if the AI they develop and/or use could have potentially serious negative impacts on employees, customers, or the general public. This is especially true in healthcare, where patients' lives will become more dependent on AI-interconnected systems that perform clinical decision-making support, medical imaging and clinical note-taking, predictive analytics, symptom checking, and more. Advances in these AI applications show no sign of slowing, and innovation will continue to outpace the development regulation for the foreseeable future. In the absence of regulation, it seems only a matter of time before some adverse action takes place. Organizations should take steps to voluntarily implement ethical AI oversight programs, and the time to act is *now.*

Five phases to ethical AI oversight implementation

The good news is that compliance professionals are uniquely positioned in our organizations—and have the necessary skill sets—to lay the foundations for an ethical AI oversight program. With some ingenuity and support from others in your organization who care about ethical AI, implementing ethical AI oversight can be achieved in five phases:

- Identify
- Buy-in from the top
- Define and establish controls
- Assess and remediate
- Oversee

During the "identify" phase, define your overall strategy for oversight of AI. What are your objectives? Obviously, to mitigate risk and ensure ethical use of AI, but how does that tie to the organization's strategic goals? What processes currently use AI, and how will oversight impact those processes? How will you (or whoever else is tasked with oversight) get plugged into those processes? This is also a good opportunity to identify stakeholders: it's a great idea to bring on board leaders over functions that develop or use the AI, such as operations and technology, along with subject matter experts you will want involved in the oversight, like privacy, security, clinical safety, and compliance functions. Finally, do some fact-finding across the organization to complete an AI inventory, documenting specific uses of AI; if you don't know about it, you can't oversee it. Remember to incorporate data governance and how your third-party partners are plugged into your overall ecosystem.

The second phase involves getting "buy-in from the top." But first, determine the best path to this for your organization. Consider the culture of the company: How embedded is AI into the fabric of the company? A bleeding-edge health tech startup developing AI products will have a different view of AI from a traditional hospital system that just happens to use some AI algorithms to aid in coding processes. Make sure the proposed oversight strategy resonates with company culture. Then, evaluate the tone from the top regarding ethical AI. Is

senior leadership committed to ethical use of AI? Is it mentioned in stand-ups or company mission statements and values? What about your code of conduct? If ethical AI is not specifically addressed in the code, consider whether it should be and what kind of training should be required around it. Consider whether to include reporting for ethical AI risks to the board and compliance committee. These measures will help ensure that the oversight approach has the commitment and support it needs to be successful, from the top down.

In the next phase, "define and establish controls." Determine what controls will be most meaningful for your organization: Do you want to start with a high-level set of principles (the Blueprint for an AI Bill of Rights would make a great model for this) or a more detailed set of risks and controls like the NIST AI RMF? Regardless of your approach, the controls should be responsive to specific risks or concerns identified in the first phase. The types of controls will be similar to those implemented by compliance programs approved by the U.S. Department of Health and Human Services Office of Inspector General or Centers for Medicare & Medicaid: policies and procedures, an AI ethics work plan, monitoring and audit, and education and training for employees developing and using AI. It's worth noting that in October 2022, Congress passed a law requiring training for federal employees on AI's benefits and risks, which the Office of Management and Budget is now developing.[12] If the federal government is training its employees on this topic, it's likely only a matter of time before government contractors will be expected to train on these topics as well.

Once controls are established, it's time for phase four: "assess and remediate." Risk assessments, monitoring, and auditing are powerful tools for ethical oversight of AI, just as they are for compliance oversight, and the NIST AI RMF includes specific recommendations for monitoring AI for potential issues. Risk assessments should not only consider obvious safety and operational impacts but also reputational impacts and more. Monitoring should be incorporated into the Secure Systems Development Life Cycle[13] for organizations developing AI or operational processes utilizing AI. Auditing can take many forms, from algorithm audits that look at AI outputs to make sure algorithms perform as intended to audits that identify unintentional AI bias to enterprise audits that examine whether AI is performing to the organization's expectations and achieving its intended business goals. Monitoring and auditing are also recommended to address specific risks, such as privacy and security.

Finally, it's time to put your ethical AI oversight program into action and "oversee" AI functions across the organization by establishing an appropriate mechanism for oversight. An AI ethics committee can enable visibility of AI development and use across the organization and facilitate reporting to the board and its relevant subcommittees. However, a formal ethics oversight committee may not be the best fit for every organization: some organizations may benefit from a less formal ethics council authorized to evaluate risks and set policy or even an informal ethics forum that answers employee questions and makes recommendations. AI ethics can also be part of compliance oversight committees and activities if that is the right fit for the organization. Whichever oversight body model is chosen, it's strongly advised to assemble a multidisciplinary team of stakeholders and advisors: invite technology and operations leaders, compliance, privacy, risk management, information security, legal, and even environmental, social, and governance (ESG) and DEI representatives—at the right level to make decisions and influence organizational leadership—to participate. This will help ensure that the oversight body has visibility to all potential AI processes and can make recommendations from a comprehensive web of perspectives and standards.

Conclusion

AI is everywhere, and AI usage will only increase in the years to come. Regulations are on lawmakers' radar, but they are coming slowly; innovation happens faster every day. And while new AI systems have the potential to do great good, if used or developed unethically, they can also cause great harm to human beings in ways and at scales that cannot be fully anticipated.

This is not cause for alarm, but it *is* a reason to take action to start implementing oversight *now.* Currently, there are voluntary guidelines available, like the NIST AI RMF and the Blueprint for an AI Bill of Rights, as well as laws and regulations in other countries that provide a guide to designing paths to the ethical development and use of AI. And because the approaches are similar, compliance professionals already have the skill sets needed to implement oversight effectively. Ethical AI oversight can be implemented and set up for success with the right approach—such as the five-phased approach previously recommended—and the right mix of supporting stakeholders.

Compliance professionals are natural leaders within their organizations for this task, and there are many clear benefits to establishing ethical AI oversight in your organization now. Support the strategic goals of your organization via the use of AI. Build trust with internal stakeholders and reinforce your reputation as a collaborator, a problem-solver, and a solution-finder. Build trust with external stakeholders as well and offer competitive advantages with opportunities for brand building and other ethical, quality, compliance, and "good citizen" initiatives like ESG or DEI. Help your organization gain assurance that it is prepared for regulations when they are enacted. And, of course, you'll sleep better at night.

Establishing ethical AI oversight has so many benefits that the real question isn't *why.* The real question is, *why not?*

Takeaways

- Regulation of the development and impacts of artificial intelligence (AI) in healthcare is years away, and advancements in AI are coming faster every day.
- In the absence of regulation, U.S. agencies have released voluntary guidelines for ethical AI oversight and managing AI risks.
- Compliance professionals' experiences and oversight make them uniquely qualified to lead the charge in implementing ethical AI oversight programs.
- An ethical AI oversight program can be established at any organization by following a simple five-phased approach.
- There are many benefits of implementing AI oversight now, and it can support strategic business initiatives as well as "good citizen" initiatives.

1 The AHSN Network, Department of Health & Social Care, and National Health Service, *Accelerating Artificial Intelligence in health and care: results from a state of the nation survey,* Autumn 2018, https://wessexahsn.org.uk/img/news/AHSN%20Network%20AI%20Report-1536078823.pdf.
2 Alex Engler, "The EU AI Act will have global impact, but a limited Brussels Effect," Brookings Institution, June 8, 2022, https://www.brookings.edu/articles/the-eu-ai-act-will-have-global-impact-but-a-limited-brussels-effect/.
3 For more information about Canada's proposed law, see "The Artificial Intelligence and Data Act (AIDA) – Companion document," Government of Canada, modified March 13, 2023, https://ised-isde.canada.ca/site/innovation-better-canada/en/artificial-intelligence-and-data-act-aida-companion-document.
4 Federal Trade Commission Act, Section 5: Unfair or Deceptive Acts or Practices, *Consumer Compliance Handbook*, accessed November 17, 2023, https://www.federalreserve.gov/boarddocs/supmanual/cch/200806/ftca.pdf.

5 Equal Credit Opportunity Act, 76 Fed. Reg. 41,590 (July 15, 2011), https://www.govinfo.gov/content/pkg/FR-2011-07-15/pdf/2011-17585.pdf.
6 Michael Atleson, "Keep your AI claims in check," Business Blog, Federal Trade Commission, February 27, 2023, https://www.ftc.gov/business-guidance/blog/2023/02/keep-your-ai-claims-check.
7 Staff in the Bureau of Competition & Office of Technology, "Generative AI Raises Competition Concerns," Technology Blog, Federal Trade Commission, January 29, 2023, https://www.ftc.gov/policy/advocacy-research/tech-at-ftc/2023/06/generative-ai-raises-competition-concerns.
8 Katrina Zhu, "The State of State AI Laws: 2023," Electronic Privacy Information Center, August 3, 2023, https://epic.org/the-state-of-state-ai-laws-2023/.
9 "Blueprint for an AI Bill of Rights," The White House, accessed November 9, 2023, https://www.whitehouse.gov/ostp/ai-bill-of-rights/.
10 National Institute of Standards and Technology, "AI Risk Management Framework," accessed November 9, 2023, https://www.nist.gov/itl/ai-risk-management-framework.
11 National Institute of Standards and Technology, "NIST AI RMF Playbook," accessed November 9, 2023, https://airc.nist.gov/AI_RMF_Knowledge_Base/Playbook.
12 Artificial Intelligence Training for the Acquisition Workforce Act or the AI Training Act, Pub. L. No. 117-207 (2022), https://www.congress.gov/bill/117th-congress/senate-bill/2551.
13 National Institute of Standards and Technology, "Secure Software Development Framework (SSDF)," updated January 10, 2023, https://csrc.nist.gov/projects/ssdf.