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.