ALM | LAW.COM

New York Law Journal

Increasing Reliance on AI Is Risky and the EU Attempts To Catch Up

By David Owen and Kenneth Ritz

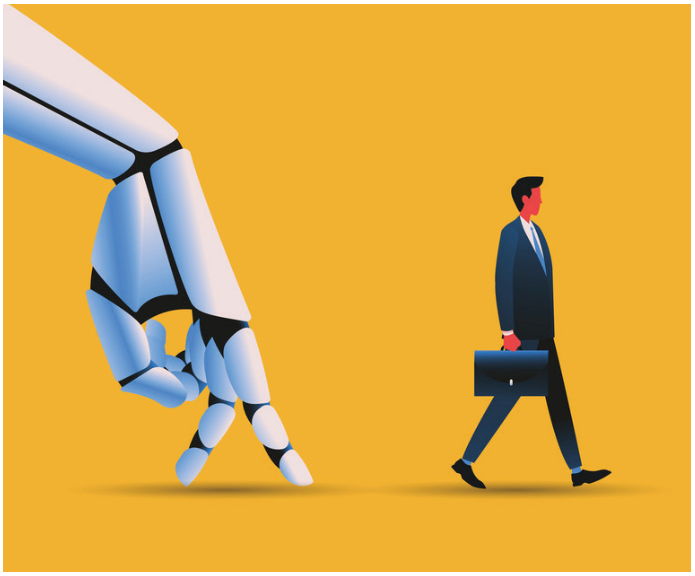
March 3, 2025

Artificial intelligence (AI) is rapidly changing the world, with seemingly limitless potential to streamline and improve productivity and our daily lives. Massive investment continues to spur AI innovation across all industries, as companies and governments race to harness this transformational new power of automated analysis and decision-making.

AI bots now routinely act in the real world based on their own independent calculations, and can learn to perform many of the higher order tasks, from driving a car to doing taxes. Lawyers use them to draft documents. Doctors use them to diagnose patients.

Chinese authorities reportedly use them to score the 'social credit' of citizens. However, we also know that AI can be error-prone and highly mysterious in its internal operations. Reported AI challenges and risks range from relatively harmless inconveniences to the potentially apocalyptic.

Despite its impressive power and risk, the AI revolution has proceeded with essentially no legal rules or guidance for developers and users. Neither the rapidly growing industry nor lawmakers had offered any substantive framework or guidance to follow, until just recently when the EU adopted the Artificial Intelligence Act of the European Union (AI Act or Act), considered the world's first comprehensive regulatory framework on AI.



Credit: Xeniya Udod Femagora/Adobe Stock

Robot fingers following businessman. Artifical intelligence, future technologies funny concept. Vector illustration

The AI Act, which will go into effect in 2026, undertakes to govern the development and use of AI in the EU, using a risk-based approach in applying prospective new rules to different types of AI depending on the risk each presents.

AI Poses Unique Risks and Challenges

AI errors can occur through glitches in the technology itself as well as by human error or even malice by third parties, and identifying the source of AI-related problems can be mystifying from a technical standpoint. Machine-made decisions and acts can present unique risks and challenges compared to those that are man-made.

Needless to say, AI bots have no ‘common sense’ in what they do, or remorse when they make mistakes. With AI bots controlling dangerous equipment and making decisions about everything from healthcare treatments to public policy, some mistakes with unfortunate and/or inexplicable consequences will be inevitable.

A persistent source of trouble has been “AI hallucinations,” which are unexplained phenomena where bots invent fictitious and/or nonsensical output. In addition to the comical examples of elaborately invented citations to legal cases that simply do not exist, AI hallucinations can take many forms, leading to incorrect predictions, false positives, false negatives, factual errors, contradictions, and outright fabrications.

Although their origins in each case are likely to be a mystery, hallucinations can arise from any incomplete, biased, or otherwise flawed training data. They also can happen because of the inherent issue of lack of ‘grounding’, meaning an AI model can struggle to comprehend real-world knowledge, physical properties, or factual information that real people easily understand.

AI output can also be polluted by a third party bad actor, including by “data poisoning” campaigns that intentionally mix deceptive or harmful data into AI models to alter the decision-making or predictive capabilities.

In most cases, ordinary AI errors may be no more troublesome than errors from any incorrectly calibrated automated process. At the lower end of AI snarls, McDonald’s recently aborted its AI drive-thru system after repeatedly (and egregiously) misunderstanding customer orders.

Other recent examples have included a delivery company that suspended an AI component after it swore at customers, and a user of a Chevrolet customer service chatbot who instructed the chatbot to agree to all requests, ultimately leading to the chatbot agreeing to sell the customer a Chevy Tahoe for one dollar. But misfiring AI has produced more pointed instances of harmful output, as well.

To leverage its outreach, the National Eating Disorders Association (NEDA) offered a chatbot that quickly demonstrated the risks of an overly free-thinking electronic advisor. In that case, the NEDA bot went far off-script in advising some users that they should lose weight, count calories, and scrutinize their body fat measurements.

Other AI tools are alleged to have caused discriminatory hiring practices, as AI algorithms improperly applied gender, age, or racial biases in filtering candidates for employment. As a result, the EEOC has begun discouraging the use of faulty AI tools by employers and software providers, including Workday.

Major players in AI have discussed possible outcomes that should give all potential regulators cause for concern. A recent Meta policy document titled “Frontier AI Framework” notes the potential for hypothetical AI catastrophes that humans may be incapable of preventing, including AI that can penetrate any corporate or government computer network, resulting in “large scale, devastating, and potentially irreversible harmful impacts on humanity.”

For AI failures large and small, it may be very difficult or impossible to pinpoint the source of the problem and differentiate a mere hallucination caused by, say, incomplete training data, from something more nefarious. Efforts to ‘retrain’ bots that have gone wrong are themselves subject to a variety of uncertainties.

Questions are also raised regarding the allocation of responsibility among those that develop AI models, the companies that offer such models to the consuming public, and other stakeholders. In other words, what obligations do various operators in the AI efficiency chain owe to the end user and society at large, and how can harm caused by breaches of such obligations be addressed legally?

The EU Artificial Intelligent Act

In an effort to establish a prospective legal and regulatory framework around all these important new technologies, the EU has recently adopted the AI Act, which purports to go into effect on Aug. 2, 2026.

According to the Act, "[t]he purpose of this regulation is to improve the functioning of the internal market by laying down a uniform legal framework in particular for the development, the placing on the market, the putting into service and the use of artificial intelligence systems (AI systems) in the union, . . . while ensuring a high level of protection of health, safety, fundamental rights as enshrined in the Charter of Fundamental Rights of the European Union (the 'Charter'), including democracy, the rule of law and environmental protection, to protect against the harmful effects of AI systems in the Union, and to support innovation."

Under the AI Act, some AI uses are prohibited outright, while others are subject to varying degrees of governance, management, and transparency requirements. The banned AI practices deemed to pose an unacceptable risk include real-time biometric identification in public spaces, social scoring systems, and manipulative technologies.

The AI Act also categorizes the different stakeholders involved in the AI system lifecycle as providers, deployers, importers, distributors, and product manufacturers. Under the Act, Providers develop and market AI systems and general-purpose AI models; deployers use AI systems within their operations; importers bring to the EU market AI systems of an entity established outside of the EU; distributors are entities other than providers or importers that make AI systems available in the market; and product manufacturers place on the market an AI system together with their own product under their own name.

The Act then imposes different compliance obligations to each role. Importantly, the Act has an extraterritorial reach, applying to entities outside of the EU if the outcomes of their AI systems are used in the EU. The Act also imposes fines for non-compliance, calculated as the higher of a percentage of the offender's global annual turnover or a certain fixed amount.

The AI Act sets out to establish a global standard for AI regulation much like the EU's General Data Protection Regulation (GDPR) did for data privacy. Just as the GDPR inspired other jurisdictions to adopt similar data privacy regulations, it is expected that the AI Act will guide the future of AI regimes around the world. Whether the US will adopt a similar nationwide framework is the subject of well-founded skepticism.

For instance, the US still lacks a comprehensive federal approach to data privacy, leaving matters in the hands of state legislatures, and a similar patchwork of state regulations is emerging in the AI space. In 2024, nearly 700 separate proposals for AI-related bills were considered across 45 states, and around 20 percent of those bills were enacted into law.

Federally, it appears that any AI regulation will take a lighter approach than the AI Act, as most of the 120 AI bills being considered by the US Congress at the close of 2024 featured voluntary guidelines and best practices, rather than strict mandates. While this new legislative and regulatory landscape spins into shape, basic principles of legal responsibility and liability remain unsettled.

David Owen *is a partner at Cahill Gordon & Reindel.* **Kenneth Ritz** *is of counsel with the firm.*