Five Things Lawyers Need to Know About AI

By Aaina Agarwal, Patrick Hall, Sara Jordan, Brenda Leong
September 2021

Note: This article is part of a larger series focused on managing the risks of artificial intelligence (AI) and analytics, tailored toward legal and privacy personnel. The series is a joint collaboration between bnh.ai, a boutique law firm specializing in AI and analytics, and the Future of Privacy Forum, a non-profit focusing on data governance for emerging technologies.

Behind all the hype, AI is an early-stage, high-risk technology that creates complex grounds for discrimination while also posing privacy, security, and other liability concerns. Given recent EU proposals and FTC guidance, AI is fast becoming a major topic of concern for lawyers. Because AI has the potential to transform industries and entire markets, those at the cutting edge of legal practice are naturally bullish about the opportunity to help their clients capture its economic value. Yet to act effectively as counsel, lawyers must also be vigilant of the very real challenges of AI. Lawyers are trained to respond to risks that threaten the market position or operating capital of their clients. However, when it comes to AI, it can be difficult for lawyers to provide the best guidance without some basic technical knowledge. This article shares some key insights from our shared experiences to help lawyers feel more at ease responding to AI questions when they arise.

I. AI Is Probabilistic, Complex, and Dynamic

There are many different types of AI, but over the past few decades, machine learning (ML) has become the dominant paradigm.[1] ML algorithms identify patterns in recorded data and apply those patterns to new data to try to make accurate decisions. This means that ML-based decisions are probabilistic in nature. Even if an ML system could be perfectly designed and implemented, it is statistically certain that at some point it will produce a wrong result. All ML systems incorporate probabilistic statistics, and those systems can make incorrect classifications, recommendations, or other outputs.

ML systems are also fantastically complex. Contemporary ML systems can learn billions or more rules from data and apply those rules on a myriad of interacting data inputs to arrive at an

[1] Commentators have often used the image of Russian nesting (Matryoshka) dolls to illustrate these relationships: AI includes machine learning, and machine learning, in turn, includes deep learning. Machine learning and deep learning have risen to the forefront of commercial adoption of AI in applications areas such as fraud detection, e-commerce, and computer vision. *See, e.g.*, The Definitive Glossary of Higher Mathematical Jargon, MATH VAULT (last accessed Mar. 4, 2021), https://mathvault.ca/math-glossary/#algo; Eda Kavlakoglu, AI vs. Machine Learning vs. Deep Learning vs. Neural Networks: What's the Difference?, IBM BLOG (May 27, 2020), https://www.ibm.com/cloud/blog/ai-vs-machine-learning-vs-deep-learning-vs-neural-networks.

output recommendation. Embed that billion-rule ML system into an already-complex enterprise software application and even the most skilled engineers can lose track of precisely how the system works. To make matters worse, ML systems decay over time, losing their use-case fitness based on their initial training data. Most ML systems are trained on a snapshot of a dynamic world as represented by a static training dataset. When events in the real world drift, change, or crash (as in the case of COVID-19) away from the patterns reflected by that training dataset, ML systems are likely to become wrong more frequently and cause issues that require legal and technical attention. Even in the moment of the "snapshot," there are other qualifiers for the reliability, effectiveness, and appropriateness of training data. How it's collected, processed, and labeled all bear on whether it is sufficient to inform an AI system in a way fit for a given application or population.

While all this may sound intimidating, an existing regulatory framework addresses many of these basic performance risks. Large financial institutions have been deploying complex decision-making models for decades, and the Federal Reserve's model risk management guidance (SR 11-7) lays out specific process and technical controls that are a useful starting point for handling the probabilistic, complex, and dynamic characteristics of AI systems. Most commercial AI projects would benefit from some aspect of model risk management, whether it's being monitored by federal regulators or not. Lawyers at firms and in-house alike, who find themselves needing to consider AI-based systems, would do well to understand options and best practices for model risk management, starting with understanding and generalizing the guidance offered by SR 11-7.

II. Make Transparency an Actionable Priority

Immense complexity and unavoidable statistical probabilities in ML systems makes transparency a difficult task. Alas, parties deploying—and thereby profiting from—AI can nonetheless be held liable for issues relating to a lack of transparency. Governance frameworks should include steps to promote transparency, whether preemptively or as required by industry- or jurisdiction-specific regulations. For example, the Equal Credit Opportunity Act (ECOA) and the Fair Credit Reporting Act (FCRA) mandate customer-level explanations known as "adverse action notices" for automated decisions in the consumer finance space. These laws set an example for the content and timing of notifications relating to AI decisions that could adversely affect customers, as well as establish the terms of an appeals process against those decisions. Explanations that include a logical consumer recourse process dramatically decrease risks associated with AI-based products and help prepare organizations for future AI transparency requirements. New laws, like the California Privacy Rights Act (CPRA) and the proposed EU AI rules for high-risk AI systems, will likely require high levels of transparency, even for applications outside of financial services.

Some AI system decisions may be sufficiently interpretable to nontechnical stakeholders today, like the written adverse action notices mentioned above, in which reasons for certain decisions

are spelled out in plain English to consumers. But oftentimes the more realistic goal for an AI system is to be explainable to its operators and direct overseers.[2]

The import of a system that's not fully understood by its operators is that it is much harder to identify and sufficiently mitigate risks. One of the best strategies for promoting transparency, particularly in light of the challenges around "black-box" systems that are unfortunately common in the US today, is to rigorously pursue best practices with respect to AI system documentation. This is good news for lawyers who are adept in the skill and attention to detail that is required to institute and enforce such documentation practices. Standardized documentation of AI systems, with emphasis on development, measurement, and testing processes, is crucial to enable ongoing and effective governance of AI systems. Attorneys can help by creating templates for such documentation and by assuring that documented technology and development processes are legally defensible.

III. Bias is a Major Problem—But Not the Only Problem

Algorithmic bias can generally be thought of as outputs of an AI system that exhibits an unjustified differential treatment between two groups. AI systems learn from data, including its biases, and can perpetuate that bias on a massive scale. The racism, sexism, ageism, and other biases that permeate our culture also permeate the data collected about us and in turn the AI systems that are trained on that data.

On a conceptual level, it is important to note that although algorithmic bias often reflects unlawful discrimination, it does not constitute unlawful discrimination per se. Bias also includes the broader category of unfair or unexpected inequitable outcomes. While these may not amount to illegal discrimination of protected classes, they may still be problematic for organizations, leading to other types of liability or significant reputational damage. And unlawful algorithmic bias puts companies at risk of serious liability under cross-jurisdictional anti-discrimination laws.[3] This highlights the need for organizations to adopt methods that test for and mitigate bias on the basis of legal precedent.

---

[2] In recent work by the National Institute for Standards and Technology (NIST), interpretation is defined as a high-level, meaningful mental representation that contextualizes a stimulus and leverages human background knowledge. An interpretable AI system should provide users with a description of what a data point or model output means. An explanation is a low-level, detailed mental representation that seeks to describe some complex process. An AI system explanation is a description of how some system mechanism or output came to be. *See* David A. Broniatowski, Psychological Foundations of Explainability and Interpretability in Artificial Intelligence (2021), https://tsapps.nist.gov/publication/get_pdf.cfm?pub_id=931426.

[3] For example, The Equal Credit Opportunity Act (ECOA), The Fair Credit Reporting Act (FCRA), The Fair Housing Act (FHA), and regulatory guidance, such as the Interagency Guidance on Model Risk Management (Federal Reserve Board, SR Letter 11–7). The EU Consumer Credit Directive, Guidance on Annual Percentage Rates (APR), and General Data Protection Regulation (GDPR) serve to provide similar protections for European consumers.

Because today's AI systems learn from data generated—in some way—by people and existing systems, there can be no unbiased AI system. If an organization is using AI systems to make decisions that could potentially be discriminatory under law, attorneys should be involved in the development process alongside data scientists. Those anti-discrimination laws, while imperfect, provide some of the clearest guidance available for AI bias problems. While data scientists might find the stipulations in those laws burdensome, the law offers some answers in a space where answers are very hard to find. Moreover, academic research and open-source software addressing algorithmic bias is often published without serious consideration of applicable laws. So, organizations should take care to ensure that their code and governance practices with respect to identifying and mitigating bias have a firm basis in applicable law.

Organizations are also at risk of over-indexing on bias while overlooking other important types of risk. Issues of data privacy, information security, product liability, and third-party risks, as well as the performance and transparency problems discussed in previous sections, are all critical risks that firms should, and eventually must, address in bringing robust AI systems to market. Is the system secure? Is the system using data without consent? Many organizations are operating AI systems without clear answers to these questions. Look for bias problems first, but don't get outflanked by privacy and security concerns or an unscrupulous third party.

IV. There Is More to AI System Performance Than Accuracy

Over decades of academic research and countless hackathons and Kaggle competitions, demonstrating accuracy on public benchmark datasets became the gold standard by which a new AI algorithm's quality is measured. ML performance contests such as the KDD Cup, Kaggle, and MLPerf have played an outsized role in setting the parameters for what constitutes "data science."[4] These contests have undoubtedly contributed to the breakneck pace of innovation in the field. But they've also led to a doubling-down on accuracy as the yardstick by which all applied data science and AI projects are measured.

In the real world, however, using accuracy to measure all AI is like using a yardstick to measure the ocean. It is woefully inadequate to capture the broad risks associated with making impactful decisions quickly and at web-scale. The industry's current conception of accuracy tells us nothing about a system's transparency, fairness, privacy, or security, in addition to presenting a limited representation of what the construction of "accuracy" itself claims to measure. In a seemingly shocking admission, forty research scientists added their names to a paper demonstrating that accuracy on test data benchmarks often does not translate to accuracy on live data.

[4] "Data science" tends to refer to the practice of using data to train ML algorithms, and the phrase has become common parlance for companies implementing AI. The term dates back to 1974 (or perhaps further), coined then by the prominent Danish computer scientist Peter Naur. Data science, despite the moniker, is yet to be fully established as a distinct academic discipline.

What does this mean for attorneys? Attorneys and data scientists need to work together to create more robust ways of benchmarking AI performance that focus on real-world performance and harm. While AI performance and legality will not always be the same, both professions can revise current thinking to imagine performance beyond high scores for accuracy on benchmark datasets.

V. The Hard Work Is Just Beginning

Unfortunately at this stage of industry and development, there are few professional standards for AI practitioners. Although AI has been the subject of academic research since at least the 1950s, and it has been used commercially for decades in financial services, telecommunications, and e-commerce, AI is still in its infancy throughout the broader economy. This too presents an opportunity for lawyers. Your organization probably needs AI documentation templates, policies that govern the development and use of AI, and ad hoc guidance to ensure different types of AI systems comply with existing and near-future regulations. If you're not providing this counsel, technical practitioners are likely operating in the dark when it comes to their legal obligations.

Some researchers, practitioners, journalists, activists, and even attorneys have started the work of mitigating the risks and liabilities posed by today's AI systems. Indeed, there are statistical tests to detect algorithmic discrimination and even hope for future technical wizardry to help mitigate against it. Businesses are beginning to define and implement AI principles and make serious attempts at diversity and inclusion for tech teams. And laws like ECOA, GDPR, CPRA, the proposed EU AI regulation, and others form the legal foundation for regulating AI. However, technical mitigation attempts still falter, many fledgling risk mitigations have proven ineffective, and the FTC and other regulatory agencies are still relying on general antitrust and unfair and deceptive practice (UDAP) standards to keep the worst AI offenders in line. As more organizations begin to entrust AI with high-stakes decisions, there is a reckoning on the horizon.

Author Information

Aaina Agarwal is Counsel at bnh.ai, where she works across the board on matters of business guidance and client representation. She began her career as a corporate lawyer for emerging companies at a boutique Silicon Valley law firm. She later trained in international law at NYU Law, to focus on global markets for data-driven technologies. She helped to build the AI policy team at the World Economic Forum and was a part of the founding team at the Algorithmic Justice League, which spearheads research on facial recognition technology.

Patrick Hall is the Principal Scientist and Co-Founder of bnh.ai, a DC-based law firm specializing at the intersection of AI and data analytics. Patrick also serves as visiting faculty at the George Washington University School of Business. Prior to co-founding bnh.ai, Patrick led responsible AI efforts at the high-profile machine learning software firm H2O.ai, where his work resulted in one of the world's first commercial solutions for explainable and fair machine learning.

Sara Jordan is Senior Researcher of AI and Ethics at the Future of Privacy Forum. Her profile includes privacy implications of data sharing, data and AI review boards, privacy analysis of AI/ML technologies, and analysis of the ethics challenges of AI/ML. Sara is an active member of the IEEE Global Initiative on Ethics for Autonomous and Intelligent Systems. Prior to working at FPF, Sara was faculty in the Center for Public Administration and Policy at Virginia Tech and in the Department of Politics and Public Administration at the University of Hong Kong. She is a graduate of Texas A&M University and University of South Florida.

Brenda Leong is Senior Counsel and Director of AI and Ethics at the Future of Privacy Forum. She oversees development of privacy analysis of AI and ML technologies, and manages the FPF portfolio on biometrics and digital identity, particularly facial recognition and facial analysis. She on privacy and responsible data management by partnering with stakeholders and advocates to reach practical solutions for consumer and commercial data uses. Prior to working at FPF, Brenda served in the U.S. Air Force. She is a 2014 graduate of George Mason University School of Law.

***Disclaimer:** bnh.ai leverages a unique blend of legal and technical expertise to protect and advance clients' data, analytics, and AI investments. Not all firm personnel, including named partners, are authorized to practice law.*