Sana Halwani
416-865-3733
shalwani@litigate.com

Jonathan Chen
416-865-3553
jchen@litigate.com

May 22, 2024

# AI Competence in the Courtroom: Four Things Judges Need to Understand Now About AI

As artificial intelligence continues to permeate every aspect of our lives, legal challenges involving AI will proliferate. Parts 1 to 3 in our series explored many of these potential questions. AI will create new legal problems and change the texture of old ones. As always, the judiciary, with the assistance of counsel, will assume a pivotal role in navigating this landscape.

Grappling with technology is nothing new for judges, but the combination of complexity, rapid evolution, and expected ubiquity of AI means that judges are at risk of getting it very wrong, very easily. With that in mind, we provide basic answers to four questions judges need to understand about AI before grappling with any case involving this technology.

**1. How is AI different from other sophisticated software?**

AI differs from other sophisticated software in several fundamental ways. AI is capable of learning, adapting, and performing complex tasks autonomously, distinguishing it from traditional software.

- **Learning and Adaptability**: AI systems, particularly those using machine learning, can learn from data and improve their performance over time without being explicitly programmed for each task. For instance, a machine learning model can improve its accuracy in predicting outcomes as it is exposed to more data. AI systems can adapt to new situations by retraining on new data. This adaptability allows AI to function in dynamic environments and solve complex, variable problems.

  On the other hand, traditional software follows a predefined set of instructions written by programmers and is less or not adaptable at all. It performs tasks exactly as programmed and does not improve or adapt unless explicitly updated by developers. Changes in its functionality or environment typically require manual code updates by developers.

- **Data-Driven**: AI relies heavily on data for training and

decision-making. The performance of AI models often correlates with the quality and quantity of data they are trained on. One such quality issue is data bias which we discuss below. Traditional software is not inherently data-driven. While it can process data, its functionality is more dependent on the specific code and logic defined by programmers rather than on data analysis and learning.

- **Decision-Making and Autonomy**: AI can make decisions based on data analysis and pattern recognition. It can handle unstructured data (like images, text, and voice) and make decisions that mimic human reasoning. AI systems can operate with a high degree of autonomy, performing complex tasks with minimal human intervention. Traditional software makes decisions based on fixed logic and predefined rules; it lacks the flexibility to interpret unstructured data. It requires ongoing human input and supervision, executing tasks based on specific user commands.

- **Human-Like Interaction**: AI enables more natural interactions with humans through technologies like chatbots, virtual assistants, and voice recognition systems. These systems can understand and generate human language to some extent. Traditional software interactions are typically more rigid and limited to predefined interfaces and commands, lacking the nuanced understanding of human language.

As addressed below, these differences create some of the thorny problems that judges will have to grapple with as they address cases involving AI.

## 2. What is AI bias, and why does it exist?

Just like humans, AI can be biased too. It is the human that develops and trains the artificial intelligence model. AI bias refers to artificial intelligence models that produce results which reflect human biases. These biases can in turn perpetuate historical social inequities. Take, for example, an AI recruiting tool that unintentionally favors candidates with a certain background or interest or of a specific gender. The result is a discriminatory one (even if unintended).

Bias can seep into artificial intelligence in several ways. Two common examples are through the training data and the algorithm. An AI system is only as good as the data input. Artificial intelligence models learn to make decisions based on that data. For example, generative AI models are built to

Lenczner
Slaght

generate text (which word should come next?) by relying on probabilities based on the dataset the model was trained on (which word usually comes next in this context?). If the dataset itself is incomplete (e.g. a certain variable is over or underrepresented in the dataset), skewed or outdated, then the probabilities and therefore the predictions will reflect those limitations. The algorithm employed can also be tainted by the developer who may inject personal preferences or weight certain attributes more heavily than others.

So, even with the best of intentions, the artificial intelligence model may produce a biased result that is perpetuated and amplified by someone also with the best of intentions. Awareness of this reality is imperative before evaluating any allegations relating to AI models, accepting results from AI models (e.g., through expert evidence), or incorporating artificial intelligence into legal decision making.

### 3. What is the difference between explainable and non-explainable AI, and why does it matter?

Explainability is about building trust in the artificial intelligence model. When we are using AI models to make predictions, there is a natural tendency for lawyers to ask how it came to that result. Developing tools and processes to understand that result is explainability.

Explainability is tied to the concept of responsible use of artificial intelligence. Whether used in a business or for legal matters, the artificial intelligence model should not be a "black box". ChatGPT, for example, is a black box because you do not know how it came to the conclusion that it did. As end users, we need to know that the model is competent, trustworthy, safe to use, up to date and accountable. To be accountable, the model must be understandable and able to be subject to human oversight and scrutinization. This in turn allows the user (or the judge) to determine whether the model meets, for example, background requirements (e.g., company policies, regulatory standards, or practice directions) or has been tested and validated. Explainability therefore imbues the result with reliability. Judges will have to grapple with what level of reliability is required in a given situation.

In addition, if a situation where harm is alleged to have been caused by a "faulty" AI, whether that AI is explainable may affect the ability of the Court or parties to evaluate fault. Self-driving cars are often used as the example here. If a self-driving car gets into an accident and its decision-making is impugned, how will a Court evaluate fault if the decision-making cannot be explained?

## 4. Can generative AIs lie, and can a human tell if this is happening?

Generative AI models like ChatGPT can produce outputs or answers that are incorrect or misleading. While an AI may not have the intent for such misleading content to be called "lying", the impact may be similar.

Courts have already started to grapple with such fabrications, sometimes called "hallucinations." For example, earlier this year, the Supreme Court of British Columbia issued a decision addressing a notice of application containing fabricated legal authorities that had been "hallucinated" by ChatGPT. The lawyer who included them gave evidence at the hearing that she did not know ChatGPT could generate fake authorities.

Generative AI models create text based on the statistical likelihood of word sequences. This means they can produce plausible-sounding but incorrect or nonsensical responses if the data suggests such patterns. Generative AI models lack the contextual understanding that a human might have. If a user inputs a prompt that is ambiguous or open to interpretation, the AI might generate a response that might fit the prompt but may not be factually accurate. The AI does not understand the context in the way humans do.

Determining whether an AI is providing accurate information or producing hallucinations can be extremely challenging for humans. This difficulty is compounded if the AI is not explainable (which is the case currently for most if not all iterations of generative AI models). Understanding the limitations of an AI (including the data on which it was trained); fact-checking; and consulting multiple sources not just AI sources, are three strategies that can help. But to employ any strategies, a Court will need to know if generative AI was used, whether in the context of a lawyer's brief or the facts of a case.

This instalment of our AI in the Courtroom series strives to provide judges with a basic understanding of some technical background and issues that may arise when AI is used in the courtroom. Judges do not need to become data scientists or coders to manage the use of AI in the courtroom, or to evaluate cases involving AI. It is important, however, for judges to at

Lenczner Slaght

least be aware of how the AI before them was developed, how it works, its application to the particular case, and the risks and implications. When these key issues are kept in mind, judges will be able to play their role as gatekeeper and properly assess when to ask questions, what questions to ask, and at what level of detail. Counsel would of course be wise to prepare responses in advance to effectively assist the Court.

***This is Part 4 of our 5-Part Series on AI in the Courtroom , which includes the below blogs.***

- Part 1 (Introduction) – AI in the Courtroom: The Quest for Legal Precedents
- Part 2 – Bars or Bytes? Exploring the Implications of a Track that Drake Might (or Might Not) Have Created
- Part 3 – On the Horizon: Legal Complexities Intersecting Generative AI, Class Actions, and IP Law
- Part 4 – AI Competence in the Courtroom: Four Things Judges Need to Understand Now About AI
- Part 5 – AI Here, AI There, AI Everywhere: Practical Challenges Litigating in an AI World

Lenczner Slaght