The Importance of Artificial Intelligence

Judges and court leaders must understand how Artificial Intelligence[1] (AI) works, its applications, its implications for the fact-finding process, and its risks. They should be able to answer the following questions in context:

> 1. How is AI being used in court or to inform judicial or administrative decisions?

> 2. Does the user understand the AI's strengths, limitations, and risks, such as bias?

> 3. Is the AI application (and the underlying data inputs on which it is based) authentic, relevant, reliable, and material to the issue at hand, and is its use or admission consistent with federal and state Constitutions, statutes, and the Rules of Evidence? [2]

Nine Pillars of AI

1. <u>AI Comes in Many Forms</u>: Just like there are many methods to solve a problem, there are many ways AI operates. Each AI method should be examined carefully to ensure it functions correctly in every situation in which it is used.

- **More:** Court systems must ensure the specific AI being utilized is suitable for its intended purpose. They need to review how the AI was developed, the data that trained it, and for AI that learns by itself, how its learning process was designed.

---

[1] Artificial intelligence (AI) refers to a machine-based system that can make predictions, recommendations, or decisions. AI systems use machine and human-based inputs to perceive environments, abstract such perceptions into models through automated analysis, and use model inference to formulate options.

[2] This resource draws substantially from the overview provided by An Introduction to Artificial Intelligence for Federal Judges, https://www.fjc.gov/sites/default/files/materials/47/An_Introduction_to_Artificial_Intelligence_for_Federal_Judges.pdf.

2. <u>AI Learns and Improves Over Time</u>: AI improves by repeatedly trying and learning. It is crucial to continually test it with different types of information. It also is important to recognize that AI makes mistakes, including misclassifications, biases, and "hallucinations", and is a tool that requires human oversight and review.

- **More:** Experts stress the importance of training, validating, and testing AI to help improve accuracy on an ongoing basis and also know how AI is expected to react on an ongoing basis.

3. <u>Humans Must Remain Fully Involved in AI</u>: Humans make decisions on how AI operates, what it looks for, and what is important in the data. To validate AI applications and outputs, court systems may look to individuals who can explain how the particular AI is configured, whether it is functioning accurately and help the system discern whether it was trained properly.

4. <u>AI Makes Predictions, Not Foolproof Conclusions</u>: AI can search through vast amounts of information quickly to identify patterns that might not be visible to humans. However, it only makes predictions based on data, and its conclusions are probabilistic and must be reviewed or contextualized by human experts. Further, sometimes it makes mistakes or has "hallucinations" (AI-contrived, false findings).

5. <u>The Quality and Quantity of Data Affect AI's Accuracy</u>: The better and more data AI has, the more accurate its predictions will be. Outdated or flawed data (and perhaps also data that is otherwise inadmissible, depending on the application) can lead to incorrect outcomes.

6. <u>Algorithms are the Heart of AI</u>: The effectiveness of an AI application often relies on the algorithm it uses, and it may be important to confirm the algorithm is accurate, free from improper bias, and using reliable inputs/data. Disputes may arise over the accuracy of algorithms, and companies may not want to reveal their proprietary information. In such

cases, courts might consider a special master or *in camera* submissions to assist in its determinations.

> **More:** Where AI is involved in court matters, the judge and attorneys may need to consider discovery regarding an AI's algorithm and the data on which it was trained.

7. <u>Specific AI Can Be Limited</u>: AI designed for a specific purpose might not perform well when applied to help understand a different context. Users should take this into account when determining whether AI is suitable for certain cases.

8. <u>AI is Adaptable</u>: One of AI's strengths is its ability to identify, aggregate, and interpret data in ways humans cannot. Whether this capability is beneficial depends on the context and must be tested, verified and challenged with human oversight and review.

9. <u>AI is Biased</u>: Similar to humans, AI can have biases. Users must be mindful of how these biases could influence the AI's predictions and transparency.

<u>Questions to Consider in the Evaluation of Proposed AI Evidence</u>

An understanding of AI driven tools and evidence <u>requires consideration of</u> (1) the AI inputs, and (2) also the weights that were attached and allocated to each input, (3) what data the AI was trained on (including what data was in the "similar" data group and how the corresponding outputs were compared to the "similar" data group), and (4) an explanation of the methodologies used for prediction.

To gather such information, in addition to the questions suggested in the sections above, the following questions might be considered by proponent (and any opponent) of the evidence:[3]

---

[3] This portion draws substantially from the Checklist for Reliable Data Review Standard developed by Stanley C. Ahalt, Dean, School of Data Science and Society at the University of North Carolina.

- For what purpose was the AI designed, trained, tested, and validated? Is that the purpose for which consideration is proposed? If not, why should the court consider admitting or using the AI for an alternative purpose?
- What safeguards could ensure appropriate use in context? Will they suffice?
- What data or features did the algorithm consider, and what weights were assigned to those features in determining the AI's output? Are the details of those features and their associated weights accessible for review? Specifically, do the considered features include sensitive attributes such as race, gender, or other protected characteristics -- or proxies for such characteristics? If these sensitive attributes are included, what is the justification for their inclusion, and how has the use of such features been evaluated to ensure compliance with ethical standards and constitutional protections?
- Does the AI have equal or disparate error rates across different racial, gender, or other suspect categories?

- **More on the AI Lifecycle:** In assessing AI evidence, it may be useful to consider the AI lifecycle – collection, processing, management, analysis, visualization/presentation – with a focus on the questions that arise at each phase. For that purpose, consider this breakdown:

Collection:
  - What was the intent behind the data collection? E.g., was it collected by a neutral academic or paid-for expert? Was it collected in anticipation of litigation?
  - How was the data collected? Were there potential collection methods that were rejected? If so, why?
  - Does the data collected adequately represent the world of potential data that could be collected? Are there areas of information that were not included that are necessary for a complete data picture?
  - Are there groups or populations that are excluded or disadvantaged by the collection method?
  - Is the data collected based on past discriminatory or biased behaviors?

Processing:
- o What assurances are there that the dataset in issue does not have incorrect or missing data?
- o Have the dataset creators adequately addressed potential differences in data types and formatting?

Management:
- o Does the data have clear, traceable sources?
- o Can the dataset owners provide sufficient evidence that the data was stored in a secure manner, protected from accidental or intentional manipulation?
- o Is the dataset maintained in a way that, to the extent relevant, protects privacy of affected individuals or groups?
  If necessary, can the dataset owners provide access (i.e., to opposing counsel, a special master, etc.) to the underlying data to allow for meaningful review?

Analysis:
- o How was the model validated? Can the expert show that the model is neutral and not based on a biased model derived from biased beliefs or data?
- o What are potential sources of systematic error in this model?
- o For stated nominal errors, who will be affected by those errors? What is the potential cost associated with the error?

Visualization/Presentation:
- o Does the visualization/presentation of the results accurately reflect the underlying data and analysis? Is the visualization/presentation and related narrative backed up by the data?
- o Is the visualization/presentation clear and easily understood?
- o Does the probative value of the visualization/presentation substantially outweigh the danger, if any, of unfair prejudice, confusing the issues, misleading the jury, undue delay, wasting time, or needlessly presenting cumulative evidence?