IV. Conclusion and Drafting Alternatives

This memo has covered a number of possible changes to address deepfakes and machine learning. Assuming, again, that any change is necessary, the most straightforward and effective changes are the following:

1. *Changes to Rule 901(b):* [ASSUMING NO ADDITION OF RULE 707]

[901](b) **Examples**. The following are examples only—not a complete list—of evidence that satisfies the requirement [of Rule 901(a)]:

(9) *Evidence about a Process or System*. For an item generated by a process or system:

(A) evidence describing it and showing that it produces ~~an accurate~~ **a reliable** result; and

(B) if the proponent acknowledges that the item was generated by artificial intelligence, additional evidence that:

(i) describes the training data and software or program that was used; and

(ii) shows that they produced reliable results in this instance.

2. Proposed New Rule 901(c) to address "Deepfakes":

901(c): *Potentially Fabricated or Altered Evidence Created By Artificial Intelligence [By an Automated System].*

If a party challenging the authenticity of computer-generated or other electronic evidence demonstrates to the court that a jury reasonably could find that the evidence has been altered or fabricated, in whole or in part, by artificial intelligence [by an automated system], the evidence is admissible only if the proponent demonstrates to the court that it is more likely than not authentic.

Draft Committee Note

This new subdivision is intended to set forth guidance and standards when the opponent alleges that an audio or video item is a "deepfake" --- i.e., that it has been altered by artificial intelligence so that it is not what the proponent says it is.

The term "artificial intelligence" can have several meanings, and it is not a static term. In this rule, "artificial intelligence" means software used to perform tasks or produce output previously thought to require human intelligence.

The rule sets out a two-step process for regulating claims of deepfakes. First, the opponent must set forth enough information for a reasonable person to find that the item has been altered by the use of artificial intelligence. Thus, a broad claim of "deepfake" is not enough to put the court and the proponent to the time and expense of showing that the item has not been manipulated by artificial intelligence. Second, assuming that the opponent has shown enough to merit the enquiry, the proponent must show to the court that the item is more likely than not genuine. While that Rule 104(a) standard is higher than ordinarily required for a showing of authenticity, it is justified given that any member of the public has the capacity to make a deepfake, with little effort and expense, and deepfakes have become more difficult to detect. It is therefore reasonable for the court to require a showing, by a preponderance of the evidence, that the item is not a deepfake, once the opponent has met its burden of going forward.

3. New Rule 707

Rule 707. Machine-generated Evidence

Where the output of a process or system would be subject to Rule 702 if testified to by a human witness, the court must find that the output satisfies the requirements of Rule 702 (a)-(d). This rule does not apply to the output of basic scientific instruments or routinely relied upon commercial software.

Draft Committee Note

Expert testimony in modern trials increasingly relies on software- or other machine-based conveyances of information, from software-driven blood-alcohol concentration results to probabilistic genotyping software. Machine-generated evidence can involve the use of a computer-based process or system to make predictions or draw inferences from existing data. When a machine draws inferences and makes predictions, there are concerns about the reliability of that process, akin to the reliability concerns about expert witnesses. Problems include using the process for purposes that were not intended (function creep); analytical error or incompleteness; inaccuracy or bias built into the underlying data or formulas; and lack of interpretability of the machine's process. Where an expert relies on such a method, the method – and the expert's reliance on it – will be scrutinized pursuant to Rule 702. But if machine or software output is presented on its own, without the accompaniment of a human expert, Rule 702 is not obviously applicable. Yet it cannot be that a proponent can evade the reliability requirements of Rule 702 by offering machine output directly, where the output would be subject to 702 if rendered as an opinion by a human expert. Therefore, new Rule 707 provides that if machine output is offered directly, it is subject to the requirements of Rule 702 (a)-(d).

It is anticipated that a Rule 707 analysis will involve the following, where applicable:

• Considering whether the inputs into the process are sufficient for purposes of ensuring the validity of the resulting output. For example, the court should consider whether the training

data for a machine learning process is sufficiently representative to render an accurate output for the population involved in the case at hand.

• [Ensuring that the opponent has been provided sufficient access to the program, and that independent researchers have had sufficient access to the program, to allow both adversarial scrutiny and sufficient peer review beyond simply validation studies conducted by the developer or related entities. Where a developer has declined to make a research license or equivalent access widely available to independent researchers, courts should be wary of allowing output from such a process.]

• Considering whether the process has been validated in circumstances sufficiently similar to the case at hand. For example, if the case at hand involves a DNA mixture of several contributors, likely related to each other, and a low quantity of DNA, the software should be shown to be valid in those circumstances before being admitted.

The final sentence of the rule is intended to give trial courts sufficient latitude to avoid unnecessary litigation over machine output that is regularly relied upon in commercial contexts outside litigation and that, as a result, is not likely to render output that is invalid for the purpose it is offered. Examples might include the results of a mercury-based thermometer, battery-operated digital thermometer, or automated averaging of data in a spreadsheet, in the absence of evidence of untrustworthiness.

The Rule 702(b) requirement of sufficient facts and data, as applied to machine-generated evidence, should focus on the information entered into the process or system that leads to the output offered into evidence.