# Guidance for Canadian Judges on Evaluating Digital Media Evidence

September 2026

Prepared by Martin Felsky, PhD, JD

## CONTENTS


**1. Acknowledgment** 4

**2. Purpose and Scope** 4

**3. Synthetic Media and the Challenge for Courts** 4

3.1 The Deepfake Defence and the Liar's Dividend 6

**4. Judicial Gatekeeping: Authentication, Reliability and Weight** 7

**5. The Presumption of System Reliability and its Limits** 8

**6. Accountability** 8

6.1 Identifying Who Is Standing Behind the Evidence 8

6.2 Ambiguous or Evasive Authentication Evidence 9

6.3 Cases Without an Authenticating Witness 9

**7. Interim Proceedings and Relaxed Evidentiary Standards** 10

**8. Distinguishing Different Forms of AI-Related Evidence** 10

8.1 Where AI Use Is Acknowledged 11

8.2 Where AI Use Is Unacknowledged (or Disputed) 12

8.2.1 Sanctioning Deliberate Use of Synthetic Evidence 12

8.3 Indirect Evidence Prepared with AI 13

**9. Calibrating the Judicial Response: Legal and Procedural Approaches** 14

9.1 Early Disclosure 15

9.2 Expert Evidence and Explainability 15

**10. Calibrating the Judicial Response: Technical and Forensic Approaches** 15

10.1 Chain of Custody 15

10.2 System Controls 16

10.3 AI Detection Tools 16

**11. Access to Justice and Procedural Fairness** 17

**12. Checklist for Acknowledged Use of AI** 18

12.1 Identify the Purpose of the Evidence 18

12.2 Understand the AI Process 18

12.3 Assess the Underlying Data 18

12.4 Assess Reliability of the Output 19

12.5 Where AI Has Enhanced Existing Evidence 19
12.6 Where AI Was Used for Identification 20
12.7 Consider Potential Prejudice 20
12.8 Consider Expert Evidence 20
12.9 Determine Appropriate Judicial Response 20
**13. Checklist for Unacknowledged Use of AI 21**
13.1 Identify the Issue 21
13.2 Accountability 21
13.3 Establish Provenance 22
13.4 Examine Chain of Custody 22
13.5 Determine Whether Alterations Occurred 22
13.6 Assess Reliability 23
13.7 Consider Expert Evidence 23
13.8 Determine Appropriate Judicial Response 23
**14. Checklist for AI-Assisted Indirect Evidence 24**
14.1 Identify the Nature and Extent of AI Use 24
14.2 Determine Who Is Standing Behind the Evidence 24
14.3 Assess Personal Knowledge 24
14.4 Consider Reliability Concerns 25
14.5 Assess the Ability to Test the Evidence 25
14.6 Consider Expert-Specific Issues (if applicable) 25
14.7 Determine Appropriate Judicial Response 25

## 1. ACKNOWLEDGMENT

The author would like to acknowledge with thanks the valuable contributions of Professor Karen Eltis in the preparation of this paper.

## 2. PURPOSE AND SCOPE

This guidance is intended to assist judges who may encounter digital media evidence in circumstances where there is an allegation, or even a possibility, that the material has been fabricated or materially altered using generative artificial intelligence (AI) tools.

The guidance is not limited to contested allegations of fabrication. It is equally relevant whenever digital media evidence is tendered and requires authentication under the *Canada Evidence Act* or applicable provincial evidence legislation, even where the parties have stipulated as to admissibility. As gatekeepers of critical evidence, courts retain responsibility for ensuring that evidence admitted into the record is what it purports to be and for determining the weight it should receive.

The proliferation of AI-generated evidence may create risks extending beyond individual cases, including diminished public confidence in adjudication, challenges to procedural fairness, and increased difficulty in explaining and scrutinizing digital evidence.

The objective of this document is not to propose new evidentiary doctrine, but to clarify how existing Canadian legal principles should be applied thoughtfully and rigorously in light of rapidly evolving technological developments. The advent of generative AI and the evolving risks associated with its use in court requires careful attention to how digital recordings were created and how they are proposed to be used.

This guidance considers issues relating to the introduction of evidence, including expert testimony, and excludes discussion of the issues around so-called hallucinations in the legal submissions of counsel.

## 3. SYNTHETIC MEDIA AND THE CHALLENGE FOR COURTS

Courts have long dealt with forged signatures, altered photographs, fabricated documents, and false testimony. Generative AI does not render traditional evidentiary principles obsolete (although it may necessitate a reconsideration of the law of evidence).[1] Rather, it requires courts

[1] See for example Delfino, Rebecca, *Deepfakes on Trial 2.0: A Revised Proposal for a New Federal Rule of Evidence to Mitigate Deepfake Deceptions in Court* (February 15, 2025), Loyola Law School, Los Angeles Legal Studies Research Paper No. 2025-10. The contrary view is expressed in Pfefferkorn, Riana, *'Deepfakes' in the Courtroom* (October 1, 2020), Boston University Public Interest Law Journal, Vol. 29, No. 2, 2020.

to apply those principles in circumstances where fabrication and manipulation are much easier, less expensive, and more difficult to detect than in the past.

Readily available tools can be prompted to create convincing footage portraying events that never occurred, people who don't exist, and words that were never spoken. Where such "synthetic media" are presented with an intent to deceive, the term "deepfake" is often used.[2] At the same time, authentic digital recordings may be altered by AI in increasingly subtle ways that are difficult to detect by unaided observation.

Courts may now encounter examples and allegations of synthetic media, including AI-generated or AI-altered digital evidence in a broad range of proceedings, from criminal prosecutions and family disputes to civil litigation, administrative proceedings, and professional discipline matters. The challenge is both technological and evidentiary. While digital media remain powerful as evidence, judges can no longer confidently assume that what appears genuine is prima facie admissible.[3]

**Importantly, synthetic media create two distinct risks. The first is that fabricated or manipulated evidence may be presented as authentic. The second is that authentic evidence may be challenged as fabricated. Both risks have the potential to undermine the truth-seeking function of the courts and public confidence in the administration of justice**.

The consequence is not that digital media have become inherently unreliable. Nor does it mean that courts should approach every recording with suspicion. Rather, judges should recognize that video realism and apparent coherence are no longer sufficient, by themselves, to establish authenticity. "In practical terms, synthetic media have become indistinguishable from authentic recordings for ordinary people and, in some cases, even for institutions."[4] According to some, deepfakes are soon expected to evolve into "interactive … actors whose faces, voices and mannerisms adapt instantly to a prompt" replicating real humans to a fault.[5]

Judges need not become technical experts in artificial intelligence or digital forensics. Their role remains fundamentally the same: to assess evidence, determine credibility, apply established

[2] Financial and Consumer Services Commission of New Brunswick, *Protecting yourself from AI and deepfake scams*.
[3] As recently stated by Chief Justice Wagner, "In an environment where information can be generated and shared rapidly, distinguishing fact from fiction has become more difficult and more important. New technologies offer opportunities to improve access and efficiency, but they also require courts to adapt quickly and thoughtfully. That adaptation requires investment and time. And it is necessary, again, if we want to maintain public trust in the justice system." Supreme Court of Canada, Annual news conference with the Chief Justice of Canada (June 9, 2026).
[4] The Conversation, *Deepfakes leveled up in 2025 – here's what's coming next* (December 26, 2025).
[5] Ibid.

legal principles, and ensure procedural fairness. The challenge is not to master rapidly changing technologies, but to understand their implications sufficiently to evaluate the evidence presented and to determine when additional scrutiny, expertise, procedural or technical safeguards may be required.

## 3.1 THE DEEPFAKE DEFENCE AND THE LIAR'S DIVIDEND

Much public discussion of synthetic media focuses on the possibility that fabricated evidence will be introduced and admitted into legal proceedings. Equally significant, however, is the possibility that *authentic evidence* will be successfully challenged as fabricated, and thus excluded from evidence. In either case the result may lead to a miscarriage of justice.

This latter phenomenon has become known as the "deepfake defence" or, in some academic literature, the "liar's dividend."[6] As public awareness of and concern about deepfakes increases, parties confronted with damaging evidence may claim that recordings (for example video surveillance or wiretaps) have been generated using artificial intelligence. In some cases, such allegations may be warranted. In others, they may be advanced strategically to delay proceedings or undermine otherwise reliable evidence.[7]

The existence of deepfake technology therefore creates a paradox. The more widely the concept of synthetic media becomes known, the easier it becomes both to fabricate evidence and to challenge authentic evidence.

Courts must remain alert to both possibilities. Obviously, an allegation that evidence is synthetic should neither be accepted nor rejected simply because deepfake technology exists. As with any challenge to authenticity, the allegation itself must be evaluated based on evidence rather than speculation.

Canadian courts have already begun confronting this issue. In *R. v. Medow* (2025 ONCJ 661), the court accepted that deepfake technology exists and that its increasing sophistication is a matter capable of judicial notice. At the same time, the court rejected unsupported allegations that police body-worn camera recordings had been manipulated, emphasizing that speculation

---

[6] "The Liar's Dividend in Legal Proceedings Is 'The Deepfake Defense.'" Delfino, Rebecca, *The Deepfake Defense—Exploring the Limits of the Law and Ethical Norms in Protecting Legal Proceedings from Lying Lawyers* (February 12, 2023), Loyola Law School, Los Angeles Legal Studies Research Paper No. 2023-02, 84.5 Ohio St. L.J. 1068 (2024), page 3.

[7] Raffaella Spone, a "cheerleader mom," was arrested and internationally vilified in 2021 after being accused of using cutting-edge "deepfake" AI technology to doctor photos and videos of her daughter's teenage rivals vaping, drinking, and posing nude to get them kicked off their elite squad. In court it was revealed that the police had relied entirely on a visual "eye test" to declare the footage fake; in fact the videos of the teenagers vaping and partying were entirely real. The Guardian, *She was accused of faking an incriminating video of teenage cheerleaders. She was arrested, outcast and condemned. The problem? Nothing was fake after all* (May 11, 2024).

alone could not displace otherwise admissible evidence. The decision confirms the principle that courts should neither ignore the reality of synthetic media nor assume that every allegation of manipulation warrants extensive inquiry.[8]

## 4. JUDICIAL GATEKEEPING: AUTHENTICATION, RELIABILITY AND WEIGHT

Judges remain responsible for determining whether digital evidence is admissible, whether it is sufficiently reliable to be considered, and ultimately what weight it should receive.

Authentication addresses a threshold question: is there evidence capable of supporting a finding that the digital material is what it purports to be? Reliability concerns whether the evidence can reasonably be relied upon for its intended purpose. Weight concerns the significance ultimately assigned to the evidence in light of the entire record.

In the context of synthetic media, it is important not to collapse these inquiries into one another. A recording may satisfy the threshold requirements for authentication yet still attract little weight because of concerns regarding provenance, continuity, unexplained modifications, or the circumstances in which it was created. Conversely, concerns about possible manipulation do not necessarily prevent admissibility where there is sufficient evidence to permit the trier of fact to assess authenticity and reliability.

Judges should also distinguish between challenges directed at the recording itself and challenges directed at what the recording purports to depict. In some deepfake cases, the central issue may not be whether a particular file exists or whether it has been preserved correctly. Rather, the issue is whether the events, statements, images, or conduct represented by the recording occurred.

Evidence concerning system integrity, storage practices, metadata, or chain of custody may assist in establishing that a file has not been altered during its lifecycle. Such evidence may not resolve whether the content itself was fabricated before entering the system or generated using artificial intelligence.

The gatekeeping function therefore requires more than technical verification. It requires judges to consider the totality of the circumstances, including provenance, continuity, corroboration, witness testimony, and where necessary, expert evidence. The objective is not to eliminate all uncertainty. Rather, it is to ensure that digital evidence is subjected to the same disciplined scrutiny that courts have long applied to other forms of potentially unreliable evidence.

[8] Another case referencing the deepfake defence is *R. v. Cheng*, 2025 ONCJ 252, where Justice Kenkel noted: "I agree with Mr. Cheng that the emergence of "deepfake" evidence and the ability to manipulate electronic records means this court must carefully consider the weight to be given these exhibits."

The existence of synthetic media technology should not lead courts to impose heightened authentication requirements in every case. For now, most digital evidence will continue to present no genuine authenticity concern. At the same time, courts should remain attentive to circumstances that justify closer scrutiny, particularly where digital media are central to the issues in dispute, appears unusually persuasive, or is challenged on a plausible evidentiary basis.

## 5. THE PRESUMPTION OF SYSTEM RELIABILITY AND ITS LIMITS

Canadian law has long recognized a presumption that electronic systems operating in the ordinary course of business are reliable, though this presumption has not been without its critics.[9] In the UK, the recent Horizon Post Office scandal has prompted a reconsideration of a similar presumption.[10]

Deepfake technology complicates this presumption, because even if (for example) a recording device is functioning properly and faithfully captured real incidents, the content may nevertheless have been surreptitiously altered after capture. Judges should therefore avoid equating routine system use with guaranteed authenticity, and distinguish between the reliability of recording systems and the authenticity of depicted events.

## 6. ACCOUNTABILITY

Digital evidence does not enter the record on its own. Rather, it is introduced through witnesses, affiants, investigators, experts, or other individuals who provide evidence supporting its admissibility. Traditional evidentiary procedures require those individuals to assume responsibility for the assertions they make regarding the evidence.

Viewed in this way, many deepfake concerns are not as unprecedented as they might appear. Fabrication has always been a possibility, and the obligation on parties to establish an evidentiary foundation remains unchanged by the arrival of synthetic media. Cases involving alleged deepfakes thus often turn on accountability.

### 6.1 IDENTIFYING WHO IS STANDING BEHIND THE EVIDENCE

When digital evidence is tendered, judges may find it useful to ask a fundamental procedural question: Who, if anyone, is standing behind this evidence? In this context "standing behind the

[9] Chasse, Ken, *Challenging Electronic Systems' and Devices' Ability to Produce Reliable Evidence* (April 25, 2019), Law Society of Ontario. Felsky, Martin, *Reconsidering the Presumption of Reliability for Digital Evidence in Canadian Law – A Cautionary Reflection on the UK Horizon Scandal* (August 21, 2025), Heuristica.
[10] The Guardian, *Update law on computer evidence to avoid Horizon repeat, ministers urged*, (January 21, 2024).

evidence" refers to a person who assumes responsibility for its authenticity, accuracy, and provenance.

In many cases, a witness or affiant will expressly attest to matters such as:

- the source of the record
- the circumstances of its creation
- the manner in which it was obtained
- its continuity and preservation
- their belief that it accurately depicts the events in question

Where such evidence is given clearly and unequivocally, the witness assumes responsibility for those assertions and may be subject to the ordinary consequences associated with false testimony, including findings affecting credibility and, where appropriate, potential perjury or related proceedings. This accountability framework remains an important safeguard in the era of synthetic media.

## 6.2 AMBIGUOUS OR EVASIVE AUTHENTICATION EVIDENCE

Judges should also remain alert to situations in which evidence is presented through language that is vague, indirect, or carefully qualified. Examples may include statements such as:

- "I received this video from another person."
- "I have no reason to doubt its authenticity."
- "This appears to show the events in question."
- "I understand this recording to be genuine."

Such statements may establish possession or transmission of a recording without necessarily establishing authenticity or personal knowledge.

Of course, ambiguity may simply reflect the limits of a witness's knowledge. On the other hand, imprecise language may have the (intended) effect of distancing the witness from responsibility for the truthfulness of the underlying evidence.

Judges should be careful to distinguish testimony that merely describes the existence or receipt of digital evidence from testimony that affirmatively supports authenticity.

## 6.3 CASES WITHOUT AN AUTHENTICATING WITNESS

Not all forms of digital evidence require testimony from a witness with direct knowledge of the events depicted. Canadian law recognizes that electronic evidence may sometimes be

authenticated through circumstantial evidence, metadata, system evidence, as business records, or through other indirect means.

Nevertheless, judges may wish to consider whether someone has provided evidence capable of establishing:

- where the material originated
- how it was obtained
- whether it remained intact
- why it should be regarded as authentic

The absence of a witness capable of addressing these matters does not necessarily render evidence inadmissible. It may, however, affect the weight that can properly be assigned to it and may justify closer scrutiny of its origin and history.

## 7. INTERIM PROCEEDINGS AND RELAXED EVIDENTIARY STANDARDS

Deepfake concerns are not confined to trials. Digital materials may also arise:

- on interlocutory motions
- in applications proceeding largely by affidavit
- before administrative tribunals
- in disciplinary proceedings
- in other contexts where strict evidentiary rules are relaxed

In such settings, courts and tribunals frequently receive evidence that would not necessarily be admissible at trial. While the existence of relaxed evidentiary standards does not eliminate the need to assess authenticity, it may require heightened awareness of the limitations of the material being considered.

Judges may find it useful to approach alleged deepfake evidence in these contexts in a manner analogous to hearsay evidence: the issue may not be whether the material can be considered at all, but rather what weight can safely be assigned to it given the available indicia of reliability.

## 8. DISTINGUISHING DIFFERENT FORMS OF AI-RELATED EVIDENCE

In an often-cited paper, AI-generated evidence is differentiated, for legal purposes, as acknowledged and unacknowledged.[11] For the purposes of this guidance, it may be useful to distinguish among three broad forms of AI-related evidence: acknowledged AI evidence,

[11] Grossman, Maura and Grimm, Paul, *Acknowledged and Unacknowledged AI-Generated Evidence* (May 23, 2025), Science and Technology Law Review.

unacknowledged AI evidence, and indirect AI evidence. Although these three categories all involve artificial intelligence, they engage different evidentiary concerns.

**Acknowledged AI evidence primarily raises questions of reliability. Unacknowledged AI evidence primarily raises questions of authenticity. Indirect AI evidence primarily raises questions of accountability.**

Identifying which category is engaged may assist courts in determining the most appropriate evidentiary framework and procedural response.

## 8.1 WHERE AI USE IS ACKNOWLEDGED

Acknowledged AI evidence consists of synthetic media where the involvement of AI is disclosed to the court. Examples may include an expert who relies on AI-assisted analytical software, a witness who uses AI-based image enhancement tools, a translator who employs AI-assisted translation software, or a professional who incorporates AI-enabled systems into an established workflow.

***The focus here is reliability, not deception***. The court may need to determine whether the AI-assisted process is sufficiently dependable for its intended purpose and whether the witness can explain and defend the methodology employed.

In many cases, the appropriate analytical framework may resemble the approach already applied to expert evidence. Existing principles governing the admissibility of expert opinion evidence, including considerations derived from cases such as *R. v. Mohan* and the reliability factors commonly associated with *Daubert*,[12] may provide useful guidance. Depending on the circumstances, courts may consider such factors as whether:

- the AI tool is accepted within the relevant professional community
- the methodology can be explained and independently evaluated
- the limitations, error rates, or known weaknesses of the tool are understood
- appropriate human oversight was exercised
- the witness can meaningfully explain how the AI-assisted process contributed to the final result

[12] A Mohan/Daubert screening is a legal evaluation used by Canadian courts to determine if expert witness testimony and novel scientific evidence are reliable enough to be admitted at trial. This framework blends Canada's core admissibility standard established in *R. v. Mohan*, [1994] 2 S.C.R. 9 with the scientific reliability criteria from the landmark U.S. case *Daubert v. Merrell Dow Pharmaceuticals Inc.*, 509 U.S. 579 (1993), which was formally adopted into Canadian law through *R. v. J.-L.J.*, [2000] 2 S.C.R. 600.

AI involvement is not, standing alone, a basis for excluding evidence. Some forms of AI assistance are becoming standard professional practice, and the analysis should remain tied to reliability, transparency, and the opposing party's ability to mount a meaningful challenge.

## 8.2 WHERE AI USE IS UNACKNOWLEDGED (OR DISPUTED)

Unacknowledged AI evidence arises when AI-generated or AI-manipulated material is presented to the court as though it were authentic human-created content or an accurate record of actual events.

Here, authenticity is the central concern. Unlike acknowledged use, undisclosed AI involvement creates a risk that the court will be misled about the true origin or nature of the material. Courts may encounter at least two distinct situations raising these authenticity concerns.

In the first, a party knowingly presents fabricated or manipulated evidence as genuine. Such conduct may engage serious concerns relating to fraud upon the court, perjury, abuse of process, professional misconduct, or contempt.

In the second, a party, witness, or counsel may present AI-generated material believing it to be genuine. The evidentiary issue in such cases is not intentional deception but mistaken reliance on material whose authenticity cannot be established.

As generative AI tools become increasingly sophisticated, courts will need to place greater emphasis on traditional methods of authentication. Once again, matters that may bear on the inquiry include:

- the origin and chain of custody of the evidence
- the availability of original source files and associated metadata
- evidence concerning the circumstances in which the material was created, stored, transmitted, and preserved
- expert evidence addressing authenticity where appropriate
- the presence of anomalies or inconsistencies suggesting manipulation

### 8.2.1 SANCTIONING DELIBERATE USE OF SYNTHETIC EVIDENCE

Existing evidentiary and procedural rules provide courts with significant tools to respond when synthetic media are intentionally used to mislead the court. The challenge posed by deepfakes is not limited to authentication and reliability; it also raises issues of abuse of process, professional responsibility, and protection of the integrity of judicial proceedings. Where a court concludes that a party has knowingly submitted fabricated or materially altered AI-generated evidence, traditional sanctions may be available without the need for new statutory powers.

An example from the United States is *Mendones v. Cushman & Wakefield* (California Superior Court, 2025 [*Mendones*]). In *Mendones*, the court concluded that several videos and other exhibits submitted in support of a summary judgment motion appeared to have been generated or altered using artificial intelligence. The court noted characteristics commonly associated with synthetic media, including unnatural facial expressions, looping video sequences, and inconsistencies between purportedly authentic recordings and the challenged exhibits. After issuing an order to show cause and providing an opportunity to respond, the court imposed terminating sanctions and considered (but did not order) referral for potential criminal prosecution based on the submission of allegedly fabricated evidence.

*Mendones* illustrates that courts need not view deepfakes solely as a technological problem. In appropriate cases, they may be addressed through traditional judicial powers designed to protect the integrity of the adjudicative process. Nevertheless, the case should not be understood as endorsing visual inspection as a reliable long-term method of identifying synthetic media. The apparent artefacts and inconsistencies that may reveal a deepfake today may disappear as generative technologies improve. Future cases may involve fabricated recordings that are indistinguishable from authentic recordings to the human observer. For that reason, courts should be cautious about relying exclusively on their own perception of a recording's appearance and should instead focus on the broader evidentiary foundation supporting authenticity and reliability.

## 8.3 INDIRECT EVIDENCE PREPARED WITH AI

Indirect AI evidence arises where artificial intelligence is used in the preparation or formulation of evidence, but the AI-generated output itself is not being tendered as evidence. Indirect AI evidence raises concerns that differ from those associated with acknowledged or unacknowledged AI evidence. The primary issue is neither authenticity nor the reliability of the AI system itself. Rather, the concern is whether the human witness remains genuinely accountable for the content being presented to the court.

Examples may include affidavits, witness statements, expert reports, investigative summaries, chronologies, or other documents that have been drafted, edited, summarized, translated, or otherwise assisted by generative AI tools. In such cases, the evidence presented to the court remains the evidence of the witness, expert, or party who adopts the document, rather than the output of the AI system itself.

Generative AI systems may produce inaccurate statements, fabricated facts, fictitious citations, misleading summaries, or subtle distortions that are not readily apparent to the user. A witness who relies extensively on AI-generated content may inadvertently adopt statements that do not reflect their own knowledge, observations, analysis, or opinions. In extreme cases, the court may be left uncertain whether the document reflects the witness's evidence or merely the output of a

generative system. Accordingly, where significant AI assistance is suspected or acknowledged, judges may wish to consider whether:

- the witness personally reviewed and adopted the contents of the document
- the witness can explain the factual basis for the statements made
- the witness can distinguish personal knowledge from information generated by AI
- the use of AI has affected the ability of opposing parties to test the evidence through cross-examination
- the document remains a reliable expression of the witness's own evidence

Where a demonstrative aid effectively becomes substantive proof of a disputed fact, the issue may be less one of artificial intelligence and more one of evidentiary procedure. Courts should remain vigilant against the introduction of evidence through procedural mechanisms that bypass ordinary requirements for authentication, reliability, and accountability.

## 9. CALIBRATING THE JUDICIAL RESPONSE: LEGAL AND PROCEDURAL APPROACHES

A fabricated or materially altered recording may function as primary evidence of fact. It may be emotionally persuasive, difficult to rebut, and capable of influencing determinations of credibility, liability, or guilt. In such circumstances, a focused and structured judicial response may be required.[13] This may include:

- conducting a voir dire on authenticity or admissibility
- considering duties of the court generally to self-represented parties
- requiring production of original files, metadata, or related technical information
- considering whether expert evidence is necessary to resolve technical disputes
- providing clear reasons explaining how authenticity and reliability were assessed

Not every allegation of deepfake manipulation will be credible. Courts must avoid allowing speculative assertions to derail proceedings. Factors that may bear on this assessment include:

- evidence of chain of custody and system controls (see below)
- presence of unexplained edits, redactions, or enhancements
- plausible motive to fabricate
- materiality of the evidence to the case

[13] Chief Justice Wagner has repeatedly emphasized that the reliability of information is now central to maintaining public confidence in courts. In remarks delivered during the Supreme Court of Canada's 150th-anniversary events in 2025, he noted that declining trust in traditional media and increasing reliance on social media have been accompanied by the rapid spread of misinformation, warning that courts must remain vigilant in protecting the values that sustain public confidence in the justice system. Supreme Court of Canada, 150th Anniversary Symposium – Words of Welcome by the Right Honourable Richard Wagner, P.C., Chief Justice of Canada (April 10, 2025).

- ability of opposing parties to meaningfully challenge authenticity

### 9.1 EARLY DISCLOSURE

In cases where concerns regarding authenticity are reasonably anticipated, counsel should be encouraged to promote the timely disclosure and conservation of primary, native-format files. While the execution of these steps rests with the parties and their legal representatives, evidentiary matters should be routinely dealt with during the discovery stage of civil litigation or the disclosure phase of criminal matters, well in advance of trial. Where existing rules do not adequately address the misuse of synthetic media, courts may consider issuing notices, practice directions, or policies to promote disclosure and preservation.

### 9.2 EXPERT EVIDENCE AND EXPLAINABILITY

In some cases, forensic expertise may be necessary to address alleged digital generation or manipulation. This highlights the importance of explainability where AI-generated analysis is involved. Courts should be cautious about accepting expert conclusions derived from opaque or non-reproducible AI systems, or admitting evidence that cannot be meaningfully contested.

## 10. CALIBRATING THE JUDICIAL RESPONSE: TECHNICAL AND FORENSIC APPROACHES

### 10.1 CHAIN OF CUSTODY

Chain of custody refers to the continuity of possession, handling, transfer, and storage of digital evidence from creation or seizure to its presentation in court. In the digital context, relevant considerations may include:

- evidence of when and how the recording was captured
- identification of the recording device
- upload, transfer, and storage logs
- records of access or handling
- use of hash values (digital fingerprints) to verify file integrity
- documentation of any redactions or edits, including their purpose and method

The absence of comprehensive continuity documentation does not automatically render evidence inadmissible. However, where the recording is central to the case, a documented chain of custody may significantly assist the court in assessing reliability.

## 10.2 SYSTEM CONTROLS

System controls refer to technological safeguards embedded in the recording, storage, and retrieval environment. These may include:

- automated logging and audit trails
- encryption and access controls
- tamper-evident storage architecture
- restricted editing permissions
- secure cloud or server environments with documented integrity protections

The purpose of examining system controls is to understand whether the environment in which the recording was generated and maintained makes material tampering plausible, detectable, or unlikely. Judges should neither presume reliability from digital storage alone nor infer manipulation from the absence of elaborate safeguards. The assessment must remain fact specific.

## 10.3 AI DETECTION TOOLS

One of the curious aspects of deepfake awareness, especially in the United States, is the so-called "CSI effect," a situation in which juries assume, based on the fictionalized television series, that every prosecuting attorney has available the most sophisticated forensic tools, and uses them on every case. With respect to deepfakes, there is a similar concern that juries – or indeed any fact finders, including judges – might assume that prosecutors have at their fingertips the most sophisticated deepfake detection tools.

Current commercial tools are useful for screening, but they are not sufficiently reliable to conclusively determine whether an image, voice or video is authentic or synthetic, particularly when the video is of high quality. Nor do they eliminate the need for human expertise when the results are disputed. To date it has been easier to create a new generation of deepfakes than it to build a reliable detector.

Where a party relies on a commercial detection tool, the court should regard the tool's output as an analytical aid rather than a definitive determination of authenticity. The ultimate assessment of authenticity, reliability, and weight remains a matter for the court based on the entirety of the evidence.

Detection systems may rely on proprietary algorithms, undisclosed training data, or probabilistic assessments. If relied upon, the basis for the tool's operation, methodology, and error rates should be sufficiently explained, typically through expert evidence, to permit meaningful judicial evaluation.

## 11. ACCESS TO JUSTICE AND PROCEDURAL FAIRNESS

Even mere allegations of synthetic media deployment raise potential access to justice concerns. Many litigants, particularly self-represented or economically disadvantaged individuals, will lack the resources to retain forensic experts capable of detecting synthetic media or analyzing digital evidence. Litigants, particularly vulnerable litigants, cannot simply be left to challenge potentially synthetic evidence without meaningful support.

Judges should remain attentive to fairness and access to justice considerations in such circumstances. While courts must maintain impartiality, they may consider appropriate procedural measures to ensure that allegations of manipulation receive genuine scrutiny without allowing unsubstantiated claims to consume disproportionate resources or undermine trust.

Courts must find a middle ground when weighing these factors, avoiding both uncritical acceptance on one hand, and unwarranted distrust on the other.

# Checklists for Responding to Digital Media Evidence

As with all bench aids, the checklists are intended as practical tools and should be applied in a manner consistent with the governing law, the particular circumstances of each case, and judicial discretion.[14]

## 12. CHECKLIST FOR ACKNOWLEDGED USE OF AI

### 12.1 IDENTIFY THE PURPOSE OF THE EVIDENCE

☐ Is the evidence being tendered as:

- ☐ Real evidence
- ☐ Expert evidence
- ☐ Demonstrative evidence
- ☐ Analytical evidence
- ☐ Investigative aid
- ☐ Other

☐ Has the use of AI been disclosed? In what phase of the proceedings?

### 12.2 UNDERSTAND THE AI PROCESS

☐ What AI tool was used?

☐ What was the purpose of the AI tool?

☐ What data was provided to the system?

☐ What prompts or instructions were used?

☐ What human review occurred?

☐ Can the process be adequately explained?

### 12.3 ASSESS THE UNDERLYING DATA

☐ Is the source data known?

☐ Is the source data reliable?

[14] The Canadian Judicial Council acknowledges the work undertaken by the National Center for State Courts in developing "bench cards" addressing AI-generated and AI-manipulated evidence. While the checklists presented here have benefitted from that work, they have been adapted to reflect the needs of Canadian judges.

☐ Has the source data been preserved?

☐ Are important inputs missing?

☐ Could different inputs have produced materially different outputs?

## 12.4 ASSESS RELIABILITY OF THE OUTPUT

☐ Can the result be independently verified?

☐ Is the methodology reproducible?

☐ Can witnesses or experts explain the output?

☐ Is there evidence of known limitations or error rates?

☐ Does the output accurately reflect the underlying evidence?

## 12.5 WHERE AI HAS ENHANCED EXISTING EVIDENCE

☐ Is the original, unenhanced version available?

☐ Has the enhancement process been documented?

☐ Can the court compare the original and enhanced versions side-by-side?

☐ Have all modifications and AI tools used been disclosed?

☐ What was the purpose of the enhancement (e.g., noise reduction, image sharpening, colour correction, voice isolation, upscaling)?

☐ Did the enhancement merely improve clarity, or did it reconstruct, infer, generate, or fill in missing information?

☐ Could the enhancement process have altered, distorted, exaggerated, or introduced features that were not reliably present in the original evidence?

☐ For audio evidence, could the enhancement have affected words, sounds, speaker characteristics, timing, or context?

☐ For image or video evidence, could the enhancement have affected facial features, objects, colours, movements, lighting, or other material details?

☐ Has any witness or expert explained the limitations, accuracy, and potential error rate of the enhancement process?

☐ Is the enhancement methodology sufficiently reliable for the purpose for which the evidence is tendered?

☐ Does the enhancement raise concerns affecting admissibility, weight, or both?

## 12.6 WHERE AI WAS USED FOR IDENTIFICATION

☐ Does the evidence involve:

- ☐ Facial recognition
- ☐ Voice recognition
- ☐ Biometric matching
- ☐ Pattern recognition
- ☐ Other identification technology

☐ Is there evidence concerning reliability?

☐ Are known limitations or biases addressed?

☐ Is independent validation available?

## 12.7 CONSIDER POTENTIAL PREJUDICE

☐ Could the presentation have disproportionate persuasive impact?

☐ Does the technology create an impression of precision that exceeds its actual reliability?

☐ Would limiting instructions assist?

☐ Should use be restricted to demonstrative purposes?

## 12.8 CONSIDER EXPERT EVIDENCE

☐ Is expert evidence required?

☐ Is the methodology sufficiently transparent?

☐ Can the expert explain the process in understandable terms?

☐ Can opposing parties meaningfully challenge the methodology?

## 12.9 DETERMINE APPROPRIATE JUDICIAL RESPONSE

☐ Additional disclosure required

☐ Production of underlying data

☐ Voir dire

☐ Expert evidence

☐ Limiting instruction

☐ Demonstrative-use restriction

☐ Exclusion

## 13. CHECKLIST FOR UNACKNOWLEDGED USE OF AI

### 13.1 IDENTIFY THE ISSUE

☐ Has a party alleged that the evidence has been fabricated, manipulated, or altered?

☐ Was the evidence not challenged during the discovery/disclosure phase?

☐ Is the challenge supported by evidence or circumstances beyond mere speculation?

☐ Is the allegation directed at:

- ☐ Video
- ☐ Audio
- ☐ Photograph
- ☐ Electronic document
- ☐ Text message / email
- ☐ Social media content
- ☐ Other digital evidence

### 13.2 ACCOUNTABILITY

*(Note that the absence of a witness who can meaningfully attest to authenticity, source, or handling does not necessarily render digital evidence inadmissible. However, it may affect the fairness of the proceeding insofar as it restricts opposing parties to test the evidence through cross-examination.)*

☐ Has an identifiable witness assumed responsibility for the evidence?

☐ Has a witness sworn or affirmed that the evidence is authentic?

☐ Has a witness explained how the evidence was created, obtained, or preserved?

☐ Does the witness have direct knowledge of the events depicted or recorded?

☐ Does the witness have direct knowledge of the process by which the evidence was created or obtained?

☐ Can the witness meaningfully answer questions concerning authenticity, provenance, and handling of the evidence?

☐ Is the witness able to identify the original source of the evidence?

☐ Is the witness prepared to accept responsibility for the accuracy and completeness of the evidence?

☐ Is the evidence being tendered through a person with no personal knowledge of its creation, source, or handling?

☐ Is the evidence effectively "orphaned" (i.e., no witness can adequately explain where it came from, how it was created, or why it should be trusted)?

## 13.3 ESTABLISH PROVENANCE

☐ Who created or obtained the evidence?

☐ When was it created or obtained?

☐ Is the original source known and verified?

☐ Is the original file available?

☐ Has the evidence been preserved in native format?

☐ Is metadata available?

## 13.4 EXAMINE CHAIN OF CUSTODY

☐ Who had possession or control of the evidence?

☐ Is there a documented chain of custody?

☐ Are there unexplained gaps in possession or handling?

☐ Was the evidence stored in a manner that protected against alteration?

## 13.5 DETERMINE WHETHER ALTERATIONS OCCURRED

☐ Has the evidence been:

- ☐ Edited
- ☐ Cropped
- ☐ Enhanced
- ☐ Redacted
- ☐ Converted to another format
- ☐ Processed using AI tools
- ☐ Otherwise modified

☐ Have any alterations been disclosed and justified?

☐ Are original versions available for comparison?

## 13.6 ASSESS RELIABILITY

☐ Is there corroborating evidence?

☐ Does other evidence support the authenticity of the recording or document?

☐ Are there inconsistencies that require explanation?

☐ Does the evidence appear unusually persuasive or unusually convenient?

☐ Is there evidence of motive or opportunity to fabricate?

## 13.7 CONSIDER EXPERT EVIDENCE

☐ Can the issue be resolved without technical expertise?

☐ Would expert evidence materially assist the court?

☐ Is a forensic examination required?

☐ Is the proposed expert methodology sufficiently transparent and explainable?

## 13.8 DETERMINE APPROPRIATE JUDICIAL RESPONSE

☐ Additional disclosure required

☐ Production of original or native files

☐ Production of metadata or audit information

☐ Preservation order

☐ Additional witness evidence concerning authenticity or provenance

☐ Voir dire

☐ Expert evidence

☐ Further inquiry into chain of custody

☐ Authentication concern

☐ Adverse inference (where appropriate)

☐ Credibility concern

☐ Exclusion

☐ Other procedural or remedial order

# 14. CHECKLIST FOR AI-ASSISTED INDIRECT EVIDENCE

## 14.1 IDENTIFY THE NATURE AND EXTENT OF AI USE

☐ Has AI assistance been disclosed?

☐ What AI tool was used?

☐ How was the AI used?

- ☐ Drafting
- ☐ Editing
- ☐ Summarization
- ☐ Translation
- ☐ Research
- ☐ Data analysis
- ☐ Other

☐ Was AI used for all or only part of the document?

☐ Is the extent of AI involvement adequately explained?

## 14.2 DETERMINE WHO IS STANDING BEHIND THE EVIDENCE

☐ Has an identifiable witness adopted the document?

☐ Has the witness sworn or affirmed its contents?

☐ Is the witness prepared to accept responsibility for the entire document?

☐ Can the witness distinguish between personal knowledge and AI-generated content?

☐ Is there any uncertainty regarding authorship or responsibility?

## 14.3 ASSESS PERSONAL KNOWLEDGE

☐ Does the witness have personal knowledge of the facts asserted?

☐ Can the witness explain the basis for key factual statements?

☐ Can the witness explain conclusions or opinions contained in the document?

☐ Does the witness appear to understand the contents independently of the AI-generated text?

☐ Are there indications that the witness adopted language or conclusions without verification?

## 14.4 CONSIDER RELIABILITY CONCERNS

☐ Are there indications of:

- ☐ Hallucinated facts
- ☐ Fabricated citations
- ☐ Incorrect quotations
- ☐ Unsupported assertions
- ☐ Mischaracterizations of evidence
- ☐ Omissions of material facts

☐ Has the document been independently reviewed and verified?

☐ Are source materials available for comparison?

☐ Do other portions of the evidence corroborate the statements made?

## 14.5 ASSESS THE ABILITY TO TEST THE EVIDENCE

☐ Can opposing parties effectively cross-examine the witness concerning the contents?

☐ Can the witness explain how particular statements came to appear in the document?

☐ Can the witness identify what was contributed by AI and what originated from the witness?

☐ Has AI use impaired the ability to assess credibility or reliability?

☐ Does the AI involvement create uncertainty regarding whose evidence is actually being presented?

## 14.6 CONSIDER EXPERT-SPECIFIC ISSUES (IF APPLICABLE)

☐ Has the expert disclosed the use of AI tools?

☐ Was AI used merely as an aid, or did it contribute substantively to the analysis?

☐ Can the expert explain and defend the reasoning process independently of the AI tool?

☐ Would the opinion remain the same absent the AI-generated content?

☐ Has the expert exercised independent judgment?

## 14.7 DETERMINE APPROPRIATE JUDICIAL RESPONSE

☐ Require additional disclosure regarding AI use

☐ Require production of drafts or source materials

☐ Permit further cross-examination

☐ Conduct a voir dire

☐ Restrict use of portions of the document

☐ Assign reduced weight

☐ Exclude portions of the evidence

☐ Other procedural direction