

ARE THE MACHINES IN CHARGE?

Using AI to Review Trusts and Other Things

Mason Clark

Assistant Professor of Law, Englehardt Research Fellow

St. Mary's University School of Law

Mclarkr16@stmarytx.edu

2026 National Conference on Special Needs Planning
and Special Needs Trusts

Thursday, October 22, 2026

I. INTRODUCTION AND SCOPE

These materials¹ accompany the session “Are the Machines in Charge? Using AI to Review Trusts and Other Things.” They are intended to stand on their own as a reference for practitioners evaluating whether, and how, to use generative artificial intelligence in the review of trust instruments and related estate planning documents.

The scope is deliberately narrow. The session addresses capability and verification, and it asks: What can these tools can do? How can they fail? What should a working verification protocol look like?

The professional responsibility framework, the ethical analysis, the problem of bias against beneficiaries with disabilities, and the profession’s broader response to AI are addressed in the companion materials prepared for the afternoon breakout session, “Navigating AI’s Wild West: Trends, Ethics, and Backlash in Legal

¹ As many jurisdictions now require attorneys to provide disclosure and/or verification statements for documents prepared with assistance from or containing information generated by generative artificial intelligence, it seems fitting to include a specific and conspicuous disclosure and verification statement here: These materials were prepared with the assistance of generative artificial intelligence. The author used various large language models to organize some parts and subsections; to rephrase some content for brevity; to format citations; to summarize lengthy reports or articles, particularly because they pertain to an area of the law that is NOT the author’s area of expertise; and to compile research material. The author certifies that all citations have been manually verified to ensure their existence and confirm their propositions as accurate representations.

However, in Part V.A, the author intentionally included unverified citations to certain changes in the law impacting conference participants. The author invites participants to find incorrect AI outputs!

Practice.” To avoid duplication, those subjects are cross-referenced here rather than repeated.

The organizing question is not whether artificial intelligence is useful. It plainly is . . . for some tasks. The question is: for which tasks is the error rate acceptable, given the cost of an error that goes undetected? In special needs practice that cost is unusually high.

II. WHAT THESE TOOLS ACTUALLY DO

A. Prediction v. Retrieval

A large language model generates text by predicting probable continuations of the input it is given. It does not look up authority unless it has been connected to a database and instructed to do so. This matters because the mechanism that produces an accurate summary of Article IV is the identical mechanism that produces a fabricated subsection of the POMS. The practical consequence is that apparent fluency carries no information about accuracy. Practitioners accustomed to evaluating written work product by its polish—a reasonable heuristic when the author is a human partner or associate whose confidence roughly tracks competence.

B. Three Architectures

Practitioners encounter at least three distinct configurations, whose reliability differs substantially. Vendor marketing frequently blurs the distinctions.

1. **General-purpose chat interfaces.** The model responds from its training data alone. No source grounding. Highest fabrication risk, particularly for citations.
2. **Retrieval-augmented legal research tools.** The model is constrained to a curated database and returns citations to source documents the user can open. Fabrication risk is materially lower but not zero; the characterization of a retrieved authority may still be wrong.
3. **Document-grounded review.** The corpus is the instrument the practitioner uploaded. The model's task is extraction and analysis of text in front of it. This is the configuration most relevant to trust review and, for extraction tasks, the most reliable of the three. But the model will still supply legal conclusions from its training data, and those conclusions are ungrounded even though the document analysis is not.

A practitioner who does not know which of these three they are using cannot calibrate their skepticism.

C. Non-Determinism

The same prompt submitted twice will not produce identical output. For reliability this is a defect, but it can also aid in verification. Where two runs of the same prompt diverge, the model is generating rather than reporting, and the divergent passage warrants double-checking.

III. A TASK TAXONOMY FOR TRUSTS AND ESTATES

Organizing by task and not by tool is helpful. The following tiers track the distance between the output and the source text. The further the model must travel from language actually in front of it, the higher the risk.

A. Tier One (Suitable for a First Pass)

Summarizing a lengthy instrument; extracting dispositive terms into a chart; constructing a chronology from a file; translating trust terms into plain language for a family or a beneficiary; comparing two versions of a document; drafting routine correspondence; applying a checklist the practitioner wrote herself. These are extraction and reorganization tasks. The answer is already in the document.

B. Tier Two (Useful, But Requires Independent Review)

Consistency checking across a trust, will, beneficiary designations, and powers of attorney; identifying provisions missing relative to a form; issue-spotting an instrument the practitioner did not draft; generating questions for a client interview. Here, the model can be helpful because checking cross-document consistency is tedious work that humans perform poorly late in the day (or after lunch!), but the output is a range of useful information and not necessarily a conclusion.

C. Tier Three (High Risk)

Conclusions regarding SSI or Medicaid eligibility; application of the POMS; state-specific payback mechanics and state agency practice; early termination analysis; tax characterization. These require the model to supply law from training

data of unknown vintage, apply it to facts, and reach a conclusion. Each step compounds the error rate of the last, and the output is delivered with no indication of accumulated uncertainty.

IV. WHY SPECIAL NEEDS PRACTICE INVERTS THE USUAL DRAFTING NORM

Special needs practice presents a risk that general AI guidance does not anticipate because the drafting conventions look like mistakes to a general-purpose model. They're often the minority pattern in its training data.

A. Discretionary Standards

Ordinary trust drafting favors an ascertainable standard. Health, education, maintenance, and support language appears in an enormous volume of trust instruments, treatises, and form books. Special needs drafting often requires the opposite. A model trained predominantly on ordinary drafting will therefore identify a fully discretionary standard as a drafting defect and helpfully propose to cure it by adding a support standard. This is not a “hallucination” in the usual sense.² The model is accurately reproducing the dominant pattern in its training data. It is simply the wrong pattern for this instrument, and following the recommendation can defeat the trust.

² We as attorneys should also be cognizant of anthropomorphizing AI. See Eric Sentell, *We Must Stop Anthropomorphizing AI*, SUBSTACK (Sept. 23, 2026), <https://ericsentell.substack.com/p/we-must-stop-anthropomorphizing-ai>.

B. Sole Benefit Restrictions

Sole benefit language reads to a general-purpose model as an unreasonable limitation on the trustee, of the kind that should be relaxed. The same dynamic applies to spendthrift and anti-compulsion provisions.

C. First-Party and Third-Party Trusts

Both trust types appear under similar headings across the secondary literature, and models collapse them with some regularity. The consequences run in both directions. On the one hand, advising that a self-settled trust needs no payback provision would render it countable; on the other, advising that a third-party supplemental needs trust should include one, for consistency, gratuitously surrenders family assets to the state Medicaid agency at the beneficiary's death.

The unifying point is that the model's instinct to improve the document is itself the hazard. The provisions it flags as defects are frequently the provisions that are intentionally nuanced and/or unusual in other fields of the law.

V. RECURRING FAILURES

A. Stale Law

A model reports the law as of some point in the past and will not volunteer which point. Three areas of special needs practice have changed recently enough to make this a live hazard.

1. Establishment by the Individual

The Special Needs Trust Fairness Act, enacted as part of the 21st Century Cures Act, Pub. L. No. 114-255, § 5007, 130 Stat. 1033 (2016), amended 42 U.S.C. § 1396p(d)(4)(A) to permit a competent adult beneficiary to establish her own first-party trust.³ Models continue to recite the pre-2016 list (i.e., parent, grandparent, legal guardian, or court) because that formulation was correct for roughly two decades and appears throughout the older secondary literature. The resulting advice generates unnecessary court petitions, client expense, and delay to cure a defect that does not exist.

2. In-Kind Support and Maintenance

The Social Security Administration's final rule omitting food from ISM calculations took effect September 30, 2024.⁴ Only shelter-related items are now counted.⁵ Models routinely advise trustees to avoid paying for groceries or restaurant meals in order to protect a benefit that is no longer at risk—a restriction on the beneficiary's quality of life grounded in repealed law.

For 2026, the federal benefit rate is \$994 for an individual and \$1,491 for an eligible couple; the presumed maximum value cap is \$351.33; the value of the one-

³ Ctr. for Medicaid & CHIP Servs., Ctrs. for Medicare & Medicaid Servs., SMD # 17-001, Implications of the Cures Act for Special Needs Trusts (Aug. 2, 2017), <https://www.medicaid.gov/federal-policy-guidance/downloads/smd17001.pdf>.

⁴ Omitting Food From In-Kind Support and Maintenance Calculations, 89 Fed. Reg. 21,199 (Mar. 27, 2024) (codified as amended at 20 C.F.R. pt. 416).

⁵ 20 C.F.R. § 416.1130(b) (2025) (defining shelter).

third reduction is \$331.33. Practitioners should confirm current figures against the annual cost-of-living adjustment before relying on any of them.⁶

3. *ABLE Accounts*

The ABLE Age Adjustment Act took effect January 1, 2026, extending eligibility under 26 U.S.C. § 529A to individuals whose disability had onset before age 46, rather than age 26.⁷ The 2026 annual contribution limit is \$20,000.⁸ Up to \$100,000 remains excluded from the SSI resource limit.⁹ Models commonly report the pre-2026 age threshold, understate the contribution limit, and — more seriously — state that ABLE balances are free of Medicaid recovery. Federal law permits a state claim against the account at the beneficiary’s death;¹⁰ a number of states have elected not to pursue it. That state-by-state variation is the kind of nuance a model flattens into a false universal.

B. Fabricated and Miscited Authority

Fabricated citations are the best-known failure mode, and the reported decisions are now numerous.¹¹

⁶ *Id.* §§ 416.1131–416.1133 (one-third reduction rule); *id.* §§ 416.1140–416.1145 (presumed value rule).

⁷ SECURE 2.0 Act of 2022, Pub. L. No. 117-328, div. T, § 124 (2022) (codified at 26 U.S.C. § 529A(e)(1)).

⁸ 26 U.S.C. § 529A(b)(2)(B)(i); Rev. Proc. 2025-32, 2025-44 I.R.B. 1.

⁹ Stephen Beck, Jr., Achieving a Better Life Experience Act of 2014, Pub. L. No. 113-295, div. B, § 103, 128 Stat. 4010 (2014); POMS SI 01130.740(D)(1).

¹⁰ 26 U.S.C. § 529A(f); Ctr. for Medicaid & CHIP Servs., Ctrs. for Medicare & Medicaid Servs., SMD # 17-002, Implications of the ABLE Act for State Medicaid Programs (Sept. 7, 2017), <https://www.medicaid.gov/federal-policy-guidance/downloads/smd17002.pdf>.

¹¹ For two well-known examples, see *Mata v. Avianca, Inc.*, 678 F. Supp. 3d 443 (S.D.N.Y. 2023) (sanctioning counsel for a filing containing fictitious AI-generated citations); *Park v. Kim*, 91 F.4th 610 (2d Cir. 2024) (referring counsel for discipline on similar facts).

First, fabricated authority does not look strange. It carries a correct court abbreviation, a plausible reporter volume, a plausible date, and a proposition that fits the argument. Second, a misattributed citation to a real authority is more dangerous than an invented one. A practitioner who recognizes a POMS section number as genuine will often let recognition substitute for verification, and the output attached to it goes unchecked.

C. Invented Rules Without Citation

The hardest failure mode to catch is a confidently stated, state-specific rule supported by no citation at all: For example, a supposed requirement that a trustee obtain court approval for distributions above a specified dollar threshold. Out-of-state practitioners cannot evaluate it. In-state practitioners may assume they have forgotten something. There is no citation to pull.

VI. A WORKING VERIFICATION PROTOCOL

A. Citation Discipline

No citation is accepted that has not been personally opened in the reporter, the statute, or the POMS. This rule admits no exceptions and does not bend for authority that looks familiar.

B. Force the Model Back into the Document

For document review, require the model to quote the operative language verbatim and identify where it appears. Then compare the quotation against the

instrument. This converts an ungrounded assertion into a checkable one and surfaces the cases in which the model is describing a provision that is not there.

C. Re-Run and Compare

Submit the identical prompt in a fresh session and compare the outputs. Passages that are stable across runs are generally grounded in the source. Passages that vary are being generated. This is a low-cost triage mechanism that tells the reviewer where to spend attention.

D. Adversarial Second Pass

Ask the model to argue against its own output: what did you miss; argue that this provision is defective; identify what a state Medicaid agency reviewer would object to. Models are markedly better at critique than at first-pass analysis, and the second pass frequently surfaces what the first omitted.

E. Sequence Matters

AI used as a second reader, after the practitioner's own review, catches the practitioner's misses. AI used as a first reader anchors the practitioner's judgment to the model's framing, and the anchoring effect is substantial. The same tool produces materially different value depending only on when in the workflow it is introduced.

F. Document the Verification

A contemporaneous file note recording what was AI-assisted and what was independently verified costs very little and is the practitioner's best evidence in a later malpractice or surcharge proceeding.

VII. THE OMISSION PROBLEM

Every element of the protocol in Part VI checks what the model said. None of them checks what the model did not say, and omissions are invisible by construction.

Consider a review of a first-party trust funded with a substantial personal injury settlement. A model asked to identify defects will generate a list of defects. It will not reliably raise whether a Medicare Set-Aside or conditional payment resolution was required, whether the settlement was properly allocated among lien claimants, whether the instrument is irrevocable on its face, whether the state agency and the Social Security Administration have reviewed it, whether the beneficiary has capacity, whether the trustee is bonded, what the successor trustee provisions say, or whether anyone asked the beneficiary what she wants. None of these will appear as a gap in the output, because the output contains no gaps—it is a complete-looking list of the issues the model happened to generate.

A practitioner working from the model's issue list inherits the model's blind spots and has no way to detect them. So, generate your own issue list first, then compare. What the model raised that you did not is *the value*. What you raised that the model did not is the measure of *how much you should have trusted it*.

VIII. A DECISION RULE

A single principle organizes everything above: Trust the machine where errors are cheap to detect and cheap to reverse.

Much of ordinary practice satisfies both conditions. A garbled summary is obvious on reading and costs nothing to regenerate. A clumsy first draft of a client letter is visibly clumsy.

Special needs work frequently fails both. An ascertainable standard inserted into a discretionary trust, a payback provision added to a third-party instrument, an unprotected addition after age 65, a defective early termination clause . . . none of these announces itself. Each may remain undetected until a redetermination, an agency review, or the beneficiary's death, by which point the error is often irreversible and the consequence falls on a person with limited ability to detect or contest it.

That asymmetry—and not just generalized skepticism about the technology—is the reason this practice area warrants heightened care. The same tool that is appropriately used with light verification in a commercial matter requires a different discipline here, because the cost of being wrong is borne by someone else and surfaces too late to fix.

IX. CONCLUSION

These tools are genuinely useful for extraction, summarization, translation, and cross-document consistency checking, and practitioners who refuse them entirely will do some tasks worse and slower than colleagues who adopt them

thoughtfully. They are unreliable for legal conclusions in an area whose governing authority changes frequently, whose drafting conventions invert the general pattern, and whose errors are detected late. The professional obligations that attach to that judgment, and the broader question of what the profession owes clients and beneficiaries when it uses these systems, are taken up in the companion materials for the afternoon session.

See Appendix for Verification Checklist

APPENDIX — VERIFICATION CHECKLIST

1. Identify which architecture produced the output: ungrounded chat, retrieval-augmented research tool, or document-grounded review.
2. Open every citation personally. No exceptions for familiar authority.
3. Require verbatim quotation of operative language and confirm it against the instrument.
4. Re-run the identical prompt in a fresh session; treat divergence as a flag.
5. Run an adversarial pass asking the model to argue against its own conclusions.
6. Independently verify any statement of current benefit figures, contribution limits, or age thresholds.
7. Treat every uncited, state-specific assertion as unverified until confirmed in primary authority.
8. Generate your own issue list before reading the model's, and compare the two.
9. Confirm the trust type — first-party or third-party — before accepting any payback analysis.
10. Record in the file what was AI-assisted and what was independently verified.