

IN THE UNITED STATES DISTRICT COURT
FOR THE DISTRICT OF MINNESOTA

CHRISTOPHER KOHLS and MARY
FRANSON,

Plaintiffs,

v.

KEITH ELLISON, in his official capacity
as Attorney General of Minnesota, and
CHAD LARSON, in his official capacity
as County Attorney of Douglas County,

Defendants.

Court File No. 0:24-cv-3754-LMP-DLM

**PLAINTIFFS' MEMORANDUM OF
LAW IN SUPPORT
OF DAUBERT MOTION TO
EXCLUDE EXPERT DECLARATIONS**

INTRODUCTION

Defendant Ellison filed two expert declarations he contends demonstrate Minn. Stat. §609.771 “actually necessary” and counterspeech “insufficient” to address AI-generated deepfakes.

But the Declaration of Prof. Jeff Hancock cites a *study that does not exist*. No article by the title exists. The publication exists, but the cited pages belong to unrelated articles. **Likely, the study was a “hallucination” generated by an AI large language model like ChatGPT.** A part-fabricated declaration is unreliable.

Separately, both declarations rely on *ipse dixit* conclusory legal opinion with no evident methodology, and these paragraphs should be excluded under *Daubert*.

TABLE OF CONTENTS

INTRODUCTION	1
TABLE OF CONTENTS.....	2
ARGUMENT	3
I. The “hallucination” in Hancock’s declaration calls the entire document into question; at minimum, the fabrication warrants exclusion.	4
A. The AI-imagined “study” calls the reliability of the entire declaration into question.	7
B. The Court should consider additional actions.	8
II. Conclusory opinions are not proper expert testimony.....	9
A. Hancock’s testimony and conclusory opinions.....	11
B. West’s testimony and conclusory opinions.	17
C. The conclusory material in both declarations, including their final paragraphs, does not constitute reliable expert evidence because it was not produced by a reliable methodology.....	20
D. The conclusory material should be excluded as legal opinion.	22
III. The experts’ other publications undermine their conclusory paragraphs. ...	25
A. In his cited editorial, Hancock is more equivocal about the potential harm of deepfakes.	25
B. In his book, West supports counterspeech to combat misinformation and notes the vague line between parody and misinformation.....	27
IV. If not excluded, the Court should only consider the declarations, not exaggerations contradicted by them.	30
CONCLUSION.....	35

ARGUMENT

Defendant Attorney General Keith Ellison's Memorandum Opposing Preliminary Injunction, Dkt.19 ("Ellison"), cites and relies on two expert declarations spanning 28 pages, and filed along with his 11,737-word opposition. The Declarations of Prof. Jeff Hancock, Dkt.23 ("Hancock Decl.") and Prof. Javin West, Dkt.24 ("West Decl.") purport to cite studies showing the dangers of deepfakes, offer policy recommendations with no apparent connecting analysis, and legal conclusions that counterspeech is "insufficient" to combat AI deepfakes.

Ironically, the Hancock declaration, a report submitted by an expert on "misinformation and deepfakes" (Hancock Decl. ¶1), *cites a study that does not exist*. As explained below, the citation bears the hallmarks of being an artificial intelligence (AI) "hallucination," suggesting that at least the citation was generated by a large language model like ChatGPT. Plaintiffs do not know how this hallucination wound up in Hancock's declaration, but it calls the entire document into question, especially when much of the commentary contains no methodology or analytic logic whatsoever. At minimum, the Hancock Declaration should be excluded from consideration in resolving Plaintiffs' motion for preliminary injunction. Further actions may be prudent to protect other courts, in

which Hancock testifies, from the “risks associated with deepfakes.” Hancock Decl. ¶32.

Independently, the unsupported conclusory paragraphs offering *de facto* opinions of law should be excluded from each declaration. These paragraphs include Hancock Decl. ¶¶24, 26-28, 30, 32 and West Decl. ¶23.

While the professors are well-published in their fields, their bottom-line conclusions do not derive from any apparent methodology, so they are improper expert opinion. Further, the opinions attempt to overturn binding precedents, and testimony about legal conclusions is inappropriate because every court comes equipped with a legal expert—the judge. Additionally, cited references and other publications by these experts, attached as exhibits to the contemporaneously-filed Declaration of M. Frank Bednarz, contradict their conclusory opinions. Finally, even if they are not excluded, the Court should confine the conclusory opinions to the statements actually contained within the declarations, and not exaggerated characterizations introduced in Ellison’s opposition.

I. The “hallucination” in Hancock’s declaration calls the entire document into question; at minimum, the fabrication warrants exclusion.

Hancock cites an imaginary study in paragraph 21 of his declaration, which purports to describe persistently harmful effects from deepfakes. It reads:

Moreover, the difficulty in disbelieving deepfakes stems from the sophisticated technology used to create seamless and lifelike reproductions of a person's appearance and voice. One study found that even when individuals are informed about the existence of deepfakes, they may still struggle to distinguish between real and manipulated content. This challenge is exacerbated on social media platforms, where deepfakes can spread rapidly before they are identified and removed (Hwang et al., 2023).

Hancock Decl. ¶21.

Hancock declared under penalty of perjury: “I have further identified the academic, scientific, and other materials referenced in this declaration in the references attached as Exhibit C.” *Id.* ¶4. This document identifies the cited reference as follows:

Hwang, J., Zhang, X., & Wang, Y. (2023). The Influence of Deepfake Videos on Political Attitudes and Behavior. *Journal of Information Technology & Politics*, 20(2), 165-182.
<https://doi.org/10.1080/19331681.2022.2151234>

Dkt. 23-1 at 39.

The “doi” url is supposed to be a “Digital Object Identifier,” which academics use to provide permanent links to studies. Such links normally redirect users to the current location of the publication, but a DOI Foundation error page appears for this link: “DOI NOT FOUND.” Bednarz Ex. 1.

Perhaps this was simply a copy-paste error? It’s not.

The article doesn’t exist.

The title of the alleged article, and even a snippet of it, does not appear on anywhere on the internet as indexed by Google and Bing, the most commonly-used search engines.¹ Searching Google Scholar, a specialized search engine for academic papers and patent publications, reveals no articles matching the description of the citation authored by “Hwang” that includes the term “deepfake.” Bednarz Ex. 2.

The JOURNAL OF INFORMATION TECHNOLOGY & POLITICS exists. But the cited pages, 165-182, at Volume 20(2), do not contain the article. Instead, these pages contain the last part of Meeks (2023), *Blue bird in a coal mine: How 2020 Democratic presidential candidates framed climate change on Twitter*, 20(2), 154-168 and almost the entirety of Kartsounidou et al. (2023), *Measuring the impact of candidates’ tweets on their electoral results*, 20(2), 169-183. See Bednarz Ex. 3.

This sort of citation—with a plausible-sounding title, alleged publication in a real journal, and fictitious “doi,” is characteristic of an artificial intelligence “hallucination,” which academic researchers have warned their colleagues about.

1

See search results for Google:
<https://google.com/search?q=%22influence+of+deepfake+videos+on+political%22;>
and Bing:
<https://bing.com/search?q=%22influence+of+deepfake+videos+on+political%22.>

See Goddard, J, *Hallucinations in ChatGPT: A Cautionary Tale for Biomedical Researchers* (2023), AM. JOURNAL OF MED., 136(11), 1059-1060, Bednarz Ex. 4.

A. The AI-imagined “study” calls the reliability of the entire declaration into question.

Plaintiffs may argue that Hancock could have cited a real study similar to the proposition in paragraph 21. But the existence of a fictional citation Hancock (or his assistants) didn’t even bother to click calls into question the quality and veracity of the entire declaration. This is especially true given that the conclusions that Ellison most relies on have no methodology behind them and consist entirely of expert say-so. *See* Section II, below.

The purpose of the Court’s gatekeeping function under *Daubert* “is to make certain that an expert, whether basing the testimony upon professional studies or personal experience, employs in the courtroom the same level of intellectual rigor that characterizes the practice of an expert in the relevant field.” *Kumho*, 526 U.S. 137, 152 (1999). Citing to a completely fictitious study fall well below this standard. “Bottom line—in its current state of development, ... researchers should NOT ask ChatGPT for sources, references, or citations on a particular topic. Or, if they do, all such references should be carefully vetted for accuracy.” Bednarz Ex. 4.

The invention suggests that the report did not “flow[] from the expert’s research” but was merely “developed for litigation.” *Lauzon v. Senco Prods.*, 270 F.3d 681, 687 (8th Cir. 2001). It is not merely an error; it suggests other parts of the declaration may be similarly fabricated, but undetected. The uncited assertions of the declaration cannot be compared with a publication’s table of contents to verify their authenticity.

B. The Court should consider additional actions.

AI-generated hallucinations have sometimes appeared in legal filings, usually by *pro se* litigants who apparently asked ChatGPT to write them a brief.² When lawyers have submitted hallucinated legal citations, the signed filings containing them have merited sanctions proceedings. *See, e.g., In re Prob. Proceeding*, 2024 NYLJ LEXIS 238 (N.Y. Sur. Ct. Kings Cnty Jan. 26, 2024). However, Plaintiffs do not know whether counsel or Hancock introduced this error.

If counsel inserted the fictional citation as a late revision, Plaintiffs intend to move for sanctions. If this occurred, Hancock would have still submitted a declaration where he falsely represented to have reviewed the cited material, and

² *See* Eugene Volokh, *Six Federal Cases of Self-Represented Litigants Citing Fake Cases in Briefs, Likely Because They Used AI Programs*, THE VOLOKH CONSPIRACY (Nov. 13, 2023), <https://reason.com/volokh/2023/11/13/self-represented-litigants-use-ai-to-write-briefs-produce-hallucinated-citations/>.

this militates in favor of excluding the entire declaration in any event, but an attorney-introduced error may be isolated to this proceeding.

If, on the other hand, Hancock (or assistants under his supervision) generated the content, this seems to be a more serious matter to pending litigation. Hancock reports serving as an expert witness in twelve different matters in the last five years. Dkt. 23-1 at 34-35. His declaration speaks of “combat[ting] the risks associated with deepfakes.” Hancock Decl. ¶32. While Plaintiffs believe that the First Amendment ought not be trampled to protect citizens against speech, truth-seeking courts and tribunals should guard against potential falsehoods. If Hancock originated the hallucination, this Court might order him to inform the parties and courts in any pending litigation about the possibility his testimony may contain content generated by a large language model. The Court has inherent authority to prevent fraud by the parties appearing before it. *Cf. Chambers v. NASCO, Inc.*, 501 U.S. 32, 45 (1991).

Plaintiffs reserve all rights.

II. Conclusory opinions are not proper expert testimony.

This Court is the gatekeeper for expert testimony. *See, e.g., Daubert; see also* Fed. R. Evid. 702 Advisory Committee Notes. The Federal Rules of Evidence apply

to pre-trial proceedings including when hearing a motion for preliminary injunction. *See* Fed. R. Evid. 1101(b).

Ellison argues (at 34) that “multiple experts agree that counterspeech is insufficient to combat the unique harms that deepfakes inflict.” But Ellison makes no effort to qualify the declarations as expert testimony employing a reliable methodology. The proponent of the expert testimony bears the burden to prove its admissibility by a preponderance of the evidence. *Daubert v. Merrell Dow Pharm., Inc.*, 509 U.S. 597, 592 n.10 (1993).

“Under *Daubert*, district courts must make a preliminary assessment of whether the reasoning or methodology underlying an expert’s testimony is scientifically valid and of whether that reasoning or methodology properly can be applied to the facts in issue.” *Lancaster v. BNSF Ry. Co.*, 75 F.4th 967, 969 (8th Cir. 2023) (cleaned up).

The conclusory opinions in declarations do not make it this far because their essentially-legal conclusions employ no discernable methodology at all. The “hallmark for reliability of an expert’s opinion is the scientific method, i.e., the generation of testable hypotheses that are then subjected to the real world crucible of experimentation, falsification/validation, and replication.” *Somnis v. Country*

Mut. Ins. Co., 840 F. Supp. 2d 1166, 1172 n.2 (D. Minn. 2012) (cleaned up). Nothing like this exists in the declarations.

Each summarizes generative AI and recount problems of online information before declaring conclusions that First Amendment caselaw favoring counterspeech effectively no longer applies.

A. Hancock’s testimony and conclusory opinions.

Hancock describes generative AI and its increasing uses. Hancock Decl. ¶¶7-10. He describes deepfakes and how they are “significantly more challenging to detect without specialized tools.” *Id.* ¶¶11-13. Deepfakes are more widely available raising “serious concerns due to [their] potential misuse in spreading disinformation and manipulating public perception.” *Id.* ¶¶14-15. “Deepfakes contribute to the spread of misinformation, create confusion about the authenticity of legitimate content, and potentially harm individuals’ reputations and wellbeing.” Hancock Decl. ¶15.

But Hancock says that in response to these concerns “researchers and platforms are working to develop advanced detection technologies and establish policy frameworks to counter the rise of deepfakes.” Hancock Decl. ¶16.

Deepfakes are transforming “the landscape of misinformation” due to reduced costs, personalization, difficulty in detection, and higher persuasive

power of audio and video. *Id.* ¶17. Hancock cites studies purporting to show that deepfake videos are more likely to be believed, alter political beliefs with convincing narratives, and rapidly spread, but “People are more likely to doubt the authenticity of a political video if the content is inconsistent with their perceptions of the politician's typical behavior or known viewpoints.” *Id.* ¶¶19-22.

At this point, Hancock begins making unsupported conclusions. He says, without citation: “Traditional fact-checking methods are less effective in combating deepfakes due to the sophisticated and deceptive nature of these manipulated audio and video files.” Hancock Decl. ¶24. He continues that factchecking finds it difficult to deal with the “rapid speed at which [deepfakes] can spread on social media platforms, often outpacing traditional forms of content due to their sensational and visually compelling nature.” *Id.* ¶25. He describes the @deptomcruise account, which allegedly made it “difficult for viewers to

distinguish these videos from authentic footage” and amassed a large following within days. *Id.*³

Hancock reports that labels on deepfakes make people “more likely to be skeptical of it, suggesting labels can be effective.” Hancock Decl. ¶26. But then, without citation, he says labelling “can also” produce “‘liar's dividend,” where individuals exploit the existence of deepfakes to deny the authenticity of real events by claiming they are fabricated.” *Id.* ¶27. This conclusion does not follow, and Hancock does not explain how he arrived at it. Deepfakes will continue to exist whether some are labelled or not, and whether genuine content is mislabeled as deepfakes or not. Actors already implausibly claim that genuine content was created by AI. See Benj Edwards, *NC governor candidate cries AI fabrication as defense for racist porn forum posts*, ArsTechnica (Sep. 20, 2024).⁴

³ This example is curious, because the @deeptomcruise account posted videos labelled as deepfakes and for the purpose of parody. Rachel Metz, *How a deepfake Tom Cruise on TikTok turned into a very real AI company*, CNN (Aug. 6, 2021), <https://www.cnn.com/2021/08/06/tech/tom-cruise-deepfake-tiktok-company/index.html> (article cited by Hancock Decl. ¶ 25). This shows that parody deepfake content, which Defendants would atextually carve out from the statute’s ambit, has some of the same alleged problems as harm-causing deepfakes.

⁴ Available at: <https://arstechnica.com/information-technology/2024/09/nc-governor-candidate-cries-ai-fabrication-as-defense-for-racist-porn-forum-posts/>.

Hancock opines that deepfakes have a “potential to create false memories” (Hancock Decl. ¶ 28), and that repeated exposure to deepfakes can lead to the “Illusory Truth Effect.” *Id.* ¶29.⁵ “By making false political information more believable, deepfakes can potentially influence public opinion and electoral outcomes, posing a threat to democratic institutions and processes.” *Id.* ¶30.

Like most of the declaration, Hancock’s penultimate paragraph characterizes deepfakes as a challenge, but this does not suggest it to be an insurmountable one. He says that “deepfakes present *significant challenges* to the integrity of information” because they “*can* spread rapidly on social media” and “*can* manipulate perceptions, create false memories, and exploit cognitive biases, *making it difficult* for individuals to discern truth from falsehood.” Hancock Decl. ¶31 (all emphasis added).

But the final paragraph introduces a firm, but unexplained, conclusion:

Traditional methods of fact-checking and counterspeech are insufficient to address the sophisticated and rapidly spreading nature of deepfakes. Regulatory measures, combined with technological solutions and public awareness efforts, are necessary to combat the risks associated with deepfakes. In my expert view, enacting and enforcing laws that specifically target the production and

⁵ As discussed below in section IV, the cited publications replicate studies performed with other forms of misinformation years and decades earlier, before the widespread availability of generative AI; these characteristics are not unique to deepfakes.

dissemination of deceptive deepfake content during elections are critical in preserving the integrity of democratic institutions and protecting the foundational trust upon which they rely.

Hancock Decl. ¶ 32.

This conclusion, a policy recommendation or statement of law, does not follow from the foregoing evidence.⁶ It evinces no method for reaching the conclusion that counterspeech is “insufficient.” Hancock Decl. ¶ 32. Nor does Hancock explain how he determined that “enacting and enforcing laws” is “critical in preserving the integrity of democratic institutions.” *Id.* The lack of explanation is not too surprising because Hancock’s bare conclusion is ultimately a legal one, reserved for the Court’s determination.

While Hancock’s declaration shows that some deepfake speech may raise “serious concerns,” this is true about a lot of speech, and does not result in the conclusion that criminalization is “necessary.”

Speech advocating racial animus is problematic, challenging, harmful, and can spread rapidly on social media. For example, photos and videos of unrelated individuals of color apparently engaged in slaughtering a cat (due to mental illness) and apparently carrying dead wildlife resulted in a politically-charged

⁶ Hancock was much more equivocal in his cited editorial, Hancock & Bailenson, discussed in Section IV.A below.

moral panic about Haitian immigrants and their supposed antisocial behavior in Springfield, Ohio—even though none of the photos or videos were taken in that city or, apparently, even depicted Haitian immigrants. Due to social media (not because the videos were deepfakes), these spread faster than they could be rebutted. These fears spread widely on social media, and viewers, including President Trump, persistently believed them. Melissa Gira Grant, *How Lies About Pet Eating Turned Into Bomb Threats*, THE NEW REPUBLIC (Sep. 19, 2024).⁷

Under what method, then, does Hancock conclude that laws protecting “deceptive deepfake content during elections” are “critical” and counterspeech “ineffective”? Are other sorts of lies incorporating “realistic” video (which is realistic because *it is real*, just misrepresented) not similarly convincing, rapidly-spread, and politically influential? Is it less “critical” for Haitians to be protected from deceptive video than politicians? Because Hancock shows no method for reaching his conclusion, it’s impossible to determine, or even replicate his conclusion concerning political deepfakes. This testimony should be rejected because it “employed no methodology whatsoever — reliable or otherwise.” *Ahlberg v. Chrysler Corp.*, 481 F.3d 630, 635 (8th Cir. 2007) (affirming exclusion).

⁷ Available at: <https://newrepublic.com/article/186149/trump-springfield-haitian-immigrants-pets-origin>.

B. West's testimony and conclusory opinions.

West similarly describes the history and increasing use of deepfakes and a broader category of AI-generated works, synthetic media. West Decl. ¶¶7-9. Citing studies, he reports that: "Researchers have shown that deepfakes, especially those microtargeted to specific groups, can impact people's attitudes towards a politician and can impact trust in news." *Id.* ¶10. Increasingly, deepfakes cannot be reliably detected. *Id.* ¶¶11-12. Deepfakes "spread most effectively, as does most information on the internet, through social media platforms." *Id.* ¶13.

While Hancock argued that deepfakes were likely to be believed because of their realistic audio-visual nature, West describes the social media phenomenon of "participatory disinformation." West Decl. ¶14. "Users can adapt a video or image for a given community or context. This is what makes it so difficult for fact checkers and journalists trying to correct the record. It becomes embedded so quickly with buy in from the community, who are contributors and sharers themselves, that it is difficult to debunk and slow." *Id.*

West observes that "platform spread is a challenge to contain, *not just for deepfakes, but for any kind of content*, especially disinformation campaigns." West Decl. ¶15 (emphasis added). Even when one platform identifies and

suppresses an article of misinformation, it can continue to spread on another platform. *Id.*

West writes that “Deepfakes generally are designed to go viral. ... They are visually and auditorily engaging; they are plausibly realistic, so much so that users often cannot tell whether they are real or not; and, importantly, they often carry shock value that fits within a larger cultural narrative, of something that many think ‘could happen.’” West Decl. ¶¶16, 21. For this reason, deepfakes “often convey a scenario that aligns with the prevailing narrative.” *Id.* ¶17. “And even when they fact checked, deepfakes can leave a lasting effect that deepens distrust in our information systems.” *Id.*

West also describes the “Liar’s Dividend,” where actors may claim real content to be faked. West Decl. ¶18. “As we proliferate our information systems with more and more synthetic, but realistically-looking, content, the more useful and potentially dangerous the Liar’s Dividend.” *Id.*

West describes political deepfakes around the globe, “often introduced at the ‘eleventh hour’, or right before the election (days or even hours, rather than weeks) before an election.” West Decl. ¶19.

“Deepfakes are especially challenging to address,” because they can be generated and spread more quickly than they are debunked, perform well on

social media, and “[i]t is difficult to counter what people see and hear with their own eyes and ears, especially when an image or video confirms our biases or narratives of how the world works.” West Decl. ¶¶20-22.

West concludes by asserting that the market of ideas is broken:

Protecting the integrity of elections is one of the most important things we can do to preserve democracy. Minnesota's deepfake law helps achieve this. It is one of the few deterrence mechanisms for reducing the spread of a new and powerful technology that mimics real people and situations through images, videos and audio that are extremely hard and time consuming to parse from reality. Left unchecked, the ease of creation, low cost, scalability potential, and the difficulty in debunking this content could be highly disruptive to an election, especially since the technology is relatively new and much of the population will not be aware of these capabilities or have access to the tools to check the veracity of the content. Some may argue that the ‘market of ideas’—the idea that open discourse and the competition of ideas will lead to truth—can correct and properly rebut depictions of people and events. In the days prior to the internet and social media, in particular, this may have been true. In our current digital world, unfortunately the market of ideas doesn't work as John Milton and John Stuart Mills may have envisioned nearly 200 years ago when our information environments were much different. Today, information overload, bots and inauthentic account, echo chambers, algorithmic amplification and power imbalances in our digital world create a marketplace where not all voices have equal access. One of the few mechanisms for countering these challenges is the Minnesota deepfake law that can at least add some friction to the production of blatantly false, but highly convincing and potentially market-damaging content.

West Decl. ¶23.

West's conclusions are somewhat more nuanced than Hancock's. That said, here again, no method appears explaining his legal conclusion that the "market of ideas...can [not] correct and properly rebut depictions of people and events."

Notably, neither expert provides an opinion about whether Kohls' July 26 video would be deceptive to viewers, even though both experts reviewed pleadings including the Complaint. Hancock Decl. ¶6; West Decl. ¶5. Thus, the experts do not "connect[] the proposed testimony with the facts of the case." *Lauzon*, 270 F.3d at 687.

C. The conclusory material in both declarations, including their final paragraphs, does not constitute reliable expert evidence because it was not produced by a reliable methodology.

The Court must assess "whether the reasoning or methodology underlying an expert's testimony is scientifically valid and of whether that reasoning or methodology properly can be applied to the facts in issue." *Lancaster*, 75 F.4th at 969 (8th Cir. 2023). "Under *Daubert*, any step that renders the analysis unreliable renders the expert's testimony inadmissible. This is true whether the step completely changes a reliable methodology or merely misapplies that methodology." *In re Wholesale Grocery Prods. Antitrust Litig.*, 946 F.3d 995, 1001 (8th Cir. 2019) (cleaned up). In this case, both steps are irretrievably broken.

No methodology exists. Studies (at least one fictional) about deepfakes are summarized, then without any method at all, each expert delivers policy recommendations in the form of ready-made legal conclusions. The experts did not apply any theory, technique, or method whatsoever in stating that counterspeech would be “insufficient,” that the “market of ideas” no longer works, or that the statute is “necessary.” “[A] bald assertion cannot carry the *Daubert* burden.” *United States v. Pon*, 963 F.3d 1207, 1221 (11th Cir. 2020).

“[W]here an expert’s opinions are the product of *no* principles or methods, and instead are based on unsubstantiated conclusions, speculation, and mere conjecture, they cannot meet the requirements of Fed. R. Evid. 702.” *Hoekman v. Educ. Minn.*, 335 F.R.D. 219, 240 (D. Minn. 2020) (citing *Wholesale Grocery*); *Ahlberg*, 481 F.3d at 635. “[N]othing in either *Daubert* or the Federal Rules of Evidence requires a district court to admit opinion evidence that is connected to existing data only by the *ipse dixit* of the expert.” *Kumho Tire*, 526 U.S. at 157.

Nothing in the conclusions has any of the intellectual rigor, replicability, or reliability that characterizes traditional expert testimony. There appears to be no way to test or verify the conclusions. They are little more than policy preferences.

D. The conclusory material should be excluded as legal opinion.

It is well-settled that “expert testimony on legal matters is not admissible.” *S. Pine Helicopters, Inc. v. Phx. Aviation Managers, Inc.*, 320 F.3d 838, 841 (8th Cir. 2003). While the experts do not explicitly claim to interpret the First Amendment, their conclusory testimony is only intelligible as legal testimony.

Hancock says that counterspeech would be “insufficient” to address the alleged problem of political deepfakes and that regulatory measures are “necessary.” Hancock Decl. ¶32. The declaration does not define what is meant by this—insufficient *to do what*? By what standard? To what effect? In fact, these phrases are used to substitute for the legal conclusions the Court must reach about strict scrutiny (“actually necessary”) and under the most relevant controlling authority. *281 Care Comm. v. Arneson*, 766 F.3d 774, 793 (8th Cir. 2014) (“There is no reason to presume that counterspeech would not suffice to achieve the interests advanced and is a less restrictive means, certainly, to achieve the same end goal.”). Ellison (at 34) cites them for exactly this purpose.

Like the testimony excluded in *Hoekman v. Educ. Minn.*, the declarations here are attempting to deliver legal conclusions under the guise of scientific expertise. 335 F.R.D. 219, 238 (D. Minn. 2020). “[T]estimony that is ‘carefully couched in the precise language used in case law’ is ‘particularly suspect’ because it indicates that

the expert had either a ‘keen awareness . . . of the direction in which he had to head, or else careful coaching prior to his testimony.’” *Id.* (quoting *Berry v. City of Detroit*, 25 F.3d 1342, 1353 (6th Cir. 1994)); accord *Stock v. Gray*, 2024 U.S. Dist. LEXIS 94868, 2024 WL 2402116 (W.D. Mo. May 23, 2024) (excluding government’s expert testimony in First Amendment case that sought to offer legal opinion by “identifying what societal interests are purportedly implicated by [the statute at issue], or by offering a legal balancing test”).

The declarations purport to override controlling law by delivering legal conclusions that counterspeech, a well-established alternative remedy for false speech, no longer long works because it is “insufficient” (Hancock Decl. ¶32) or because the “market of ideas” went *kaput* sometime between John Stuart Mills and the present day. West Decl. ¶23.

Binding caselaw rejects this. No less in the internet era, courts side with more speech, not less. *SBA List*, 814 F.3d at 476. The supposed novelty of deepfakes does not alter the law any more than prior technological changes did. *Kohls v. Bonta*, 2024 U.S. Dist. LEXIS 179933, *14 (E.D. Cal. Oct. 2, 2024).

In fact, as discussed further below, both experts have written that counterspeech remains viable. The public learns to react to changed technologies. Hancock describes this in his cited editorial. “It is important to recall that humans

have been adapting to novel forms of deception for millenia” and they respond when “warned by a third party, or we are educated about novel deceptive techniques.” Hancock & Bailenson, 2021, Bednarz Ex. 5, at 150. “For example, email spam is much less effective than when it first emerged, in part because people are aware of it.” *Id.*

“Before exempting a category of speech from the normal prohibition on content-based restrictions... the Court must be presented with ‘persuasive evidence that a novel restriction on content is part of a long (if heretofore unrecognized) tradition of proscription.’” *Alvarez*, 567 U.S. at 722. No tradition exists to constrain parody, which has confused some listeners since Classical times. AI-aided deepfakes provide a new mode of engaging in this political impersonation, but “the First Amendment’s command [does] not vary.” *Brown v. Entm’t Merchs. Ass’n*, 564 U.S. 786, 790 (2011).

The conclusory paragraphs violate Federal Rule of Evidence 702 by offering legal conclusions contrary to First Amendment case law favoring counterspeech. “Such conclusions are impermissible legal opinions, exclusively reserved for the Court, ... and accordingly will be excluded.” *Hoekman*, 335 F.R.D. at 239. The testimony is at least unhelpful because “[e]ach courtroom comes equipped with a

'legal expert' called a judge." *Burkhart v. Washington Metro. Area Transit Auth.*, 112 F.3d 1207, 1213 (D.C. Cir. 1997).

III. The experts' other publications undermine their conclusory paragraphs.

A. In his cited editorial, Hancock is more equivocal about the potential harm of deepfakes.

Hancock's declaration cites "Hancock & Bailenson, 2021" four times, which is a guest editorial for a special edition of the journal *CYBERPSYCHOLOGY, BEHAVIOR, AND SOCIAL NETWORKING* about "The Social Impact of DeepFakes." Bednarz Ex. 5. It provides an overview of recent research, including articles published in that issue, which Hancock and his co-author were guest editors of. The piece supports several of Plaintiffs' arguments.

In contrast to the unsupported conclusions in his declaration, the editorial is much more equivocal about the potential harm of deepfakes.

Hancock notes that deepfakes are not uniquely deceptive. "Studies have shown that deception detection is approximately the same whether the message is conveyed through text (e.g., a court transcript, an Internet chat log), an audio recording (e.g., a voicemail, a radio program), or a video (e.g., an interrogation video). Although this may seem surprising given the richer detail available in video, accuracy tends to be at chance regardless of medium because there are no

reliable signals to human deception (i.e., there is no Pinocchio's nose) and we tend to trust what others say." Bednarz Ex. 5 at 149-50 (citing study).

He speculates that deepfakes might have greater impact, however. "Although the rates of detection are likely similar to other media, the impact of deception by deepfake has the potential to be greater than that of verbal deception because of the primacy of visual communication for human cognition." *Id.* at 150.

After summarizing studies showing that deepfakes could reduce trust in news (even if not believed) and inflame prejudice, the editorial adds nuance not evident in Hancock's declaration:

Although these implications paint a discouraging portrait of a future with deepfake technology, this take assumes a relatively passive consumer of media. It is important to recall that humans have been adapting to novel forms of deception for millenia. People tend to be trusting of one another until they have some reason to become suspicious or more vigilant, a state that Levine refers to as a trust default. We move out of our trust default when we learn about inconsistent information, or we are warned by a third party, or we are educated about novel deceptive techniques. For example, email spam is much less effective than when it first emerged, in part because people are aware of it.

In the same way, it is possible for people to develop resilience to novel forms of deception such as deepfakes. For example, advertising frequently relies on misleading visual information (e.g., drink this beer, have beautiful friends; smoke this cigarette, experience the great outdoors). Over time, consumers get their guard up and are not fooled by advertising, in part because they develop a schema of expectations for advertising. Indeed, we develop expectations like

this for most media we consume.

Bednarz Ex. 5 at 150 (citations omitted).

One of the studies in the special edition supports this belief that citizens can adapt to deepfakes. Hancock and his co-author summarized it like this: “In what is likely the first study evaluating a media literacy program targeting deepfakes..., the authors show efficacy for a media literacy program in Korea that builds resilience to believing deepfakes.” Bednarz Ex. 5 at 151.

The editorial concludes equivocally. “Although deepfake technology has the potential to undermine our trust in media or falsely influence our beliefs about the world, it may also become more commonplace and mundane as people use deepfake technology to improve their day-to-day communication.” *Id.*

B. In his book, West supports counterspeech to combat misinformation and notes the vague line between parody and misinformation.

While West’s final conclusory paragraph suggests that the market of ideas is broken, he has previously been much less despondent about refuting misinformation. He co-wrote a book to help a “critically thinking electorate” spot and refute misinformation. Bergstrom and West, *CALLING BULLSHIT* (2020). The preface says: “we believe that adequate bullshit detection is essential for the survival of liberal democracy,” and favorably quotes an op-ed that opines “instead

of trying combat each [foreign propaganda] leak directly, the United States government should teach the public to tell when they are being manipulated.”

Bednarz Ex. 6 at xiv.

In chapter 2, West and his co-author noted that the problems with misinformation, attributed to deepfakes in his declaration, existed before AI because of social media. Bednarz Ex. 7 at 25-36. But then he remarks: “We are not so pessimistic.... Society will adjust similarly to a world of deepfakes and whatever reality-bending technologies follow.” *Id.* at 36.

CALLING BULLSHIT outlines three potential solutions for misinformation: technology, regulation, and education. *Id.* Technology unlikely helps: “Technologically, the same artificial intelligence techniques used to detect fake news can be used to get around detectors, leading to an arms race of production and detection that the detectors are unlikely to win.” *Id.* As for regulation, West and his coauthor also rejected that as a solution:

First, it runs afoul the First Amendment to the U.S. Constitution, which guarantees freedom of speech. Second, who gets to determine what is fake news? If a leader doesn’t like a story he or she could declare it fake news and pursue criminal charges against the perpetrators.

Bednarz Ex. 7 at 36 (emphasis added).

Instead, the “most powerful” approach is education. *More speech*, just as First Amendment caselaw and Plaintiffs suggest. “If we do a good job educating people in media literacy and critical thinking, the problem of misinformation and disinformation can be solved from the bottom up. That is our focus in this book, and in much of our professional lives.” *Id.* at 37.

Similarly, the concluding paragraph of CALLING BULLSHIT reads:

The rise of misinformation and disinformation keeps us up at night. *No law* or fancy new AI *is going to solve the problem*. We all have to be a little more vigilant, a little more thoughtful, a little more careful when sharing information-and every once in a while, we need to call bullshit when we see it.

Bednarz Ex. 9 (emphasis added).

In CALLING BULLSHIT, West also wrote in a roundabout way about the vague line between deliberate misinformation and parody. In chapter 8, he writes about the difficulty of training AI to spot misinformation.

Compare the challenge of teaching an algorithm to classify news stories as true or fake. This is much harder than deciding whether a handwritten numeral is a six. You don’t necessarily know the answer just by looking at it; you may have to do some research. It’s not clear where to do that research, or what sources count as authoritative. **Once you find an answer, reasonable people still might not agree whether you are looking at fake news, hyperpartisan news, satire, or some other type of misinformation.** And because fake news is continually evolving, a training set of fake news stories from 2020 might be out of date by 2021.

Bednarz Ex. 8 at 192 (emphasis added).

So, even when something is proved false, “reasonable people still might not agree” whether it’s misinformation or satire. The vagueness of parody among reasonable people is why Ellison’s proposed insertion of an atextual parody exception into Section 609.771 does not resolve the overbreadth and vagueness problems flagged by Plaintiffs. It merely moves the ambiguous and over-inclusive boundary from “realistic” to an equally vague and over-inclusive line for “parody,” which Ellison does not further define.

IV. If not excluded, the Court should only consider the declarations, not exaggerations contradicted by them.

Ellison’s memorandum exaggerates several claims not found within—and actually contradicted by—the declarations.

Ellison (at 8) says that “Deepfakes are also unique because of the impact of false memory and repeated exposure.” (Citing Hancock Decl. ¶¶28-30.) But none of these paragraphs say that deepfakes are unique in this regard. In fact, the cited studies show they are *not* unique. Paragraph 28 describes the Loftus *et al.* study. Bednarz Ex. 10. The introduction to the Loftus study explained that the false memory effect studied has been previously documented using different real photographs or manually-edited photographs. *Id.* at 2. “In a notable study, fifty percent of participants developed complete or partial false memories after

exposure to a fake childhood photograph and guided imagery exercises over three interviews.” *Id.* Loftus *et al.* replicate these results with AI-edited content, but they do not compare AI to non-AI images for their ability to implant false memories, let alone conclude that AI images are “uniquely” warping to memory. In his declaration, Hancock somewhat misleadingly describes the finding as: “When participants’ memories were assessed, the percentage of false memories were two times higher in the deepfake video condition than the control, suggesting that deepfakes can significantly increase false memories.” Hancock Decl. ¶28. But the “control” in the study was being shown the same unedited image a second time. It should not be surprising that subjects shown an altered photo or video are less able to accurately recall the unedited photo shown at the beginning of the study than subjects shown the same photo twice. These results simply replicate those caused by manually-edited images in 1998, as Loftus cites.⁸ Hancock’s next paragraph discusses the Illusory Truth Effect (ITE)—the repetition of misinformation makes it seem more credible. Hancock Decl. ¶29. ITE is a generic problem with misinformation, having been studied since the 1970s, and Hancock

⁸ Michael B Miller and Michael S Gazzaniga. 1998. *Creating false memories for visual scenes*. NEUROPSYCHOLOGIA 36, 6 (1998), 513–520, available at: <https://www.sciencedirect.com/science/article/pii/S0028393297001486>.

does not suggest that deepfakes are “unique” in this regard. *Id.* Hancock reports that “[t]he study found that individuals who had previously been exposed to deepfakes were more likely to perceive them as accurate compared to those encountering them for the first time, for both political and non-political deepfakes.” *Id.* But this is what the ITE hypothesis predicts, and is not unique to deepfakes. Indeed, the cited Ahmed *et al.* study remarks “[t]he results are consistent with the literature on earlier forms of misinformation, such as text and visual fake news.” Bednarz Ex. 11 at 17. Arguably, the Ahmed study supports Plaintiffs’ position that a ban on political deepfakes is unnecessary or even counter-productive: it found that subjects from China were most likely to believe deepfakes, while subjects from Singapore (a tech-savvy democracy) were least likely. The authors theorize the cause for this disparity:

In China, the results suggest that respondents were most likely to perceive the misinformation as accurate compared to all other contexts. Several factors could explain this. The government tightly controls the media environment in China with strict regulations on what can be published, which also extends to the social media environment.... In such an environment, the information that filters to social media may also be perceived as more credible as individuals may not doubt its authenticity given trust in government scrutiny. This would be in contrast with countries with lesser regulated media environments where misinformation can spread more freely.

Bednarz Ex. 11 at 20.

Ellison (at 32) concludes a discussion of the expert declarations by remarking “Thus, in almost all cases, it will be impossible to engage in counterspeech that will adequately, or even meaningfully, respond to a deepfake.” Ellison does not cite the declarations for this proposition, and for good reason: it does not exist in them. In fact, the West declaration supports Plaintiffs’ overbreadth argument because he says political deepfakes are “often introduced... right before the election (days or even hours, rather than weeks).” West Decl. ¶19. Plaintiffs contend Section 609.771 is not narrowly tailored precisely because it covers 90 days before a primary election—not just the immediate period before an election when it might be impossible to rebut a deepfake.

After referencing the declarations, Ellison says (at 35) that “[t]he ban established by Minnesota is actually necessary to combat that influence.” But neither declaration opines this. Hancock says “enacting and enforcing laws that specifically target the production and dissemination of deceptive deepfake content during elections are critical in preserving the integrity of democratic institutions and protecting the foundational trust upon which they rely.” Hancock Decl. ¶32. This carefully-written sentence does not actually endorse the Minnesota law, perhaps because fails to address the “liar’s dividend” of falsely claiming an authentic video is a deepfake. *Id.* ¶27. West addresses the statute, but he does not

call it *necessary*. Only that “[o]ne of the few mechanisms for countering these challenges is the Minnesota deepfake law that can at least add some friction to the production of blatantly false, but highly convincing and potentially market-damaging content.” West Decl. ¶23 (emphasis added). West thus implies other less restrictive mechanisms exist, which is not surprising because he co-wrote a book about recognizing and rebutting misinformation with education—more speech, as the Constitution demands.

Ellison (at 32) says “The record evidence instead demonstrates that even when viewers know of deepfakes, they still struggle to distinguish between real and manipulated content.” (Citing Hancock Decl. ¶¶21, 28-29.) As discussed above, Hancock Decl. ¶28-29 do not show or even suggest that deepfakes are different from earlier deceptive media in this regard. The research is based on known memory effects studied for decades, but relevant to AI due to social media image filters potentially altering people’s recollection of their own lives. *See* Bednarz Ex. 10 at 1. As for paragraph 21, this concerns the fictitious Hwang et al. (2023) study, which appears to be a fabrication invented by an AI chatbot, as discussed in Section I.

CONCLUSION

The declaration of Prof. Hancock should be excluded in its entirety because at least some of it is based on fabricated material likely generated by an AI model, which calls into question its conclusory assertions. The court may inquire into the source of the fabrication and additional action may be warranted.

Independently, the conclusory paragraphs of both declarations should be excluded because they are not based on any apparent methodology at all, but only the experts' say-so and legal conclusions. Even if not excluded, the contradictions of the report with other publications of the experts goes to weight, and only material in the declarations should be credited, not counsel's characterizations which in several places are contradicted by the cited literature.

Dated: November 15, 2024

Respectfully submitted,

/s/ M. Frank Bednarz

Douglas P. Seaton (#127759)

James V. F. Dickey (#393613)

Alexandra K. Howell (#0504850)

UPPER MIDWEST LAW CENTER

12600 Whitewater Drive, Suite 140

Minnetonka, MN 55343

Voice: 612-428-7000

Email: Doug.Seaton@umlc.org

James.Dickey@umlc.org

Allie.Howell@umlc.org

M. Frank Bednarz (*pro hac vice*)

HAMILTON LINCOLN LAW INSTITUTE

1440 W. Taylor St. # 1487

Chicago, IL 60607

Voice: 801-706-2690

Email: frank.bednarz@hlli.org

*Attorneys for Plaintiffs Christopher Kohls
and Mary Franson*