Comments of David Inserra, Jennifer Huddleston, and Juan Londoño in Response to Federal Trade Commission Request for Public Comment on Its Policy Statement Addressing AI Accuracy

We appreciate the opportunity to provide information related to the Federal Trade Commission (FTC)'s Request for "Public Comment on Policy Statement Addressing AI Accuracy." This comment does not represent the views of any particular party or special interest group but is intended to assist regulators in understanding how involving federal regulators in deciding what makes an artificial intelligence (AI) model "accurate" impacts innovation and expression. Our comments focus mainly on these key takeaways:

- The statement correctly sheds light on state laws' free speech issues;
- The commission should be restrained regarding its own authority under Section 5 and consider the broader risks of an expansive view;
- The commission must consider the impact of its own actions on innovation and expression as well as those it is considering from state laws; and
- Education, not regulation, can provide an appropriate counterbalance to many concerns regarding AI "accuracy" issues.

The Statement Correctly Sheds Light on State Laws' Free Speech Issues

The FTC's policy statement expresses concern about the growing number of state-level AI bills that would force AI companies to modify their models and either carry speech they do not wish to carry or censor speech they would otherwise carry. The statement places particular emphasis on Colorado's infamous AI bill, SB 26-189, which has faced significant hurdles in its implementation, including delays and lawsuits, largely due to its onerous requirements and "disparate impact" standard.

The main issue with Colorado's bill and other proposals that follow the same approach is their focus on what the bill considers "AI bias," which forces companies to alter their products to curtail what policymakers believe could be AI-powered discrimination. While this may be a laudable goal, these bills ultimately tell AI companies that certain outputs are prohibited, even if they could be considered protected speech.

These laws present a great opportunity to break down how AI models and their outputs are by-products of multiple editorial and creative decisions, which are ultimately expressive actions. While it is more obvious to see how end users are engaging in expressive content when generating content using generative AI, the expressive behaviors from developers cannot and should not be ignored. Developers make editorial and expressive decisions

about how to train their models, what information or sources the models will consider in their outputs, how to present these outputs, and when to withhold an output.

There are multiple examples of how these editorializing decisions manifest. In its lawsuit against the Colorado bill, AI company xAI argues that it has trained its model to seek what it defines as "objective truth" and to present these facts plainly, without consideration of political or ideological ramifications.[i] This standard has shaped how they train their model, how they fine-tune its weights, and which guardrails or controls they decide to include or omit. Similarly, OpenAI decided to rework its ChatGPT-4o model because it considered it "overly flattering or agreeable," to the point of being sycophantic.[ii] Even when presenting "objective" information, the tone the model decides to present this information with can alter how the end user perceives it. OpenAI believed that an overly sycophantic model was considered by its customers as less trustworthy, less accurate, and unsettling. Thus, they decided to alter how the model presents information to better suit what they believe serves their consumers best. This is a subjective editorial decision.

Bills like SB 26-189 would force AI companies to either train their models in a particular way, add caveats or additional commentary to their outputs, or refrain from producing any outputs that could be considered contributing to AI-powered discrimination. Putting such guardrails on protected expressive behavior would inherently violate developers' free speech rights and consumers' right to access information freely.

The FTC's AI policy statement correctly recognizes that AI will be "relied upon for research, analysis, and advice on both personal and business issues," thus, it should not be forced by these bills to "silence or censor lawful expression or dissent." Throughout the document, the agency correctly diagnoses that these state bills will shape AI technologies in ways that violate the First Amendment rights of consumers and developers, and feels compelled to step in. However, the authority invoked and the scope of the agency's intervention raise significant First Amendment questions as well.

The Commission Should Be Restrained Regarding Its Own Authority

While states can create a patchwork of problematic AI regulations, the need to preempt them does not absolve the federal government of its own responsibilities or of potential problems in its policy approaches. Federal policies may raise their own speech concerns, and policymakers must exercise appropriate restraint in their use of authority at all levels of governance, being careful not to create a more problematic framework in attempts to overcome a state patchwork.

Agencies must also not overstep the scope of their authority or claim new regulatory powers that are not established by congressional delegation to the agency. In this regard,

one of the first considerations is whether Section 5 authority is an appropriate tool for such preemption. Even in the executive order that gave rise to this, there are clearly other tools at the administration's disposal to challenge interstate burdens, such as the dormant commerce clause. Such a broad interpretation of Section 5 risks granting the FTC far more expansive authority over a wide range of speech-related actions. This would raise significant free expression concerns and constitute a significant expansion of the agency's power more generally.

The potential expansion into the speech of platforms around their decision on what constitutes truth goes beyond just AI. If the FTC is to view its own authority as expanding into being the arbiter of truth, rather than just its authority over deceptive claims as traditionally understood, this could result in a much broader intrusion into free expression and the information ecosystem.

Assuming the role of an arbiter of truth, the FTC simultaneously and selectively exaggerates and minimizes what AI companies have promised consumers to argue that their conduct is deceptive under its new interpretation of Section 5. The FTC argues that AI companies "that steer the outputs of their AI systems toward unexpected objectives, and away from the objectives set by or reasonably expected by users" likely have deceived their users.[iii] As we mentioned in commentary at the time, the FTC made a very similar argument in February 2026 when it warned Apple News that its selection of news could also be considered a deception of users.[iv]

As with Apple News, this can result in the FTC making subjective determinations regarding what counts as "the best possible output," "ideologically motivated distortions," "politically inflammatory outputs," or "their own or their employees' political agendas."[v] In doing so, the FTC claims the authority that can punish any AI company that holds different values and views of what constitutes "the truth," or "objectivity," or any company that merely makes a mistake, because a company's lack of conformity with the FTC's views can be considered a deception of consumers.

Such subjective determinations are nothing more than an elaborate attempt to argue that the FTC should have the power to determine which views are reputable, truthful, and good, and which views are deceptive, misleading, or bad because of their ideological or political nature. This allows the FTC to use political or ideological viewpoints as the basis for considering something to be deceptive. The FTC tries to frame this in terms of consumer expectations, but the FTC is unilaterally designating itself as the regulator of AI-related speech.

In addition to its potential censorious nature—which will be expanded upon later—such a proposal significantly expands the scope for government intrusion into product development and deployment, which could chill access to information. The FTC should also consider how such expansive power could be abused in the future. Even if it considers its own actions a means of asserting balance, the inclusion of a broader interpretation would allow future commissions to engage in myriad regulations that go beyond the typical scope of their authority and could burden entrepreneurship and innovation for both companies and consumers.

This emerging view of Section 5 goes well beyond the established interpretation of the agency's authority to regulate conduct that exceeds what is traditionally considered deception. The FTC's own history shows that over-expansive interpretations may initially lead to greater impact for the agency, but they are also likely to draw public and congressional ire, resulting in further restraint even on existing authority.[vi]

Federal Determinations Could Create Their Own Speech Problems

Trading many regulators for one does not solve the problem if the resulting regulatory environment still burdens innovation and expression. The FTC's proposed action could exert more government pressure on AI-related expressive decisions than the state laws it aims to preempt, even if it is in the name of establishing "truth." Many matters of "truth" have multiple viewpoints, and just as with state mandates, a federal government attempt to establish when an AI is "truthful" could allow government regulators to dictate what information can be used in either programming or querying an AI product.

The most immediate and basic challenge to the FTC's proposed policy statement is the disparity between the legal precedent the FTC uses to justify its actions and the supposed deceptions committed by AI companies. In the cases cited, the deceptions were specific and measurable: (1) a company claimed its product would help prevent disease but had no scientific study to support it; (2) a company charged consumers money but never obtained their express consent; and (3) a company promised no hidden fees but charged fees that many consumers never knew about.

These are nothing like the promises made by AI companies. For example, the FTC cites the following vague marketing statements:

- "At OpenAI, we're working hard to make AI systems more useful and reliable";[vii]
- "Helpfulness is one of Claude's most important traits…providing information grounded in evidence";[viii]
- "Grok is your truth-seeking AI companion";[ix] and

- "Claude is an artificial intelligence, trained by Anthropic using Constitutional AI to be safe, accurate, and secure."[x]

These aspirational promises of efficacy, helpfulness, safety, and truth-seeking cannot be confirmed through randomized controlled trials. These statements are in no way hiding the ball on tangible hidden fees or costs, because consumers were not enrolled in AI tools without their consent. While each of these companies certainly promotes the value and power of their tools in these marketing statements, they also regularly note their governing principles and policies, as well as their limitations. Further, the companies warn users to beware of hallucinations and other faulty AI outputs[xi] and provide explicit disclaimers in their terms of service.[xii] If the FTC were alleging that these companies were promising scientifically provable benefits without evidence, signing users up for accounts without their consent, or failing to notify users of usage fees after claiming there were no fees, then the FTC may have a case. However, it alleges no such thing.

Instead, the FTC claims that precedent from cases where the violations were discrete, measurable, and objectively verifiable acts and claims is similarly applicable to AI companies that make "explicit and implicit representations that AI systems aim to produce the best output possible given technological and resource constraints." In this sense, the FTC argues that AI companies with differing views or approaches from what the agency believes is "the best" output can be deemed deceptive. In other words, the FTC is trying to judge AI models' accuracy and performance—two largely subjective variables—in the same way it evaluates dietary supplements' medical-benefit claims or users being charged fees without proper notice or consent. This is an absurd comparison.

Even if such statements were somehow considered to be within the purview of Section 5, the FTC is in no position to objectively determine whether the outputs of AI tools are "the best" or are deceptive. As the FTC itself acknowledges, "Of course, AI systems are likely to simultaneously pursue multiple objectives." But, the FTC argues, AI companies "might steer their systems' outputs toward objectives other than those that consumers ask for or expect." For example, AI companies could be "training a model surreptitiously to produce ideologically motivated distortions," or "suppress accuracy and interpose other objectives, such as so-called 'equity.'" The FTC argues that any such distortion, suppression, or modification to advance certain "political agendas" is a deception.[xiii] Thus, the agency will inevitably have to rely on subjective criteria and discretion to enforce this policy.

As a result of this subjective and ideological basis for its proposed policy, the FTC will be unable to consistently and clearly define when an AI tool has acted deceptively. As Americans, we disagree about many issues, including race, gender, immigration, elections, gerrymandering, Supreme Court packing, economic theory, gun rights, the role of religion,

and countless others. AI tools will be asked questions regarding all of them. While some users may wish AI tools would align with their views on those issues, other users want AI tools to present a completely different view. What some users would consider a balanced or fair view, others might consider unbalanced.

In short, it is guaranteed that multiple significant minorities of users will all prefer and expect AI tools to align with their viewpoints. If a tool fails to meet the expectations and political views of each of these significant minority groups, the FTC's policy holds that the company has engaged in deception. However, this is an unreachable standard requiring companies to satisfy every group in society with their answers.

The more likely and more pernicious alternative is that the FTC expects AI companies to adhere to the current government's preferences and expectations. The FTC's own proposal cites left-wing views of historical injustices, disparate impact, discrimination, and equity, strongly suggesting that AI tools that provide or favor such viewpoints have acted deceptively. Such a standard is deeply injurious to free expression and raises serious constitutional problems.

As a practical matter, how would the FTC even determine that an AI company was motivated by a political ideology to distort its AI outputs? Is the FTC going to review each output of each AI tool across every contentious issue by probing and asking follow-up questions to ascertain the total amount of ideological distortion that each model has? Is it going to weigh how much it supports one view versus other views? If an AI tool consistently favors one viewpoint but provides strong evidence to back up that viewpoint, is that distortion or accuracy? How will the FTC know that any given result isn't a model mistake or hallucination, rather than a political agenda? Will companies have a safe harbor for alleging that an output is a product of a mistake or hallucination? How will the FTC determine where the line between "mistake" and "willful manipulation" resides? And what is the clear standard for when a model is too biased and too steered towards ideological objectives so that it has become deceptive? Even granting the implausible assumption that the FTC would apply its standard without viewpoint discrimination, the FTC offers no principled line between an acceptable degree of bias and an unlawful deception.

More broadly, how does the FTC differentiate this theory from other forms of media? The New York Times has the motto "all the news that's fit to print," but many readers would disagree that it upholds that motto.[xiv] They would contend that the paper has printed stories that were not fit to print, while leaving news that *was* fit to print unaddressed or poorly addressed. As the US' paper of record, the New York Times has made explicit and implicit claims that it aims to produce the best possible output given technological and resource constraints. Should the FTC be allowed to act against the Times? Fox News once

used the motto "Fair & Balanced," but many Americans disagreed.[xv] Should the FTC have been able to act against Fox for such deceptive representations? These examples make clear what the FTC's proposed policy does not—AI tools, like other media, are protected speech and ought to be free from government pressure and censorship.

It is also worth briefly considering the consequences of the proposed FTC policy. It would slow updates to AI systems, because making a change—even if the change is meant to improve accuracy, reduce biases, or attempt to be more even-handed—would result in some consumers believing that AI companies had "steer[ed] the outputs of their AI systems away from the objectives set by or reasonably expected by users." This could result in various unintended and perverse consequences. For example, the Trump Administration blocked the use of Anthropic's most powerful Fable 5 model due to security and safety concerns. In response, Anthropic re-released Fable 5 with greater safety classifiers to prevent users from accessing certain categories of potentially dangerous information. Users asking Fable 5 about certain topics would be routed to "less capable models" that would not provide as robust, complete, or accurate information.[xvi] But under the FTC's proposed policy statement, this would appear to be a deceptive act by steering users away from their objectives or expectations. In general, this reduces AI companies' ability to respond to current events and improve their products.

The policy would also pressure AI companies to generate content they do not wish to, or to refrain from generating content they wish to. An AI company may refuse to generate imagery associated with violent and hateful groups such as ISIS, the KKK, or Nazis. Nonetheless, other users may believe that AI companies have deceived them by making vague promises to be helpful and accurate but then refusing to generate that category of content. Similarly, a company may wish to generate outputs that emphasize the biological elements of sex and gender but then finds that a future FTC deems gender identity-affirming views more important or accurate. Faced with the threat of FTC action, AI companies will likely be pressured to align their policies with the views of the administration of the day.

The FTC's proposed statement promises subjective, viewpoint-based punishment that harms innovation. Such a policy is censorious, viewpoint-discriminatory, and deeply injurious to continued innovation.

Education on AI Is an Alternative to Help Individuals Ascertain Accuracy

A better response to concerns about the truthfulness of AI does not come from government mandates but from equipping citizens with the tools to analyze the media and information they encounter. Education, not regulation, can provide an appropriate counterbalance that

empowers consumers and innovators while ultimately making the public, not a single regulator, the arbiter of truth. The FTC has itself engaged in such educational campaigns on many fraud and deception issues.[xvii] Rather than seeking to establish a single view of which models are adequately accurate, the agency could provide information on AI literacy to help consumers evaluate the information they receive from AI, without favoring any particular outcome.

Conclusion

The FTC policy statement correctly identifies the risks posed by state AI policies to innovation and speech. But in seeking to overcome the disruption caused by the risks of a state patchwork, greater care is needed to ensure that FTC action does not replace one set of potentially censorious regulations with another. The agency should also carefully consider the full consequences of an expansive view of its authority, not only on innovation and speech in the AI sector, but also on how such an authority could be abused to create more harmful restrictions or interventions in the future.

---

[i] David Inserra, "xAI Sues Over Yet Another Colorado Law That Threatens Free Expression," *Cato at Liberty*, April 22, 2026.
[ii] "Sycophancy in GPT-4o: what happened and what we're doing about it," OpenAI, April 29, 2025.
[iii] "Federal Trade Commission's Proposed Policy Statement Concerning the Suppression of Accuracy in Artificial Intelligence Systems," Federal Trade Commission, July 1, 2026.
[iv] David Inserra, "FTC Continues to Confuse Free Expression and Censorship as It Threatens Apple News," *Cato at Liberty*, February 13, 2026.
[v] "Federal Trade Commission's Proposed Policy Statement Concerning the Suppression of Accuracy in Artificial Intelligence Systems," Federal Trade Commission, July 1, 2026.
[vi] See Merrill Brown, "FTC Temporarily Closed in Budget Dispute," *Washington Post*, April 30, 1980.
[vii] Adam Kalai, et al., "Why language models hallucinate," OpenAI, September 5, 2025.
[viii] Judy Hanwen Shen, et al., "How people ask Claude for personal guidance," Anthropic, April 30, 2026.
[ix] "The truth-seeking AI assistant," xAI.
[x] "Question what's next," Claude; Amanda Askell, et al., "Claude's Constitution," Anthropic, January 21, 2026.
[xi] See OpenAI: "Does ChatGPT Tell the Truth?," OpenAI, last modified July 29, 2026; "Why Language Models Hallucinate," OpenAI, September 5, 2025; in-product disclaimer: "ChatGPT can make mistakes. Consider checking important information." See Anthropic: "Reduce Hallucinations," Anthropic, accessed July 31, 2026; "Measuring Political Bias in Claude," Anthropic, November 13, 2025 ("There is no agreed-upon definition of political bias, and no consensus on how to measure it. Ideal behavior for AI models isn't always clear"); in-product disclaimer: "Claude is AI and can make mistakes. Please double-check cited sources." See xAI: "Announcing Grok," xAI, November 3, 2023; "xAI Consumer FAQs," xAI, May 12, 2025; in-product disclaimer: "Grok can make mistakes. Verify its outputs."
[xii] The FTC itself disclaims such disclaimers and terms of service as insufficient remedies and outliers. A review of the terms of service xAI, Anthropic, and OpenAI, backs up these statements with each including clear and sometimes all capitalized language that these services are provided "as is" with no promises of perfect accuracy or performance. Users are warned not to rely on these services as sole sources of truth or professional advice. Users that are dissatisfied are told they may terminate the service at any time. These statements are not outliers but generally align with the other statements made by AI companies regarding

their limitations and underlying principles. The FTC's selective choice of AI company statements ignores countervailing statements and evidence. See: "Terms of Service – Consumer," xAI, June 26, 2026; "Consumer Terms of Service," Anthropic, October 8, 2025; "Terms of Use," OpenAI, January 1, 2026.

[xiii] "Federal Trade Commission's Proposed Policy Statement Concerning the Suppression of Accuracy in Artificial Intelligence Systems," Federal Trade Commission, July 1, 2026.

[xiv] "Slogan for The Times on the Web: 'All the News That's Fit to Print,'" *New York Times*, October 25, 1996.

[xv] Michael M. Grynbaum, "Fox News Drops 'Fair and Balanced' Motto," *New York Times*, June 14, 2017.

[xvi] "Redeploying Fable 5," Anthropic, June 30, 2026.

[xvii] See e.g. "Pass It On," Federal Trade Commission.