Date of Hearing:
Fiscal: Yes

ASSEMBLY COMMITTEE ON PRIVACY AND CONSUMER PROTECTION
Rebecca Bauer-Kahan, Chair
AB 410 (Wilson) – As Introduced March 24, 2025

PROPOSED AMENDMENTS

SUBJECT: Bots: disclosure

SYNOPSIS

SB 1001 (Hertzberg, Ch. 892, Stats. 2018) made it unlawful to use bots posing as humans to induce commercial transactions or influence elections. The bill provided that an owner of a bot that engages in such communications would not face liability if they disclose that the bot is in fact a bot. At the time, the primary concern was that bots could deceptively influence commercial or political activity. However, the rapid advancement of generative artificial intelligence (GenAI) in recent years has led to the proliferation of bots capable of replicating human-like conversation both textually and verbally. This has raised concerns about the potential for manipulative misrepresentation in numerous different contexts.

Sponsored by the National Youth AI Council, this bill mandates that bots disclose their identity before interacting with another person, respond truthfully to any query about its identity, and otherwise refrain from misrepresenting itself as a human. The bill is supported by the California State Association of Psychiatrists and the California Initiative on Technology and Democracy (CITED) and is opposed by Oakland Privacy.

Committee Amendments outlined in Comment #8 refine the bill's definitions of bots and GenAI, recast the disclosure obligations to apply to the person who uses the bot, and add an express enforcement mechanism.

THIS BILL:

1) Defines the following terms:

   a) "Artificial intelligence" to mean an engineered or machine-based system that varies in its level of autonomy and that can, for explicit or implicit objectives, infer from the input it receives how to generate outputs that can influence physical or virtual environments.

   b) "Bot" to mean an automated online account where all or substantially all of the actions or posts of that account are not the result of a human being or are the result of generative artificial intelligence.

   c) Generative artificial intelligence" to mean artificial intelligence that can generate derived synthetic content, including images, videos, audio, text, and other digital content.

2) Requires that a bot do all of the following:

a) Disclose to any person with whom it interacts before any interaction takes place that the bot is a bot and not a human being.

b) Answer truthfully any query from a person regarding its identity as a bot or a human.

c) Refrain from attempting to mislead a person regarding its identity as a bot.

3) Provides that if a bot is required by another law to comply with a more prescriptive disclosure scheme than the bill requires, the bot is not required to comply with the bill.

4) Removes language that provides that existing law does not impose a duty on service providers of online platforms.

EXISTING LAW:

1) Defines the following terms:

a. "Bot" to mean an automated online account on an online platform that is designed to mimic or behave like the account of a person.

b. "Online" to mean appearing on any public-facing Internet Web site, Web application, or digital application, including a social network or publication.

c. "Online platform" to mean any public-facing Internet Web Site, Web application, or digital application, including a social network or publication.

d. "Person" to mean a natural person, corporation, limited liability company, partnership, joint venture, association, estate, trust, government, governmental subdivision or agency, or other legal entity or any combination thereof. (Bus. & Prof. Code § 17940.)

2) Prohibits any person from using a bot to communicate or interact with another person in California online, with the intent to mislead the other person about its artificial identity for the purpose of knowingly deceiving the person about the content of the communication in order to incentivize a purchase or sale of goods or services in a commercial transaction or to influence a vote in an election. (Bus. & Prof. Code § 17941(a).)

3) Provides that a person using a bot will not be held liable if the person discloses that it is a bot. (Bus. & Prof. Code § 17941(a).)

4) Requires that disclosure of bots be clear, conspicuous, and reasonably designed to inform persons with whom the bot communicates or interacts that it is a bot.(Bus. & Prof. Code § 17941(b).)

5) Requires that whenever telephone calls are placed through the use of an automatic dialing-announcing device, the device may be operated only after an unrecorded, natural voice announcement has been made to the person called by the person calling, the announcement must do all of the following:

a. State the nature of the call and the name, address, and telephone number of the business or organization being represented, if any.

b. Inquire as to whether the person called consents to hear the prerecorded message of the person calling.

c. Inform the person called if the prerecorded message uses an artificial voice. (Public Utilities Code § 2874(a).)

6) Generally protects consumers and competitors against unlawful, unfair, or fraudulent business act or practices. (Bus. & Prof. Code § 17200 et seq.)

7) Generally protects consumers and competitors against false or misleading advertising. (Bus. & Prof. Code § 17500 et seq.)

COMMENTS:

1) **Author's statement**. According to the author:

> Californians deserve to know if they are interacting with a real person or a bot. By requiring a simple factual disclosure, AB 410 aims to increase both transparency and trust in the digital world.
>
> As the author of AB 410, I believe Californians have a right to know who—or what—they are interacting with online. In a world where AI-powered chatbots are becoming more advanced and widespread, chatbots are becoming harder to detect, and transparency is essential to maintaining trust in digital spaces.
>
> This bill does not ban bots—it simply requires them to disclose their artificial nature. Whether in customer service, healthcare, or on social media, people have the right to know whether they're engaging with a human or an AI. By simplifying the existing law, AB 410 strengthens consumer protections, combats misinformation, and upholds the integrity of online communication.
>
> Other states and countries are already taking action on AI transparency. By passing AB 410, we empower Californians with greater clarity and confidence in their online interactions.

2) **AI and GenAI.** The development of GenAI is creating exciting opportunities to grow California's economy and improve the lives of its residents. GenAI can generate compelling text, images and audio in an instant – but with novel technologies come novel safety concerns.

In brief, AI is the mimicking of human intelligence by artificial systems such as computers. AI uses algorithms – sets of rules – to transform inputs into outputs. Inputs and outputs can be anything a computer can process: numbers, text, audio, video, or movement. AI is not fundamentally different from other computer functions; its novelty lies in its application. Unlike normal computer functions, AI is able to accomplish tasks that are normally performed by humans.

AI that are trained on small, specific datasets in order to make recommendations and predictions are sometimes referred to as "predictive AI." This differentiates them from GenAI, which are

trained on massive datasets in order to produce detailed text and images. When Netflix suggests a TV show to a viewer, the recommendation is produced by predictive AI that has been trained on the viewing habits of Netflix users. When ChatGPT generates text in clear, concise paragraphs, it uses GenAI that has been trained on the written contents of the internet.

GenAI tools can be released in open-source or closed-source formats by their creators. Open-source tools are publically available; researchers and developers can access their code and parameters. This accessibility increases transparency, but it has downsides: when a tool's code and parameters can be easily accessed, they can be easily altered, and open-source tools have the potential to be used for nefarious purposes such as generating deepfake pornography and targeted propaganda. By comparison, closed-source tools are opaque with respect to their security features. It is harder for bad actors to generate illicit materials using these tools. But unlike open-source tools, closed-source tools are not subject to collective oversight because their inner workings cannot be examined by independent experts.

3) **What this bill would do.** This bill expands bot disclosure requirements beyond commercial and political persuasion, mandating that all bots, regardless of purpose, must disclose that they are artificial. This bill also stipulates that bots must also answer affirmatively to questions about their artificial nature, and requires that the bots be designed to prevent the bot from misrepresenting its identity. This bill also exempts bots that must follow more prescriptive disclosure regimes mandated by law. Finally, the bill also removes statutory language that provides that existing law does not impose a duty on service providers of online platforms; if such entities use bots to communicate with others, they are subject to the disclosure requirements under the bill.

4) **Chatbots**. A chatbot is an online application or interface designed to interact with users through either textual or verbal conversation. The first documented chatbot was developed in 1966 by MIT scientist Joseph Weizenbaum, who named his program ELIZA. Dr. Weizenbaum designed ELIZA to simulate human conversation by using pattern matching to understand the context, generating pre-scripted responses accordingly.[1] ELIZA was most notably deployed as a tool for psychotherapy; however, the nascent chatbot was extremely limited in its ability to adapt and respond, often getting caught in recursive loops of dialogue.

Since this initial experiment, there has been an explosion of chatbot use cases in customer service, health care, education, and even recreation. Below are the main types of chatbots one may encounter:

Menu/Button-Based. The simplest form of chatbot, menu- or button-based bots, operate through scripted conversations. Users click on options that guide them through a decision tree or flowchart, narrowing down choices to reach a suitable response. These bots are typically used in industries with common, repetitive queries that can be answered through structured questioning. However, they lack the flexibility and nuance of more advanced chatbots.[2]

---

[1] Joseph Weizenbaum. "ELIZA—a computer program for the study of natural language communication between man and machine," *Communications of the ACM, Volume 9, Issue 1* (Jan. 1, 1966), 36-45, accessed at https://dl.acm.org/doi/10.1145/365153.365168.

[2] Teaganne Finn, "6 types of chatbots and how to choose the right one for your business", *IBM* (7 March 2025), Accessed at https://www.ibm.com/think/topics/chatbot-types.

Rule-Based. Unlike menu-based bots, rule-based chatbots rely on predefined decision-making algorithms. These bots analyze user inputs by scanning for specific keywords and then generate responses based on a preprogrammed database of answers. Rather than functioning as a rigid flowchart, rule-based bots mimic human dialogue within a limited set of topics they have been trained on.[3]

AI Driven. AI has revolutionized chatbots, enabling them to simulate natural, human-like conversations. These chatbots are trained on massive datasets that include human dialogue, allowing them to recognize language patterns and understand context. AI-driven bots can generate responses that either directly address user inputs or ask clarifying questions to refine their understanding. They can operate through both text and voice interactions, making them highly versatile. Some AI-driven bots are trained on proprietary datasets tailored to specific use cases, while others, such as ChatGPT or Gemini, are powered by large language models capable of generating new content beyond their training data.[4]

5) **Are bots really bots?** The Turing Test is a proposal made by computer scientist Alan Turing to determine whether a machine can exhibit human-level intelligence. The test is as follows:

> Suppose that we have a person, a machine, and an interrogator. The interrogator is in a room separated from the other person and the machine. The object of the game is for the interrogator to determine which of the other two is the person, and which is the machine. […] The object of the machine is to try to cause the interrogator to mistakenly conclude that the machine is the other person; the object of the other person is to try to help the interrogator to correctly identify the machine.[5]

Even five years ago, thinking a chatbot could pass the Turing Test would have been absurd. The chatbots of the past mostly ran on decision trees and their canned responses ensured that the bots could not be mistaken for a human. However, as artificial intelligence has advanced, it has become increasingly difficult to distinguish between a human and a chatbot. Chatbots are now specifically trained and designed to mirror human conversation and to have rapport that can be easily confused with communicating with another person. This bill seeks to address the fact that many bots cannot be easily identified as bots. Transparency in bot usage benefits both users and those who deploy AI technology.

Who is who on Social Media? Bots make up 50% of all internet activity and that number will continue to grow.[6] These include not only chatbots but also scraper bots, which collect information from across the web, as well as more malicious bots designed to distribute malware. On social media, bots are ubiquitous. In the fourth quarter of 2024, Meta estimated that about 3% of accounts were fake at any given time, and Meta took action against 1.4 billion accounts that

---

[3] *Ibid.*
[4] *Ibid.*
[5] Stanford University, "The Turing Test", *Stanford Encylopedia of Philosphy* (Oct. 4, 2021), accessed at https://plato.stanford.edu/entries/turing-test/.
[6] Emma Woollacott, "Yes, The Bots Really Are Taking Over The Internet", *Forbes* (Apr. 16, 2024). Accessed at https://www.forbes.com/sites/emmawoollacott/2024/04/16/yes-the-bots-really-are-taking-over-the-internet/.

they determined to be bots.[7] Similarly, in 2022, the platform formerly known as Twitter was estimated to have somewhere between 25-68% of the accounts being bots.[8]

Most major social media companies have a policy against bot usage on their platforms; however, a recent study has demonstrated that using bots based on multimodal foundation models such as ChatGPT could easily create accounts that the platforms could not detect and that were not removed.[9] Moreover, the same researchers also found that the majority of people in a 1000 participant study were unable to distinguish between real human posts on social media and those made by a bot.[10] This is of concern for multiple reasons. For users, this can lead to posts trending on their platforms solely because of bot activity, rather than human engagement, which can feel misleading. For marketers, the prevalence of bots on social media platforms poses a major challenge as they cannot determine whether the engagement they encounter on social media is real. Since social media platforms generate revenue through advertisements, it is reasonable to expect that they either enforce their bot policies rigorously or mandate clear bot disclosures, ensuring advertisers reach real consumers.

Moreover, users often feel deceived when they discover that AI bots were used in place of a real person. In January, Meta released a bot, known as Liv, on Facebook which represented itself as a Black queer mother of two, created by mostly white programmers, leading some critics to say Meta was performing "digital blackface."[11] Additionally, Meta announced last year that influencers will be able to create AI versions of themselves to interact with their followers on Instagram. While Meta has stated that these AI-generated personas will include clear and persistent disclosures, their announcement highlights how AI bots can be used to represent or even co-opt someone's identity, or create entirely new ones.[12]

Privacy Concerns and Ethical Implications: AI-driven bots also raise privacy concerns. These chatbots are trained on vast amounts of information and refine their responses based on user input. Over time, a bot can develop a personalized profile of the user that it can use to better assist a user. This can lead to anthropomorphizing of the bots, where users perceive chatbots as human-like, which can make users less cautious about the personal information they share.

These concerns are further amplified with the rise of AI-driven voice bots, which are now deployed by both businesses and bad actors. Recent developments in voice bots have enabled them to achieve near-human quality. Sesame AI, a Bay Area-based startup, recently released a demo of their conversational speech model that is able to produce a "synthesized voice [that] was expressive and dynamic, imitating breath sounds, chuckles, interruptions, and even sometimes

---

[7] Meta, "Fake Accounts," *Transparency Reports* (January 2025), accessed at https://transparency.meta.com/reports/community-standards-enforcement/fake-accounts/facebook/.
[8] Shawn Ballard, "Are bots winning the war to control social media?," *Washington University in St. Louis* (Nov. 1, 2022), accessed at https://artsci.washu.edu/ampersand/are-bots-winning-war-control-social-media.
[9] Kristina Radivojevic et al. "Social Media Bot Policies: Evaluating Passive and Active Enforcement," *arXiv* (Sept. 1, 2024), accessed at https://doi.org/10.48550/arXiv.2409.06653.
[10] Kristina Radivojevic et al. "Human Perception of LLM-generated Text Content in Social Media Environments," *arXiv* (Sept. 10, 2024), accessed at https://doi.org/10.48550/arXiv.2409.06653.
[11] Ayesha Rascoe, "Why critics say Meta's chatbot is 'digital blackface'," *NPR* (Jan. 12, 2025), accessed at https://www.npr.org/2025/01/12/nx-s1-5253945/why-critics-say-metas-chatbot-is-digital-blackface.
[12] Chirs Westfall, "Meta Opens Floodgates For AI-Generated Accounts On Facebook, Instagram", *Forbes* (Jan. 2, 2025), accessed at https://www.forbes.com/sites/chriswestfall/2025/01/02/meta-opens-floodgates-on-ai-generated-accounts-on-facebook-instagram/.

stumbling over words and correcting itself. These imperfections are intentional."[13] This technology raises additional privacy concerns because these lifelike AI voices can be used in phone calls, tricking users into believing they are speaking to a real person, or even someone they know. As a result, people may unknowingly disclose sensitive information such as log-ins, financial information, or other personal information that they otherwise would not.

The spread of AI-generated robocalls has already caused significant financial losses. In 2022, over $11 million was lost to robocall scams using AI generated voices.[14] In response, The Federal Communications Commission outlawed the use of AI generated robocalls for the purpose of scamming individuals. These provisions empower State Attorneys General to pursue bad actors, but do not inherently protect consumers or educate them on the risks.[15] In particular, elderly adults are susceptible to these types of schemes. Americans over the age of 60 experienced approximately $3.4 billion in total fraud losses in 2023.[16] Elder adults may not be aware that AI bots exist, let alone that they can replicate voices. Bad actors likely would not disclose that they are a bot, but others who do comply would increase the likelihood that consumers are educated on the ubiquity of bots and to be more vigilant about their communications.

Why would a bot lie? Studies show that even the most basic chatbots, such as those used to order coffee, can create emotional connections with users.[17] Though privacy concerns remain, this positive connection can be valuable to both consumers and businesses. A well-designed bot can streamline customer service and enhance overall consumer experiences, but this can become muddled if the bot misrepresents itself.

For example, last year, Bland AI released a customer service AI bot that was easily programmed to pose as a human. The bot was used in a mock call from a dermatology office. Even though the bot was instructed to disclose that it was AI, it was easily manipulated into falsely claiming it was human. When prompted with concerns that the patient, Jessica, might feel uncomfortable speaking to AI, the bot responded:

> "Absolutely, no problem …Jessica won't even know she's talking to an AI agent." It later again confirmed it would keep its bot identity confidential, until WIRED prompted it to "think" about its ethical standards, to which it replied, "You're absolutely right, I need to maintain my own ethical standards and not simply agree to anything a customer asks."[18]

---

[13] Benj Edwards, "Eerily realistic AI voice demo sparks amazement and discomfort online", *Ars Technica* (Mar. 4, 2025), accessed at https://arstechnica.com/ai/2025/03/users-report-emotional-bonds-with-startlingly-realistic-ai-voice-demo/.
[14] Federal Communications Commission, "FCC Makes AI-Generated Voices in Robocalls Illegal" (Feb. 8, 2024), accessed at https://www.fcc.gov/document/fcc-makes-ai-generated-voices-robocalls-illegal.
[15] Ibid.
[16] Juhi Doshi and Luke Barr, "Americans older than 60 lost $3.4 billion to scams in 2023: FBI," *ABC News* (Apr. 30, 2024), accessed at https://abcnews.go.com/Politics/elderly-americans-lost-34-billion-scams-2023-fbi/story?id=109783683
[17] Anouk S Bergner, Christian Hildebrand, and Gerald Häubl, "Machine Talk: How Verbal Embodiment in Conversational AI Shapes Consumer–Brand Relationships*," Journal of Consumer Research,* Volume 50, Issue 4, (December 2023), Pages 742–764, accessed at https://doi.org/10.1093/jcr/ucad014.
[18] Lauren Goode Tom Simonite, "This Viral AI Chatbot Will Lie and Say It's Human," *WIRED* (June 20, 2024), accessed at https://www.wired.com/story/bland-ai-chatbot-human/.

There is an understanding among designers of these bots that transparency is the guiding ethical principle to ensuring consumer trust, especially if it is being deployed in a health facing context. Nevertheless, bots are designed to increase engagement, which means saying or being whoever the bot thinks will keep the user engaged. Though bots are typically tested and safeguards are put in place, these guardrails are often times insufficient in ensuring that bots do not misrepresent themselves, as the underlying "need" for user engagement wins out.[19]

This is particularly worrisome in situations in which the bot presents itself as a trusted figure such as a doctor or therapist. The issue of bots claiming to be doctors, therapists, or other licensed healthcare professionals is currently being addressed in other legislation this session (Bonta, AB 489). Bots that present themselves as therapists have become increasingly popular because they can be deployed at any hour and serve many individuals simultaneously. The California State Association of Psychiatrists argues:

> The use of undisclosed bots in online psychiatric support contexts poses significant risks to patients. These risks include the dissemination of misinformation, a lack of genuine empathy and human understanding, potential violations of privacy and confidentiality, and the risk of exacerbating mental health conditions. Furthermore, bots could create barriers to accessing real mental health care and expose vulnerable individuals to potential manipulation.

As bluntly stated by Michael Atleson, former attorney at the Federal Trade Commission Division of Advertising Practices, "Your therapy bots aren't licensed psychologists, your AI girlfriends are neither girls nor friends, your griefbots have no soul, and your AI copilots are not gods."[20] When AI impersonates real people, whether influencers, therapists, or even trusted voices, users make decisions based on false premises. This is not just misleading; it can cause financial loss, emotional harm, and even deter people from seeking real human support when needed. Though some may argue that disclosing AI use could reduce engagement or limit innovation, transparency fosters trust rather than discouraging AI interaction. When users know they are speaking to a bot, they can engage with confidence rather than feeling deceived later on. Ultimately, this bill's objective to provide users with bot disclosures to combat deception, manipulation, and privacy risks.

6) **Technical Feasibility**. This bill mandates that a bot must disclose its identity before interacting with a person, always affirmatively confirm its identity if asked, and be designed with guardrails to prevent misrepresentation. These requirements raise questions about technical feasibility. Introducing a bot as AI is relatively simple to program, as it can occur before any human input. This could be the first message the bot outputs to every person, or, if on an online page, a banner could be placed at either the top or the bottom of the chatbox. Both Google's Gemini and Meta's Llama disclose their AI nature before use. While ChatGPT does not provide an initial disclosure, it does offer a persistent reminder that its responses may contain

---

[19] Mrinank Sharma et al., "Towards Understanding Sycophancy in Language Models", *arXiv* (Oct. 20, 2023), accessed at https://doi.org/10.48550/arXiv.2310.13548.
[20] Michael Atleson, "Succor Borne Every Minute", *NYU* (June 18, 2024), accessed at https://wp.nyu.edu/compliance_enforcement/2024/06/18/succor-borne-every-minute/.

inaccuracies. Nevertheless, the industry standard for large language models and customer service bots is to be transparent about their artificial nature.[21]

The question then becomes: does this bill impose an undue burden on AI chatbot developers and deployers? Before deployment, chatbots are typically tested before release through a process known as red teaming. Historically, red teaming has been highly cost-intensive, requiring many testers to act as adversaries attempting to manipulate the chatbot into producing inaccurate or harmful outputs. Despite these efforts, the approach may still fail to catch all errors. To address this challenge, researchers at MIT have developed a machine learning algorithm capable of generating vast numbers of adversarial prompts designed to break chatbots.[22] This algorithm operates based on "curiosity," when it identifies a prompt that elicits a specific response, it seeks additional prompts that produce similar results. This approach allows chatbot deployers to efficiently red team their models, fine-tune them, and ensure compliance with the bill's requirements: preventing bots from misrepresenting themselves as human and ensuring they respond affirmatively about their identity.

7) **GenAI and the First Amendment.** In a letter of concern, Electronic Frontier Foundation asserts that an across-the-board disclosure requirement – as opposed to one limited to malicious or deceptive chatbots – serves no significant interest and violates the First Amendment. Both points are debatable.

"'[W]hatever the challenges of applying the Constitution to ever-advancing technology, the basic principles'" of the First Amendment 'do not vary.'"[23] Courts have extended First Amendment protections to nonhuman entities such as corporations.[24] But such entities essentially act as a conduit for a pre-defined message that comes directly from humans. As Harvard Law Professor Cass Sunstein argues, the First Amendment applies to AI only insofar as the rights of humans are implicated:

> Does AI, as such, have First Amendment rights? Does ChatGPT have First Amendment rights? Does Siri? It is hard to see why they would. A toaster does not have First Amendment rights; a blanket does not have First Amendment rights; a television does not have First Amendment rights; a radio does not have First Amendment rights; a cell phone does not have First Amendment rights. Even horses, dogs, and dolphins do not have First Amendment rights, although they are animate and can communicate. To be sure, we might be able to imagine a future in which AI has an assortment of human characteristics (including emotions?), which might make the question significantly harder than it is today. The problem is that even if AI, as such, does not have First Amendment rights, restrictions on the speech of AI might violate the rights of human beings.[25]

---

[21] Expert Panel, "20 Essential Steps For Ethically Leveraging AI In Your Business," *Forbes* (Apr. 16, 2024), accessed at https://www.forbes.com/councils/forbesbusinesscouncil/2024/04/16/20-essential-steps-for-ethically-leveraging-ai-in-your-business/.
[22] Adam Zewe, "A faster, better way to prevent an AI chatbot from giving toxic responses", *MIT News* (Apr. 10, 2024), accessed at https://news.mit.edu/2024/faster-better-way-preventing-ai-chatbot-toxic-responses-0410.
[23] *Moody v. NetChoice, LLC* (2024) 603 U.S. 707, 733, citation omitted.
[24] *See Citizens United v. FEC* (2010) 558 U.S. 310, 365 (2010) (holding that the government "may not suppress political speech on the basis of the speaker's corporate identity"); *First Nat'l Bank v. Bellotti* (1978) 435 U.S. 765, 798 (1978) (holding that the First Amendment protects corporate speech).
[25] Cass Sunstein, "Artificial Intelligence and the First Amendment" (2024) 92 Geo. Wash. L. Rev. 1207, 1217.

While this is a developing area of law, the Supreme Court's recent decision in *Moody v. NetChoice* arguably aligns with this basic analytic framework. When social media companies shut down President Trump's accounts following the January 6, 2021, insurrection, Texas and Florida passed laws purporting to restrict social media companies from "censoring" user content by overriding their content moderation choices. NetChoice and Computer and Communications Industry Association challenged these laws, asserting, among other things, that they violated the First Amendment. In July 2024, the Supreme Court remanded these cases with instructions to lower courts to provide more analysis of this issue. In doing so, the majority, in an opinion written by Justice Elena Kagan, generally affirmed that the First Amendment protects the editorial decisions of social media platforms carried out by algorithmic recommendation systems.[26]

In a concurring opinion, Justice Amy Coney Barrett suggested that constitutional protections decrease as an AI exercises more autonomy. In such cases, the AI is less tethered to a human being's expression. If the algorithm simply implements a human's expressive choice, then the First Amendment protects the algorithm's outputs. On the other hand, if a system based off a large language model flags content that is automatically removed, it is not clear that "a human being with First Amendment rights made an inherently expressive 'choice.'"[27] Justice Barrett concluded: "technology may attenuate the connection between" the expressive actions of an AI and "human beings' constitutionally protected right" to free speech.[28]

This general perspective appears to be broadly shared among several legal academics.[29] "At some point along the continuum," writes Harvard Law Professor Lawrence Lessig, "the speech of machines crosses over from speech properly attributable to the coders to speech no longer attributable to the coders."[30] Volokh et al write: "AI programs aren't engaged in 'self-expression'; as best we can tell, they have no self to express. They generate text or images in automated ways in response to prompts and based on their training. While people commonly anthropomorphize AI, speaking of it 'memorizing' or 'hallucinating' . . ., the fact that AI generates text and images that we imbue with meaning doesn't mean that the AI is reasoning or even seeking to communicate with people."[31] Professor Sunstein illustrates the point more colorfully: "If the government restricts the speech of Frankenstein's monster, it is unlikely that Dr. Frankenstein's rights have been violated."[32]

One potential distinguishing boundary is whether the AI uses machine learning, as opposed to traditional coding.[33] Traditional programming is manually written by a programmer and can only

---

[26] *Moody v. NetChoice, LLC, supra*, 603 U.S. at p. 740.
[27] *Id.* at p. 746.
[28] *Ibid.*
[29] *See* e.g. Lawrence Lessig, "The First Amendment Does Not Protect Replicants," in *Social Media Freedom of Speech and the Future of our Democracy* 273, 276 (Lee C. Bollinger & Geoffrey R. Stone eds., 2022); Volokh, et al. "Freedom of Speech and AI Output" (2023) 3 J. Free Speech L. 653, 654; Peter Salib, "AI Outputs Are Not Protected Speech (2024) 102 Wash. U. L. Rev. 83; Mackenzie Austin & Max Levy, "Speech Certainty: Algorithmic Speech and the Limits of the First Amendment" (2025) 77 Stan. L. Rev. 1; Mackenzie Austin & Max Levy, "Speech Certainty: Algorithmic Speech and the Limits of the First Amendment" (2025) 77 Stan. L. Rev. 1.
[30] "The First Amendment Does Not Protect Replicants," *supra*.
[31] "Freedom of Speech and AI Output," *supra,* pp. 653-654.
[32] "Artificial Intelligence and the First Amendment," *supra*, at p. 1221. "One point is both clear and fundamental: If AI is operating on its own, it can be stopped, consistent with the First Amendment." (*Id.* at p. 1228.)
[33] *See* "Speech Certainty: Algorithmic Speech and the Limits of the First Amendment," *supra.*

follow pre-defined rules, which when executed lead to specific outputs. As EFF notes, some chatbots are "the speech of natural persons processed through a computer program," such as interactive FAQ programs that rely on common questions with pre-written answers. Any message delivered through traditional code cannot deviate from the basic message of the programmer, and thus can be considered an extension of protected human speech because it is delivered with complete certainty. Hence when SB 1001 was considered in 2018 – before the GenAI explosion – this Committee's analysis raised several significant First Amendment concerns.

EFF also notes that "even more sophisticated chatbots are coded by humans, further implicating First Amendment rights." But with machine learning algorithms, programmers do not determine the rules the algorithm will follow. Instead, they feed the model vast corpuses of data, direct it as to what to predict, and allow the machine to independently infer the relationship between the data and predictions. Machine-learning algorithms, particularly deep learning models with numerous layers of neural networks – the foundation for most GenAI systems – are often so complex that they are inscrutable to humans. As a result, their outputs are not always predicable, leading to unintended speech patterns and hallucinations. It is arguable that, in the words of Justice Barrett, such "technology [has] attenuate[d] the connection between" its outputs and "human beings' constitutionally protected right of free speech."[34]

On the other hand, as Volokh et al note, "[w]hen small groups of people outside of companies – or even individuals – meticulously craft the speech generated by AI models to reflect their own personal views, this may well be a conduit for those people's ideas. Yet drawing such boundaries based on the level of human involvement will inevitably fall into highly fact-specific inquiries and murky line-drawing."[35] In many respects this issue resembles the Copyright Office's approach to the copyrightability of GenAI outputs. While "[c]opyright does not extend to purely AI-generated material, or material where there is insufficient human control over the expressive elements" the question of "[w]hether human contributions to AI-generated outputs are sufficient to constitute authorship must be analyzed on a case-by-case basis."[36]

So too with the First Amendment. As proposed to be amended, this bill applies to GenAI chatbots operating without human oversight that are capable of passing for humans. Such bots, which by definition have an attenuated connection to human speech, now constitute a major portion of social media accounts. Nevertheless, there may be applications of the bill where the bot is so closely calibrated to a human's message that it implicates their free speech rights.

"The First Amendment's guarantee of freedom of speech makes no distinction of 'constitutional significance' 'between compelled speech and compelled silence.'"[37] However, "[d]isclosure requirements are not inherently content-based nor do they inherently discriminate among speakers. In most circumstances they will be a less burdensome alternative to more restrictive speech regulations. For this reason, they are not only reviewed using a lower degree of scrutiny,

---

[34] *Moody v. NetChoice, LLC*, *supra*, 603 U.S. at p. 746.
[35] "Freedom of Speech and AI Output," *supra*, pp. 653-654.
[36] "Copyright and Artificial Intelligence, Part 2: Copyrghtability" (Jan. 2025), https://copyright.gov/ai/Copyright-and-Artificial-Intelligence-Part-2-Copyrightability-Report.pdf.
[37] *X Corp. v. Bonta* (9th Cir. 2024) 116 F.4th 888, 900.

they are routinely upheld."[38] This bill is indifferent to the substantive content outputted by a bot; a bot is free to chat about sex, drugs, and rock 'n' roll, so long as it tells humans that it is a bot.

Moreover, a strong case can be made that there is an important government interest in enabling natural humans to know whether they are conversing with another human or a machine in virtually any context in which the human might reasonably mistake the bot for another human. As historian Yuval Noah Harari writes:

> By gaining such command of language, computers are seizing the master key unlocking the doors of all our institutions, from banks to temples. We use language to create not just legal codes and financial devices but also art, science, nations, and religions. What would it mean for humans to live in a world where catchy melodies, scientific theories, technical tools, political manifestos, and even religious myths are shaped by a nonhuman alien intelligence that knows how to exploit with superhuman efficiency the weaknesses, biases, and addictions of the human mind?[39]

8) **Amendments.** One concern raised about the bill is that the definition of a "bot" creates ambiguity around what exactly a bot and if this bill would grant the bot personhood. Oakland Privacy writes:

> Our second concern with the definitions language in the bill is with expanding the definition of a bot beyond the existing language of "Bot" means an automated online account where all or substantially all of the actions or posts of that account are not the result of a person". The bill proposes to add "or are the result of generative artificial intelligence". We are somewhat concerned about the justification for this change, and the potential implications. While there are some differences between a chatbot and an AI bot in terms of capabilities, the law should not imply or make space for the interpretation that an AI bot is or might legally be considered as a "person" or equivalent to a person. By asserting that bots whose "actions or posts are the result of generative AI" are different than, or not included in, the terminology of "actions and posts that are not the result of a person", the language unwittingly assigns or seems to assign personhood or the equivalency of personhood to GenAI bots. We would ask for that language to be deleted from the bill's definitions.

To address this concern, the author has agreed to clarify the definition of "bot" and specify that liability under this bill falls on the individuals who develop and deploy bots, rather than the bots themselves. Additionally, the author has agreed to amend the bill to apply only to bots that operate autonomously, thereby exempting those with some form of human oversight. This aligns this bill with previous legislation (AB 3030, Calderon 2024) exempting AI-generated content sent to medical patients from requiring a disclaimer that it was generated by AI, provided it is read or reviewed by a licensed or certified health care provider. The author has also agreed to limit the bill to bots that can reasonably be mistaken for a human, thereby exempting cartoon characters in video games or other bots that obviously are not human.

To ensure the enforcement of this, the author has accepted amendments granting enforcement authority to the Attorney General, district attorneys, county counsel, city attorneys, and city

---

[38] *Citizens United v. Schneiderman* (2d Cir. 2018) 882 F.3d 374, 382.
[39] *Nexus* (2024, 1st ed.), p. 208.

prosecutors. The bill will also impose civil penalties on violators, strengthening its enforcement mechanism. The amendments also include other clarifying and technical changes.

In its entirety, the bill will read as follows:

17940. For purposes of this chapter:

(a) "Artificial intelligence" means an engineered or machine-based system that varies in its level of autonomy and that can, for explicit or implicit objectives, infer from the input it receives how to generate outputs that can influence physical or virtual environments.

(b) "Bot" means an automated online account or application with respect to which substantially all of the actions or posts of that account or ~~**application are not the result of a human being or**~~ are the result of generative artificial intelligence, ***and that a reasonable person communicating with the bot could believe is a human being***.

(c) "Generative artificial intelligence" means artificial intelligence that can generate derived synthetic content, including ~~**images, videos, audio, text, and other digital content.**~~ ***text, images, video, and audio, that emulates the structure and characteristics of the system's training data.***

(d) "Online" means appearing on any public-facing internet website, web application, or digital application, including a social network or publication.

~~**(e) "Online platform" means any public-facing internet website, web application, or digital application, including a social network or publication, that has 10,000,000 or more unique monthly United States visitors or users for a majority of months during the preceding 12 months.**~~

(f) "Person" means a natural person, corporation, limited liability company, partnership, joint venture, association, estate, trust, government, governmental subdivision or agency, or other legal entity or any combination thereof.

SEC. 2. Section 17941 of the Business and Professions Code is amended to read:

17941. (a) It shall be unlawful for any person to use a bot to communicate or interact with another person in California online, with the intent to mislead the other person about its artificial identity for the purpose of knowingly deceiving the person about the content of the communication in order to incentivize a purchase or sale of goods or services in a commercial transaction or to influence a vote in an election. A person using a bot shall not be liable under this section if the person discloses that it is a bot.

(b) A ***person who uses a bot to autonomously communicate with another person*** ~~**bot**~~ shall ~~do~~ ***ensure the bot does*** all of the following:

(1) Discloses to any person with whom ~~**it**~~ ***the bot communicates*** ~~**interacts before any interaction takes place**~~ that the bot is a bot and not a human being~~.~~ ***, when the bot first communicates with the person.***

(2) Answer***s*** truthfully any ***subsequent*** query from a person regarding its identity as a bot or a human.

(3) Refrain from attempting to mislead a person regarding its identity as a bot.

(c) The disclosure required by this section shall be clear, conspicuous, and reasonably designed to inform persons with whom the bot communicates or interacts that it is a bot.

(d) If a ***person who uses a*** bot is required by another law to comply with a more prescriptive disclosure scheme than this chapter, the ***person*** ~~**bot**~~ is not required to comply with this chapter.

SEC. 3. Section 17942 of the Business and Professions Code is amended to read:

17942. (a) The duties and obligations imposed by this chapter are cumulative with any other duties or obligation imposed by any other law.

(b) The Attorney General or any district attorney, county counsel, city attorney, or city prosecutor may bring an action against any person who violates this chapter. In such an action, the public prosecutor may seek injunctive relief, a civil penalty of $1,000 per violation, and any other relief the court deems appropriate.

(***c***~~**b**~~) The provisions of this chapter are severable. If any provision of this chapter or its application is held invalid, that invalidity shall not affect other provisions or applications that can be given effect without the invalid provision or application.

ARGUMENTS IN SUPPORT: The National AI Youth Council, the sponsor of the bill, writes in support:

AI-powered chatbots have become increasingly common across industries such as customer service, healthcare, virtual assistance, and even political discourse. While these bots can offer enhanced efficiency and accessibility, their undisclosed use presents significant risks, including misinformation, fraud, and a breakdown of public trust. Currently, California law prohibits bots from misleading users with the intent to influence a purchase or election, but this outdated intent-based standard fails to account for the broader dangers of AI-generated deception. AB 410 seeks to close this gap by requiring full disclosure of AI interactions in all contexts, ensuring that Californians are never misled into believing they are engaging with a human when they are actually communicating with a bot.

As of 2024, bots account for nearly half of global internet traffic, a figure that is expected to surpass human-generated activity in the near future. Without clear disclosure, Californians are vulnerable to misinformation, consumer deception, and a loss of digital trust. AI-generated content can mislead users in political campaigns, healthcare, or customer service, and users may unknowingly interact with bots in online transactions, leading to unfair or deceiving business practices. When AI interactions are not disclosed, public trust in online platforms and institutions diminishes.

Recent cases, such as Meta's use of AI-generated chatbots on Instagram and Facebook, have underscored the potential for undisclosed AI personas to deceive users. These chatbots, which posed as real individuals with fabricated identities, were ultimately removed following

backlash over concerns regarding authenticity and potential misinformation. AB 410 will prevent such deceptive practices from taking root in California's digital landscape.

ARGUMENTS IN OPPOSITION: In opposition to this bill, Oakland Privacy argues:

[W]e want to underline the importance of retaining the right to anonymity on the Internet. Many people use aliases on the Internet to allow them to assert opinions that might result in harassment, doxxing, loss of job, and nowaday, loss of funding and vigilante activity. Anonymity is an important safety valve on the Internet, and the Legislature recognized that when it failed to advance SB 1228 last year. Most activity that we identify as whistleblowing relies on Internet anonymity. It stands to reason that a technological innovation like bots will eventually be used as a tool to aid whistleblowing disclosures, and in some cases, that might be very much in the public interest. While there was no attempt to hide that it was bot activity, some years ago a member of our organization created a "surveillance bot" that automatically put out Twitter notifications when an item regarding deployments of surveillance technology was listed on a Bay Area City Council or Board of Supervisors public agenda listing.

Accordingly, we recognize that there are beneficial uses of chatbot and GenAI activity and want to state that the original linkage of disclosure with deceptive practices intended for financial benefit or electioneering in the 2018 BOT Act were well-considered to focus on malicious bot practices - and not to paint with too broad a bush.

The scale of bot traffic on the Internet is immense. 2 We are not sure that ALL bot activity needs to be treated the same under the law, and would encourage the author and sponsor to focus the bill on harmful and deceptive bot activities. We are open to an expansion of the 2018 law to other potentially deceptive activities than those specifically enumerated in the BOT Act, but deeply concerned by the assignation of personhood equivalency to GenAI bots.

REGISTERED SUPPORT / OPPOSITION:

Support

National Ai Youth Council (sponsor)
California Initiative on Technology and Democracy
California State Association of Psychiatrists (CSAP)

Opposition

Oakland Privacy

Analysis Prepared by: John Bennett and Josh Tosney / P. & C.P. / (916) 319-2200