The Human Signal:

Reader-Facing AI Detection, the Disclosure Field, and the Tax on Assisted Authorship

Travis Gilly* ❧ August 2026 (v0.5)

Abstract

In July 2026, Substack integrated a commercial AI detector into its reading surface, letting any reader scan any newly published post and receive an estimate of how much of the text a machine produced, displayed as a percentage beside the author's name. Announcing a funding round the following week, the detection company's chief executive stated the mission in a sentence: "If we do not actively discriminate in favor of human content, then we're just gonna see more and more AI, and it's just gonna drown out any human signal that we have." This Article takes the sentence at its word and asks the question it leaves open: who counts as human when authorship runs through assistive technology. The instrument examined here is the third of its kind, after the invisible watermark and the crowdsourced authenticity flag, and it perfects what the first two began, because it ships with the remedy attached. Beside the detection badge sits a voluntary disclosure field where the author may explain how the work is made. The Article's central claim is that the pairing converts the field into a tax assessed by disability. A nondisabled writer clears an adverse badge with a style preference; a writer whose work runs through dictation software and a language model clears it only by publishing a diagnosis. The exculpatory answer is priced differently for the two populations, silence becomes an answer once the field exists, and rising detector accuracy raises the tax rather than curing it, because the harm was never the error rate. The harm is a taxonomy with no category for assisted authorship, enforced by an increasingly reliable instrument. The Article then shows that every accommodation-shaped repair fails a one-sentence test, since a badge, an exemption tag, or a visible opt-out that permits any inference of disability is itself the disclosure, and it specifies the universal-design remedies that pass. It closes with the doctrine: the screen-out and methods-of-administration provisions of ADA Title III, and the New York City Human Rights Law, which reaches the detection company where it lives.

Keywords: content provenance; assistive technology; compelled disclosure; disability discrimination; disparate impact; universal design; public accommodation; screen-out; disclosure penalty; New York City Human Rights Law; platform governance; forced intimacy.

---

*Travis Gilly is Founder and Executive Director of the Real Safety AI Foundation. ORCID: 0009-0007-2954-6313. Correspondence: t.gilly@ai-literacy-labs.org. The author discloses the use of speech-to-text software (Wispr Flow) and Anthropic's Claude as assistive technology for a diagnosed neurodevelopmental condition. All legal analysis, case synthesis, theoretical framing, doctrinal arguments, and conclusions are solely the author's own.

Contents

I. Introduction

On July 21, 2026, Substack turned every reader into an inspector. The platform integrated Pangram, a commercial AI detection system, into its reading surface, so that any post, note, or reply published from that day forward can be scanned in place, returning an estimate of how much of the text was written by a human being and how much with machine assistance.[1] The result renders as a percentage. The percentage sits beside the author's name.[2]

Eight days later, the detection company announced a nine million dollar funding round, and its chief executive compressed the business model into one sentence: "If we do not actively discriminate in favor of human content, then we're just gonna see more and more AI, and it's just gonna drown out any human signal that we have."[3] This Article takes the executive at his word. The product discriminates in favor of human content, which is to say it discriminates, and the only question is where the line labeled human falls. This Article's answer is that the line falls through the disabled population, and that the instrument's most celebrated feature, a voluntary disclosure field offered as the author's remedy, is where the discrimination completes itself.

The instrument examined here is the third of its kind. The first was the invisible watermark: a mark embedded by the model provider inside generated text, mandated in the European Union and shipped worldwide by at least one major provider, detectable later by institutions granted the means to look.[4] The second was the crowdsourced flag: a report option inviting readers of a professional platform to accuse a post of reading like machine

---

1. Substack (@SubstackInc), X (July 21, 2026), `https://x.com/Substack/status/2079598704424779787` (last visited Aug. 14, 2026) (announcing the integration and the author-side controls).

2. Sarah Perez, *Substack's New Tool Tells You Who's Been Writing Their Newsletters with AI*, TechCrunch (July 22, 2026), `https://techcrunch.com/2026/07/22/substacks-new-tool-tells-you-whos-been-writing-their-newsletter` (last visited Aug. 14, 2026).

3. Rebecca Bellan, *As AI Content Floods the Internet, Pangram Raises $9M to Detect It*, TechCrunch (July 29, 2026), `https://techcrunch.com/2026/07/29/as-ai-content-floods-the-internet-pangram-raises-9m-to-detect-i` (last visited Aug. 14, 2026) (quoting Pangram chief executive Max Spero).

4. Regulation (EU) 2024/1689 (Artificial Intelligence Act) art. 50(2), 2024 O.J. (L 1689); European Commission, *Guidelines on the Implementation of the Transparency Obligations for Certain AI Systems Under Article 50 of the AI Act* (July 20, 2026).

output, with the accusations fed forward as training signal.[5] The third is the detector badge: platform-integrated, reader-facing, rendered as a number, and, critically, shipped together with its own answer mechanism. Substack's integration includes a field titled "How I make this," in which an author may describe the process behind the work, and a per-post switch by which an author may disable detection entirely.[6]

The pairing of badge and field looks like fairness. An accusation is displayed; a reply channel is provided; the author speaks in her own words. This Article's central claim is that the pairing is the injury. Once the field exists, silence becomes an answer. Once the badge is reliable, the field becomes the only exit. And the exit is priced differently depending on who is leaving. A nondisabled writer facing an adverse badge clears it with a style preference: I draft fast, I use an editor, I write in a plain register. A writer whose work necessarily runs through dictation software and a language model clears it only one way, by explaining that the machine in the text is a medical fact about the author. For one population the field asks about craft. For the other it asks for a diagnosis. That is a tax assessed by disability, collected at the point of publication, and no improvement in detection accuracy reduces it. Accuracy raises it, because the more reliable the badge, the more binding the label, and the higher the cost of staying silent.

The Article proceeds in six parts. Part II describes the instrument as shipped, including the accuracy generation that arrived days after launch and the vendor's own account of what the product is for. Part III sets out the empirical record the instrument inherits: human readers cannot perform the discrimination task the badge performs for them, their confidence bears no relation to their accuracy, evaluators penalize disclosed machine assistance even where the work is accurate and the author is human, and the writers most likely to trip the instruments write in registers associated with disability, second-language acquisition, and assistive technology. Part IV states

5. The flagging instrument, shipped by LinkedIn on July 30, 2026 under the label "Seems like AI slop," collects accusations without any countervailing judgment and without ground truth; the classifier it trains learns complaint prevalence rather than authorship. The empirical record on such instruments is taken up in Part III.

6. Substack, *supra* note 1 (describing the disclosure field and per-post controls); *see also* Igor Bonifacic, *Substack Is Adding an AI Detection Feature*, Engadget (July 21, 2026), `https://www.engadget.com/2220064/substack-is-adding-an-ai-detection-feature/` (last visited Aug. 14, 2026) (noting the platform's acknowledgment of the detector's limits at launch).

the disclosure tax precisely and locates it in the doctrine of disclosure control: the settled principle, running from the statutory inquiry restrictions to the service-animal documentation rules, that disability information belongs to the disabled person, who decides when, where, and whether. Part V confronts the remedies and shows that every accommodation-shaped repair reproduces the harm, because a marked class is a disclosed class; it specifies the universal-design repairs that survive a one-sentence test. Part VI supplies the law: the screen-out and methods-of-administration provisions of Title III of the Americans with Disabilities Act, the circuit geography that governs where those provisions can reach a web-native defendant, and the New York City Human Rights Law, which reaches the detection company in the borough where it was founded. Part VII answers the strongest objection, that machine-generated flooding is a real problem, by agreeing, and then showing that flooding is an identity problem living at the account layer, which detection at the authorship-tool layer does not solve and cannot solve. The conclusion returns to the executive's sentence and answers its open question.

One note on method. The author's own publishing runs through the assistive pipeline this Article describes, a fact disclosed on the face of everything he publishes, in his own words, in the same form every time.[7] In the three weeks after the integration launched, the same workflow, applied by the same author, returned detection results ranging from entirely human to entirely machine across different posts, a spread the Article uses in Part IV not as a complaint about error but as evidence of what the instrument measures and what it cannot.

II. The Instrument as Shipped

A. The Badge

The integration operates on the reading surface. A reader viewing any post published on or after July 21, 2026 may invoke a scan; the detector returns a segment-level estimate of machine involvement, aggregated and displayed as a percentage split between human-written and

7. See author's note, *supra* (star footnote). The disclosure practice predates every instrument discussed in this Article.

AI-assisted text.[8] The scan is available to every reader, on every eligible post, at no cost, in the application itself. No institutional gatekeeping mediates access to the result. This distinguishes the badge from the watermark regime, in which detection results flow to a defined class of expert users, and from the crowd flag, in which the accusation flows to the platform's classifier rather than to the public. The badge is an accusation published to the same audience as the work.

B. The Accuracy Generation

Eight days after the Substack launch, the vendor shipped its next detection model, announcing that the new system is "over 99% accurate at finding AI-assisted writing and mixed human-AI content" and better at defeating so-called humanizer tools that lightly rewrite machine output.[9] Independent evaluation had already placed the vendor at the top of its category on the error dimension that matters most to accusers: a University of Chicago study found it "the only tool to satisfy a strict cap (FPR ≤ 0.005) without sacrificing accuracy."[10] The platform, for its part, acknowledged at launch that no detector is fully reliable and that edited machine text is harder to classify.[11]

The temptation is to litigate those numbers. This Article declines. Its argument does not require the detector to be wrong, and is in fact strongest where the detector is right. A classifier that reliably identifies machine involvement in text, deployed as a public badge, correctly identifies the output of an assistive pipeline as machine-involved, because it is. The dictated draft was cleaned by a model. The composed draft was generated at the author's direction and revised at the author's hand. Machine involvement is present; the instrument reports it; the report is accurate. The discrimination enters at the next step, in what the label is taken to mean and what it costs different authors to correct the meaning. The taxonomy has two categories, human

8. Substack, *supra* note 1; Perez, *supra* note 2.

9. Bellan, *supra* note 3.

10. Press Release, Pangram Labs, AI Transparency Company Pangram Integrated into Two Leading News Source and Resource Platforms (Sept. 15, 2025), `https://www.businesswire.com/news/home/20250915315368/en` (last visited Aug. 14, 2026) (quoting the study).

11. Bonifacic, *supra* note 6.

and machine, and assisted authorship belongs to neither. An accurate instrument enforcing an incomplete taxonomy does not become fair by being accurate. It becomes thorough.

C. The Field

Beside the badge sits the remedy as the platform conceives it: a "How I make this" field, author-controlled, freeform, displayed to readers who scan the work, in which the author may explain the process behind the writing.[12] Authors may also disable detection on individual posts.[13] The design assumes the field is neutral: a channel, available to all, costing each author the same. Part IV shows the assumption is false. The field costs different authors different things, the difference tracks disability, and the per-post disable switch is itself a signal, because a writer whose posts alone cannot be scanned wears a different badge, one the reader fills in.

III. The Record the Instrument Inherits

A. Readers Cannot Perform the Task

The badge sells a judgment readers cannot make themselves. The finding is stable across the literature: human evaluators cannot reliably distinguish machine-generated from human-written text, and their confidence in the judgment bears no measurable relation to their accuracy.[14] The failure is not confined to lay readers: trained applied linguists could not identify authorship of research abstracts, and blinded physician reviewers rating clinical abstracts returned

12. Substack, *supra* note 1.

13. *Id.*

14. Umair Ullah et al., *Breaking the Imitation Game: Can LLMs Fool Humans and Machines Alike?*, 2026 Int'l J. Intelligent Sys. art. 8117975, `https://doi.org/10.1155/int/8117975` (last visited Aug. 14, 2026) (survey respondents, most of them familiar with machine-generated content, correctly classified 15 percent of model-generated passages; the authors conclude that correct identifications reflect individual intuition rather than recognition of stable linguistic patterns, and that human evaluators lack the perceptual signals needed to differentiate fluent model text from human writing); Emmanuel Mayor, Lucas M. Bietti & Adrian Bangerter, *Can Large Language Models Simulate Spoken Human Conversations?*, 49 Cognitive Sci. art. e70106 (2025), `https://doi.org/10.1111/cogs.70106` (last visited Aug. 14, 2026) (across transcript conditions, most judgments fell at chance, and authentic human openings were judged machine-produced significantly more often than chance).

authorship confidence statistically consistent with an inability to discern authorship at all.[15] The instruments, in short, are sold as a prosthesis for a judgment their buyers have been shown not to possess.[16] Institutions that ran the experiment at scale drew the conclusion in public: Vanderbilt University disabled its vendor's AI detector after concluding the results could not be trusted for consequential decisions.[17] A reader who cannot perform the task, handed a percentage that performs it for her, does not become a better judge. She becomes a confident one. The badge does not inform the reader's judgment; it replaces it, at face value, which is the one epistemic habit the technology's own advocates say the flood of machine text should have cured.

B. The Writers the Instruments Misread

The second finding identifies who pays when the instruments err or when their taxonomy runs out. Detection systems disproportionately flag writers whose prose exhibits regularity: consistent grammar, disciplined structure, low burstiness, reduced idiomatic variation. The flagship demonstration involved non-native English writers, whose TOEFL essays were classified as machine-generated at rates exceeding sixty percent by detectors that performed near-perfectly on native-speaker text.[18] The same registers characterize autistic writing, which is frequently rule-consistent and structurally literal, and the output of assistive pipelines generally, which by

---

15. Kyle Kendro, Jenna Maloney & Scott Jarvis, *Do Large Language Models Produce Texts with "Human-Like" Lexical Diversity? Evidence from Four ChatGPT Models*, Int'l J. Applied Linguistics (2026), `https://doi.org/10.1111/ijal.70115` (last visited Aug. 14, 2026) (reviewing the experimental record, including studies in which poetry judges, applied linguists, and physics students each failed to identify machine authorship); Kyle W. Lawrence et al., *Human Versus Artificial Intelligence-Generated Arthroplasty Literature: A Single-Blinded Analysis of Perceived Communication, Quality, and Authorship Source*, 20 Int'l J. Med. Robotics & Computer Assisted Surgery art. e2621 (2024), `https://doi.org/10.1002/rcs.2621` (last visited Aug. 14, 2026).

16. *See* Debora Weber-Wulff et al., *Testing of Detection Tools for AI-Generated Text*, 19 Int'l J. for Educ. Integrity art. 26 (2023), `https://doi.org/10.1007/s40979-023-00146-z` (last visited Aug. 14, 2026) (evaluating detection accuracy across tools and conditions); Ahmed M. Elkhatat, Khaled Elsaid & Saeed Almeer, *Evaluating the Efficacy of AI Content Detection Tools in Differentiating Between Human and AI-Generated Text*, 19 Int'l J. for Educ. Integrity art. 17 (2023), `https://doi.org/10.1007/s40979-023-00140-5` (last visited Aug. 14, 2026).

17. Michael Coley, *Guidance on AI Detection and Why We're Disabling Turnitin's AI Detector*, Vanderbilt Univ. Brightspace (Aug. 16, 2023), `https://www.vanderbilt.edu/brightspace/2023/08/16/guidance-on-ai-detection-and-why-were-disabling-turnitins-ai-detector/` (last visited Aug. 14, 2026).

18. Weixin Liang et al., *GPT Detectors Are Biased Against Non-Native English Writers*, 4 Patterns 100779 (2023), `https://doi.org/10.1016/j.patter.2023.100779` (last visited Aug. 14, 2026).

design normalizes disfluency into clean prose.[19] The population sitting closest to the classifier's decision boundary is the population writing through accommodations.

C. The Disclosure Penalty

The third finding forecloses the cheerful answer that authors should simply explain themselves, and it is now among the best-replicated results in the human-AI evaluation literature. Across sixteen preregistered experiments with 27,491 participants, evaluations of creative writing fell when evaluators believed the text was produced by an AI model "or with the help of one," an effect the authors name the AI disclosure penalty, mediated by perceived authenticity, and "sticky," persisting across evaluation metrics, contexts, content types, and every mitigation intervention the literature had proposed.[20] A parallel program of thirteen experiments across tasks and evaluator roles found that actors disclosing AI use are trusted less than those who do not, that the penalty operates through diminished perceived legitimacy, and that it holds "regardless of whether disclosure is voluntary or mandatory."[21] The penalty is not confined to attitude scales. When a crowdfunding platform made AI disclosure mandatory, a study exploiting the policy as a natural experiment estimated that the disclosure reduced funds raised by 39.8 percent and backer counts by 23.9 percent.[22] And the penalty concentrates precisely where newsletter writing lives: the sharpest perception declines attach to social and interpersonal

19. The disfluency-stripping function is not incidental; it is the accommodation. A dictation pipeline that preserved every false start would not be assistive. The polish the classifier reads as machine signature is the access working as intended.

20. Manav Raj, Justin M. Berg & Rob Seamans, *The Artificial Intelligence Disclosure Penalty: Humans Persistently Devalue AI-Generated Creative Writing*, 155 J. Experimental Psych.: Gen. 896, 896–915 (2026), `https://doi.org/10.1037/xge0001889` (last visited Aug. 14, 2026).

21. Oliver Schilke & Martin Reimann, *The Transparency Dilemma: How AI Disclosure Erodes Trust*, 188 Org. Behav. & Hum. Decision Processes 104405 (2025), `https://doi.org/10.1016/j.obhdp.2025.104405` (last visited Aug. 14, 2026).

22. Ning Wang & Chen Liang, *How to Disclose? Strategic AI Disclosure in Crowdfunding*, arXiv:2602.15698 (2026), `https://doi.org/10.48550/arxiv.2602.15698` (last visited Aug. 14, 2026) (exploiting Kickstarter's 2023 mandatory AI disclosure policy as a natural experiment).

writing, the registers of personal address in which readers experience machine involvement as a loss of human sincerity.[23]

The record carries an honest qualification, and the qualification sharpens rather than softens the argument. A 2026 systematic review of the journalism literature found no uniform AI penalty for news content: for routine, data-driven reporting, provenance and disclosure cues frequently produced null or conditional effects.[24] The penalty, in other words, is domain-sensitive: weakest where writing is impersonal and informational, strongest where it is creative, personal, and voiced. A subscription newsletter is the second kind of writing by construction; the product a Substack author sells is voice. The instrument examined here therefore deploys its badge in the exact register where the penalty the badge triggers is at its documented maximum.

One further finding from the trust program deserves separate statement, because it prices the badge itself. The disclosure penalty, though robust, is "comparatively weaker than the effect of third-party exposure": being revealed by someone else damages trust more than revealing oneself.[25] A reader-invoked detection badge is third-party exposure by design. The empirical literature thus ranks the instrument's two surfaces, and the ranking runs against the accused author twice: the badge (exposure) costs more than the field (disclosure), and the field costs more than silence used to cost before the field existed. Every path through the architecture is priced, and every price is higher for the author whose machine involvement is medical.

---

23. Hiroki Nakano, Jo Takezawa & Fabrice Matulic, *Understanding Reader Perception Shifts upon Disclosure of AI Authorship*, arXiv:2510.24011 (2025), `https://doi.org/10.48550/arxiv.2510.24011` (last visited Aug. 14, 2026) (261 participants; disclosure eroded perceived trustworthiness, caring, competence, and likability, with the sharpest declines in social and interpersonal writing).

24. Lorena Liçenji & Julian Hoxha, *When News Is "Written by Artificial Intelligence": A Systematic Review of Provenance and Disclosure Cues in Journalism and Their Effects on Credibility and Trust*, 9 Frontiers A.I. art. 1815243 (2026), `https://doi.org/10.3389/frai.2026.1815243` (last visited Aug. 14, 2026).

25. Schilke & Reimann, *supra* note 19 (within-paper meta-analysis).

IV. The Tax

A. Disclosure Control as the Baseline Right

Disability law's settled architecture treats disability information as the disabled person's to control. The employment title forbids disability-related inquiries of employees except where job-related and consistent with business necessity, and walls off what is learned behind confidentiality requirements.[26] The public-accommodations regulations run access without documentation demands even where fraud is the stated fear: staff confronting a service animal may ask two narrow questions and "shall not require documentation" or ask "about the nature or extent of a person's disability."[27] The regulatory system, in other words, already knows the principle: conditioning ordinary participation on the production of one's diagnosis is itself the discrimination, prior to and independent of whatever the person was trying to access. Disability scholarship names the lived form of the same principle forced intimacy, the standing expectation that disabled people surrender private parts of themselves as the admission price of what nondisabled people receive by default.[28]

B. The Price Structure of the Field

Against that baseline, price the field. An author receives an adverse badge: the number beside her name says the machine wrote most of this. The field invites her answer. Consider the answers available.

The nondisabled author's truthful exculpatory answers are style facts. I write quickly and cleanly. I use an editor. I dictate on walks and polish later, by choice. Each answer costs nothing the law protects. Each is the kind of craft talk authors volunteer in interviews.

---

26. 42 U.S.C. § 12112(d)(4)(A) (prohibiting inquiries "as to whether such employee is an individual with a disability or as to the nature or severity of the disability" unless job-related and consistent with business necessity).

27. 28 C.F.R. § 36.302(c)(6) (current through the eCFR edition of Aug. 1, 2026; last amended Dec. 19, 2016).

28. Mia Mingus, *Forced Intimacy: An Ableist Norm*, Leaving Evidence (Aug. 6, 2017), `https://leavingevidence.wordpress.com/2017/08/06/forced-intimacy-an-ableist-norm/` (last visited Aug. 14, 2026).

The assisted author's truthful exculpatory answer is a medical fact. The machine signature in my prose is my accommodation; I write through dictation software and a language model because of a condition. There is no version of the answer that clears the badge while withholding the condition, because the condition is the explanation. A general answer, I use tools, concedes the badge's premise and absorbs the disclosure penalty without rebutting the inference that matters to the reader who scanned. Only the accommodation answer converts the machine signature from a choice into a necessity, and the accommodation answer is a diagnosis, published beside the work, permanently, to every reader who looks.

That is the tax. The field is facially neutral and open to all; the price of using it effectively is disability disclosure for one population and small talk for the other. A facially neutral demand that can be answered, in substance, only by revealing protected status operates as an inquiry into that status, and the inquiry restrictions above are the doctrine's recognition that such demands are regulated conduct, not neutral channels.[29]

C. Silence Priced, and Accuracy as Enforcement

Two features complete the mechanism. First, the field prices silence. Before the field existed, an unexplained adverse badge was an unexplained adverse badge. After, an empty field beneath a high number is itself legible: the author was offered the microphone and declined. The remedy's existence converts non-disclosure from a default into a statement. Constructive compulsion needs no order compelling speech; it needs only an architecture in which every option speaks. And the architecture defeats even the author who was always going to disclose. An author with a standing voluntary practice of describing her assistive process, in her own words, on her own timing, does not escape the mechanism, because the badge and the disclosure travel in separate channels: the scan is reader-invoked and arrives with the authority of an instrument, independent of anything the author wrote, so her voluntary account no longer introduces her process. It rebuts a finding that has already landed. The empirical ranking in Part III.C gives the channel difference its

29. *Cf.* 42 U.S.C. § 12112(d)(4)(A); 28 C.F.R. § 36.302(c)(6). The formal structures differ across titles; the shared principle, that the diagnosis is not the price of participation, does not.

price: third-party exposure damages trust more than self-disclosure, and the badge is third-party exposure every time it is invoked.

Second, accuracy enforces. The vendor's newest model closes the category in which assisted authors previously escaped notice, mixed and edited text, and advertises the closure as the headline capability.[30] The author's own three weeks under the integration illustrate the shift: identical workflow, identical author, results spanning fully human to fully machine as the deployed models turned over. The spread is not an indictment of the instrument's competence. It is a demonstration that the quantity measured, the statistical fingerprint of the prose, floats free of the quantity the badge is read to convey, the humanity of the author. As the fingerprint reading grows more reliable, the badge grows more binding, the empty field grows louder, and the tax rises. The harm was never the error rate. The harm is a two-category taxonomy, human and machine, with no category for the person whose humanity arrives by machine, enforced by an instrument that no longer misses.

D. What the Accommodation Literature Already Knows

The claim that a disclosure demand suppresses participation rather than enabling it is not a prediction. It is among the most consistent findings in the empirical literature on disability accommodation, measured in populations with every professional incentive to ask.

Among first-year resident physicians reporting a disability and a need for accommodations, more than half did not request them, and the most frequently cited reason, given by roughly six in ten non-requesters, was fear of stigma or bias.[31] A national study of internal medicine residents three years later found the same pattern with the stigma figure higher still, cited by 82 percent of those who needed accommodations and did not seek them, and found the suppression concentrated among residents with cognitive disabilities, who had markedly

30. Bellan, *supra* note 3 ("over 99% accurate at finding AI-assisted writing and mixed human-AI content").

31. Karina Pereira-Lima et al., *Barriers to Disclosure of Disability and Request for Accommodations Among First-Year Resident Physicians in the US*, 6 JAMA Network Open art. e239981 (2023), `https://doi.org/10.1001/jamanetworkopen.2023.9981` (last visited Aug. 14, 2026).

lower odds of both access and request.[32] Experimental work in management research traces the causal chain directly, showing that perceived disability stigma drives the withholding of accommodation requests.[33] And qualitative work names the structure this Article has been describing: disclosure is a double-edged sword, because the accommodation that secures access simultaneously operates as a stigma symbol, with one participant describing the act of requesting as reminding everyone that she is disabled.[34]

Two consequences follow for the instrument examined here. First, a disclosure field is not a remedy that disabled authors will simply use; the population it purports to serve is the population documented to decline such channels at rates approaching half, for precisely the reason the field's design ignores. Second, the suppression is sharpest among people with cognitive disabilities, the same population whose writing most depends on the composition tools that trigger the badge. The architecture concentrates its demand on the group the literature identifies as least able to answer it.

V. The Marked-Class Trap

A. Why the Obvious Repairs Fail

The instinctive repair is a third category. Add an assistive-technology classification; let verified assisted authors carry an exemption tag instead of a percentage; excuse their posts from scanning. Every version of this repair reproduces the harm it treats, because a marked class is a disclosed class. An assistive-technology badge is a disability announcement with the diagnosis redacted, and the redaction protects nothing that matters, because the injury was never the

32. Christopher J. Moreland et al., *Disability Accommodation Access and Requests in US Internal Medicine Residents with Disabilities*, 9 JAMA Network Open art. e263392 (2026), `https://doi.org/10.1001/jamanetworkopen.2026.3392` (last visited Aug. 14, 2026).

33. Xiji Zhu, Feng Jiang & Sabrina D. Volpone, *The Role of Inclusive Leadership in Reducing Disability Accommodation Request Withholding*, 52 J. Mgmt. 1570 (2025), `https://doi.org/10.1177/01492063251314453` (last visited Aug. 14, 2026).

34. Mairead Moloney, Robyn Lewis Brown & Gabriele Ciciurkaite, *"Going the Extra Mile": Disclosure, Accommodation, and Stigma Management Among Working Women with Disabilities*, 40 Deviant Behav. 942 (2018), `https://doi.org/10.1080/01639625.2018.1445445` (last visited Aug. 14, 2026).

particulars. It was anyone knowing at all. Category disclosure is disclosure. The same failure consumes the visible opt-out: a writer whose posts alone return no scan wears the absence as a badge, and the reader completes the inference the design withheld. Even the European watermarking regime's assistive carve-out, the most disability-aware provision in this field, draws its category line through the middle of the disabled population, exempting communication aids that leave semantics untouched while marking the composition aids on which cognitive and executive-function disabilities rely; a line that decides which disabilities may pass invisibly is still a sorting of disabled people by machine-readable mark.[35]

The test, then, is one sentence. If a reader inspecting the badge, the tag, the empty field, the dispute trail, or the opt-out can infer anything about disability, the design fails. Run the candidate repairs through that gate and nearly all of them die.

B. What Survives

Three designs pass. First, badge deletion with author primacy: no automatic percentage; the author-controlled description is the only provenance surface that exists, so the disabled author's statement is a statement among statements rather than a rebuttal beneath a number. Second, the undifferentiated bucket: if a machine-involvement indicator must exist, a single binary, tools were used, that swallows everything from a spellchecker to a grammar assistant to a full dictation-and-drafting pipeline, with no gradations and no percentages, so the assisted author stands invisibly inside the overwhelming majority of writers who touch some tool.[36] Third, the no-reason dispute path: any correction or removal mechanism open to every author, for any reason, with no reason displayed and no documentation ever collected. What unites the three is that the disabled author disappears into a crowd. Universal design is not one repair among others here; it is the only class of repair that does not itself disclose. The accessibility literature reaches

35. Regulation (EU) 2024/1689, art. 50(2) (exempting systems performing "an assistive function for standard editing"); European Commission, *supra* note 4 (construing the exception to cover, inter alia, content that only transforms authentic human input through assistive technologies allowing persons with disabilities to communicate, while returning summarization and substantive rewriting to the marking duty).

36. The bucket works precisely because it is uninformative. An indicator that nearly every author triggers supports no inference about any author. Herd anonymity is the accessibility property; granularity is the surveillance property.

the same conclusion from the other direction, observing that sole reliance on mandated individual accommodations delays support, stigmatizes those who invoke it, and misses everyone with an undisclosed disability, which is why proactive design that embeds access for all users is treated as the corrective rather than the supplement.[37]

VI. The Doctrine

A. Title III: Screen-Outs and Methods of Administration

Title III of the ADA prohibits public accommodations from imposing "eligibility criteria that screen out or tend to screen out an individual with a disability" from full and equal enjoyment of services, unless necessary,[38] and separately prohibits "utiliz[ing] standards or criteria or methods of administration" that have the effect of discriminating on the basis of disability.[39] A detection-and-disclosure architecture is a method of administration in the provision's plain sense: a standing operational system through which the platform administers the credibility of its authors. Its tendency to screen out is the mechanism Parts III and IV documented: the writers whose prose carries the machine signature of their accommodations are systematically badged, systematically pressed into the priced field, and systematically subjected to the disclosure penalty that follows, while similarly situated nondisabled writers pass or pay in craft talk. Neither provision requires intent; both are effects provisions, and the effect is established by the design's own price structure. The full-and-equal-enjoyment guarantee that both provisions serve[40] is not satisfied by formally identical access to a field whose effective use costs one class its medical privacy. The Supreme Court situates that guarantee inside a statute enacted under a "broad

37. Stephanie A. Gedzyk-Nieman & Anne Derouin, *Ensuring Access for All Nursing Students: Classroom Accommodations and Universal Design for Learning*, 31 Creative Nursing 392 (2025), `https://doi.org/10.1177/10784535251392561` (last visited Aug. 14, 2026).

38. 42 U.S.C. § 12182(b)(2)(A)(i).

39. 42 U.S.C. § 12182(b)(1)(D).

40. 42 U.S.C. § 12182(a).

mandate," whose "comprehensive character" the Court counted among the Act's most impressive strengths.[41]

B. Access, Not Content

Before geography, a distinction the defense will reach for first. The same line of cases that carries this Article's coverage authority also holds that Title III governs access to what a covered entity offers and not the content of the offering. The Seventh Circuit put it as a bookstore that must admit disabled readers without stocking Braille;[42] the Ninth Circuit said the same of an insurance policy's terms.[43] A platform will therefore argue that its detection badge is part of the product it sells readers, and that a claim aimed at the badge asks a court to redesign the offering rather than open its door.

The answer is that this Article's claim does not run through the general rule those cases construe. Both decisions interpret § 12182(a), the general prohibition. The provisions relied on here are the specific prohibitions Congress wrote separately: eligibility criteria that tend to screen out,[44] and standards, criteria, or methods of administration having the effect of discriminating.[45] Those provisions regulate exactly what the content cases carve out from the general rule: how the entity administers its offering, as distinct from what the offering contains. A badge is not an ingredient of the newsletter a reader came to read. It is the operational apparatus through which the platform administers author credibility, and the disclosure field is the procedure it supplies for contesting the apparatus's output.

Two features of this claim keep it on the access side of the line. First, the remedies specified in Part V alter no post's content, no reader's experience of any text, and no editorial

---

41. PGA Tour, Inc. v. Martin, 532 U.S. 661, 675 (2001).

42. Doe v. Mut. of Omaha Ins. Co., 179 F.3d 557, 559, 563 (7th Cir. 1999) (Title III's core meaning is that a facility open to the public may not exclude disabled persons or prevent them from using it as the nondisabled do; § 302(a) "does not require a seller to alter his product to make it equally valuable to the disabled and to the nondisabled").

43. Weyer v. Twentieth Century Fox Film Corp., 198 F.3d 1104, 1115 (9th Cir. 2000) (Title III requires nondiscriminatory enjoyment of the goods a place provides, not provision of different goods).

44. 42 U.S.C. § 12182(b)(2)(A)(i); 28 C.F.R. § 36.301(a).

45. 42 U.S.C. § 12182(b)(1)(D).

judgment about what may be published; they change what the administrative apparatus displays about how a work was made and what an author must surrender to correct it. Second, the injury is not that disabled authors receive a less valuable product. It is that they are charged a different price, in medical privacy, for the same participation, which is a condition on enjoyment rather than a complaint about what is being enjoyed. The First Circuit, which supplies the doctrine's most expansive reading, declined even to foreclose the broader theory, observing that nothing in the legislative history explicitly precludes extending the statute to the substance of what is offered.[46] This Article does not need that reading. It needs only the uncontested proposition that how a public accommodation runs its operation is governed even where what it sells is not.

C. The Circuit Geography

Whether Title III reaches a web-native platform at all depends on where the question is asked. The Ninth Circuit, home to the platform, requires a nexus between the challenged service and a physical place of public accommodation, a requirement a pure web service cannot meet.[47] The circuit's own web-accessibility landmark proves the boundary rather than softening it: Domino's was reachable because its website and app connected customers to physical pizzerias.[48] The First Circuit holds the opposite, and said so as a matter of first impression three decades ago: public accommodations are not "limited to actual physical structures."[49] The Seventh Circuit's formulation points the same direction, describing Title III's reach in terms of the goods and

---

46. Carparts Distrib. Ctr., Inc. v. Auto. Wholesaler's Ass'n of New Eng., Inc., 37 F.3d 12, 20 (1st Cir. 1994).

47. Weyer v. Twentieth Century Fox Film Corp., 198 F.3d 1104, 1114 (9th Cir. 2000) (requiring "some connection between the good or service complained of and an actual physical place"). The holding arose from an employer-provided disability policy administered by an insurer, and the opinion does not address whether a web-native service with no physical premises is covered.

48. Robles v. Domino's Pizza, LLC, 913 F.3d 898, 905 (9th Cir. 2019) (nexus satisfied because the digital channels "facilitate access to the goods and services of a place of public accommodation," the restaurants).

49. Carparts Distrib. Ctr., Inc. v. Auto. Wholesaler's Ass'n of New Eng., Inc., 37 F.3d 12, 19 (1st Cir. 1994) (holding as a matter of first impression that establishments of public accommodation are not limited to actual physical structures, and reasoning that it would be irrational to protect those who enter an office to purchase services while excluding those who purchase the same services by telephone or mail).

services offered rather than the walls around them.[50] The claim described above is therefore live or dead by geography, which is an unstable foundation for a civil right, and which motivates the second doctrinal path.

D. The New York City Human Rights Law

The detection company was founded and operates in Brooklyn.[51] That places it inside the New York City Human Rights Law, which prohibits disability discrimination by providers of public accommodations without the federal statute's physical-place baggage,[52] and which its own text commands be construed "liberally for the accomplishment of the uniquely broad and remedial purposes thereof," independently of how federal and state analogues are read.[53] Disparate-impact liability is available by statute rather than by contested implication. And the city law solves the triad problem that defeats accommodation analysis elsewhere, the vendor who creates the signal while owing the badged person nothing, because it separately prohibits aiding and abetting any practice the chapter forbids, reaching the party that builds and supplies the discriminatory instrument alongside the party that deploys it.[54] The significance is structural: the entity that designs the taxonomy, trains the classifier, and licenses the badge to platforms sits, for its own conduct, in the most plaintiff-protective public-accommodations regime in the country. The architecture examined in this Article was built inside the jurisdiction least tolerant of it.

---

50. Doe v. Mut. of Omaha Ins. Co., 179 F.3d 557, 559 (7th Cir. 1999) (reading Title III's core meaning to reach the operator of a store, theater, "Web site, or other facility (whether in physical space or in electronic space)," and citing Carparts for the proposition).

51. *See* Bellan, *supra* note 3 (describing the "New York-based AI detection startup"); Pangram Labs, company records and profiles consistently locating the firm in Brooklyn, New York.

52. N.Y.C. Admin. Code § 8-107(4)(a).

53. N.Y.C. Admin. Code § 8-130(a).

54. N.Y.C. Admin. Code § 8-107(6) (unlawful "to aid, abet, incite, compel or coerce the doing of any of the acts forbidden under this chapter, or to attempt to do so").

VII. Objections, Run in Both Directions

An argument worth making is worth stress-testing from the defense table. This Part takes the six strongest answers a platform or vendor would give, states each at full strength, and shows where each one lands.

A. The Flood Is Real

The strongest objection concedes everything above and answers: the flood is real. Machine-generated text is drowning authentic writing, some discrimination between the two is necessary for any human signal to survive, and imperfect instruments beat none. The premise is granted. The conclusion does not follow, because it mistakes the layer on which the real problem lives.

Flooding is not an authorship-tool problem. It is an identity and volume problem: one actor wearing a thousand faces, manufactured consensus, engagement farms, coordinated inauthenticity. Substance-checking catches a bad document; nothing at the document layer catches a fake crowd, and nothing at the document layer needs to. The honest countermeasures live at the account layer, where the fraud lives: rate structures, identity assurance, provenance of accounts rather than of sentences, coordinated-behavior detection. An instrument aimed at how a sentence was produced cannot distinguish the sockpuppet's ten-thousandth post from the disabled author's first, because at the sentence layer they are the same event, a model produced text. The puppeteer, who can spin a new account faster than any badge attaches, routes around the instrument; the assisted author, who cannot opt out of the tools that make her writing possible, absorbs it. A detection regime at the wrong layer is not a partial solution to flooding. It is a tax on the population that cannot dodge it, collected in service of a problem it does not solve.

B. The Author Agreed to the Terms

The platform will answer that every author accepted terms of service contemplating platform features, detection included, and that participation constitutes consent to the architecture.

Two replies. First, the civil rights obligations at issue do not travel through contract: a public accommodation cannot condition service on a waiver of the very statutory protections that regulate how it serves, and an adhesive click at signup, before the instrument existed, is not an election about an instrument shipped years later. Second, and empirically, voluntariness does not launder the penalty even on the platform's own framing. The trust literature tested the distinction directly: the disclosure penalty holds "regardless of whether disclosure is voluntary or mandatory."[55] A consent theory answers a coercion objection. The objection here is not coercion. It is price.

C. Section 230

The platform's reflexive shield fails at the threshold, and the failure is worth stating because it clears the field of the defense most often mistaken for dispositive. Section 230 protects a provider from being "treated as the publisher or speaker of any information provided by another information content provider."[56] The badge is not another's information. It is the platform's and the vendor's own output: a percentage they compute, render, and attach. An entity is unprotected as to content it creates or develops itself,[57] and the claim described in this Article does not seek to hold anyone liable for an author's post. It targets the platform's own instrument and the vendor's own classifier. The statute that immunizes intermediaries for hosting speech says nothing about instruments they build to grade it.

D. The First Amendment

The serious defense is constitutional, and candor requires ranking it first among the unresolved. The Supreme Court has recognized that a platform's compilation and curation of third-party

---

55. Schilke & Reimann, *supra* note 19.

56. 47 U.S.C. § 230(c)(1).

57. Fair Hous. Council of San Fernando Valley v. Roommates.com, LLC, 521 F.3d 1157, 1162, 1167–68 (9th Cir. 2008) (en banc) (as to content a website creates itself, or is "responsible, in whole or in part" for creating or developing, it is a content provider outside § 230's shield; development means "materially contributing to [the] alleged unlawfulness"). The distinction the en banc court drew is instructive here: the operator was immune for the open text box its users filled in, and unprotected as to the questions it composed and required users to answer as a condition of service. A percentage the platform computes and displays is the second kind of content, not the first.

speech is itself expressive activity,[58] and a platform will characterize the badge as its own protected commentary on the works it hosts, with any remedy an unconstitutional compulsion to alter its editorial voice. Three answers, offered with their limits acknowledged. First, the claim regulates an operational screening architecture's discriminatory effects on access to a commercial service, the domain anti-discrimination law has always governed even where the regulated business trades in expression; the theory does not dictate what the platform may say about any post, and the surviving remedies in Part V are a menu of the platform's own choosing, none of which prescribes a message. Second, the vendor's posture differs from the platform's: a classifier sold as infrastructure, marketed on accuracy rather than viewpoint, claims the protections of a measuring instrument, and a measuring instrument's discriminatory deployment is conduct before it is commentary. Third, to the extent the badge is speech, the disclosure field's compelled-answer dynamic runs the compelled-speech concern in the other direction, against the author, whose effective participation now requires publishing a diagnosis. The Article does not pretend this terrain is settled. It observes that the constitutional question is at least double-edged, and that the universal-design remedies were specified precisely so that relief exists which no reading of the expressive-curation cases forbids.

E. Necessity

The screen-out provision excuses eligibility criteria shown to be "necessary" for the provision of the service.[59] Necessity is where the flood objection returns in doctrinal dress, and where it fails twice. It fails on ends, because the interest asserted, defense against inauthentic volume, lives at the account layer, and a document-layer classifier does not achieve it, as Part VII.A showed. And it fails on means, because necessity cannot survive the existence of designs that serve every legitimate interest without the discriminatory surface: the author-primacy design, the undifferentiated bucket, and the no-reason dispute path of Part V each preserve whatever

58. Moody v. NetChoice, LLC, 603 U.S. 707 (2024).

59. 42 U.S.C. § 12182(b)(2)(A)(i); see also 28 C.F.R. § 36.301(a).

reader-information function the badge claims while extracting no diagnosis from anyone. A criterion is not necessary when its accessible replacement is sitting in the same paragraph.

F. Show Me the Numbers

The soberest objection is evidentiary: disparate-impact claims conventionally ride on statistics, and this Article offers mechanism, literature, and a documented individual instance rather than a platform-scale distribution of badge outcomes by disability status. Three responses, in candor. First, the distribution exists and is held entirely by the platform and the vendor; the asymmetry of access to proof is itself a feature of the architecture, and pleading standards do not require a plaintiff to possess the defendant's telemetry to state a plausible effects claim built on documented mechanism. Second, the vendor's own marketing concedes the capability at the heart of the showing: a model advertised as over 99 percent accurate at finding AI-assisted and mixed text is a model that, by its maker's description, finds the assisted class.[60] Third, in the jurisdiction identified in Part VI.C, the construction command runs independently and liberally,[61] and the local law's effects analysis was built for exactly the case where the defendant's instrument, not the plaintiff's data, is the crux. The gap is real; it is a discovery gap, not a merits gap, and it is one more reason the record-keeping obligations proposed below belong in any repair.

G. Just Publish Somewhere Else

Finally, the market answer: the aggrieved author can leave. Full and equal enjoyment has never meant enjoyment somewhere else; a segregated exit is the injury restated, not a remedy for it. And the exit leads nowhere, because the architecture is converging. The invisible watermark ships in the model layer worldwide. The crowdsourced flag ships on the professional network. The detector badge ships on the newsletter platform. An author who writes through assistive technology and leaves the badge finds the flag, and behind the flag, the mark. The instruments

60. Bellan, *supra* note 3.

61. N.Y.C. Admin. Code § 8-130(a).

differ; the taxonomy they enforce, human or machine with nothing in between, is the same, and it is the taxonomy, not any single product, that this Article has been about.

VIII. Conclusion

The executive's sentence deserves to be quoted one last time, in full: "If we do not actively discriminate in favor of human content, then we're just gonna see more and more AI, and it's just gonna drown out any human signal that we have." The sentence assumes the human signal is a property of text, measurable in the rhythm of sentences, and that protecting humanity means protecting the texts that carry the right fingerprint. Every page above has pressed on the assumption's excluded middle: the human being whose signal reaches the page through a machine, because the machine is how that human writes. Her thoughts are not less hers for the carrying. A regime that reads the carrier and badges the person has not detected artificiality. It has defined humanity as the unassisted, published the definition beside her name, and offered her, as remedy, a field in which to prove her personhood with a diagnosis.

The doctrine already knows better. It knows that the diagnosis is not the price of admission, that the two questions at the door are the most a gatekeeper may ask, that effects discriminate without anyone intending them, and, in at least one city, the city where this instrument was built, that public accommodation follows the service and not the walls. The design principle follows from the doctrine and fits in a sentence: no reader may be able to infer disability from any surface of a provenance system, and the only architectures that satisfy the sentence are the ones in which assisted authors vanish into the crowd of everyone. The human signal was never in the fingerprint. It is the person. Any instrument that cannot find her there is not detecting the machine. It is discriminating in favor of a narrower humanity, exactly as advertised.

References

Americans with Disabilities Act of 1990, 42 U.S.C. §§ 12101–12213.

Bellan, R. (2026). As AI content floods the internet, Pangram raises $9M to detect it. *TechCrunch*. https://techcrunch.com/2026/07/29/as-ai-content-floods-the-internet-pangram-raises-9m-to-detect-it/.

Bonifacic, I. (2026). Substack is adding an AI detection feature. *Engadget*. https://www.engadget.com/2220064/substack-is-adding-an-ai-detection-feature/.

Carparts Distribution Center, Inc. v. Automotive Wholesaler's Association of New England, Inc., 37 F.3d 12 (1st Cir. 1994).

Coley, M. (2023). Guidance on AI detection and why we're disabling Turnitin's AI detector. *Vanderbilt University Brightspace*. https://www.vanderbilt.edu/brightspace/2023/08/16/guidance-on-ai-detection-and-why-were-disabling-turnitins-ai-detector/.

Doe v. Mutual of Omaha Insurance Co., 179 F.3d 557 (7th Cir. 1999).

Fair Housing Council of San Fernando Valley v. Roommates.com, LLC, 521 F.3d 1157 (9th Cir. 2008) (en banc).

Gedzyk-Nieman, S. A., & Derouin, A. (2025). Ensuring access for all nursing students: Classroom accommodations and universal design for learning. *Creative Nursing*, 31(4), 392–398. https://doi.org/10.1177/10784535251392561.

Kendro, K., Maloney, J., & Jarvis, S. (2026). Do large language models produce texts with "human-like" lexical diversity? Evidence from four ChatGPT models. *International Journal of Applied Linguistics*. https://doi.org/10.1111/ijal.70115.

Lawrence, K. W., Habibi, A. A., Ward, S. A., Lajam, C. M., Schwarzkopf, R., & Rozell, J. C. (2024). Human versus artificial intelligence-generated arthroplasty literature: A single-blinded analysis of perceived communication, quality, and authorship source. *The International Journal of Medical Robotics and Computer Assisted Surgery*, 20(1), e2621. https://doi.org/10.1002/rcs.2621.

Elkhatat, A. M., Elsaid, K., & Almeer, S. (2023). Evaluating the efficacy of AI content detection tools in differentiating between human and AI-generated text. *International Journal for Educational Integrity*, 19, art. 17. https://doi.org/10.1007/s40979-023-00140-5.

European Commission. (2026). Guidelines on the implementation of the transparency obligations for certain AI systems under Article 50 of the AI Act. https://digital-strategy.ec.europa.eu/en/library/guidelines-transparency-obligations-providers-and-deployers-ai-systems.

Liang, W., Yuksekgonul, M., Mao, Y., Wu, E., & Zou, J. (2023). GPT detectors are biased against non-native English writers. *Patterns*, 4(7), 100779. https://doi.org/10.1016/j.patter.2023.100779.

Liçenji, L., & Hoxha, J. (2026). When news is "written by artificial intelligence": A systematic review of provenance and disclosure cues in journalism and their effects on credibility and trust. *Frontiers in Artificial Intelligence*, 9, art. 1815243. https://doi.org/10.3389/frai.2026.1815243.

Moloney, M., Brown, R. L., & Ciciurkaite, G. (2018). "Going the extra mile": Disclosure, accommodation, and stigma management among working women with disabilities. *Deviant Behavior*, 40(8), 942–956. https://doi.org/10.1080/01639625.2018.1445445.

Moody v. NetChoice, LLC, 603 U.S. 707 (2024).

Moreland, C. J., Pereira-Lima, K., Lee, K. T., et al. (2026). Disability accommodation access and requests in US internal medicine residents with disabilities. *JAMA Network Open*, 9(3), e263392. https://doi.org/10.1001/jamanetworkopen.2026.3392.

Nakano, H., Takezawa, J., & Matulic, F. (2025). Understanding reader perception shifts upon disclosure of AI authorship. *arXiv*. https://doi.org/10.48550/arxiv.2510.24011.

Mayor, E., Bietti, L. M., & Bangerter, A. (2025). Can large language models simulate spoken human conversations? *Cognitive Science*, 49(9), e70106. https://doi.org/10.1111/cogs.70106.

Mingus, M. (2017). Forced intimacy: An ableist norm. *Leaving Evidence*. https://leavingevidence.wordpress.com/2017/08/06/forced-intimacy-an-ableist-norm/.

New York City Human Rights Law, N.Y.C. Admin. Code §§ 8-101 to 8-131.

Pangram Labs. (2025). AI transparency company Pangram integrated into two leading news source and resource platforms [Press release]. *Business Wire*. https://www.businesswire.com/news/home/20250915315368/en.

Pereira-Lima, K., Meeks, L. M., Ross, K. E. T., et al. (2023). Barriers to disclosure of disability and request for accommodations among first-year resident physicians in the US. *JAMA Network Open*, 6(5), e239981. https://doi.org/10.1001/jamanetworkopen.2023.9981.

PGA Tour, Inc. v. Martin, 532 U.S. 661 (2001).

Raj, M., Berg, J. M., & Seamans, R. (2026). The artificial intelligence disclosure penalty: Humans persistently devalue AI-generated creative writing. *Journal of Experimental Psychology: General*, 155(4), 896–915. https://doi.org/10.1037/xge0001889.

Robles v. Domino's Pizza, LLC, 913 F.3d 898 (9th Cir. 2019).

Schilke, O., & Reimann, M. (2025). The transparency dilemma: How AI disclosure erodes trust. *Organizational Behavior and Human Decision Processes*, 188, 104405. https://doi.org/10.1016/j.obhdp.2025.104405.

Perez, S. (2026). Substack's new tool tells you who's been writing their newsletters with AI. *TechCrunch*. https://techcrunch.com/2026/07/22/substacks-new-tool-tells-you-whos-been-writing-their-newsletters-with-ai/.

Regulation (EU) 2024/1689 (Artificial Intelligence Act), art. 50(2), 2024 O.J. (L 1689).

Substack. (2026). Announcement of Pangram AI detection integration. *X*. https://x.com/Substack/status/2079598704424779787.

Communications Decency Act § 230, 47 U.S.C. § 230.

Ullah, U., Laudanna, S., Di Sorbo, A., Visaggio, C. A., & Murray, R. (2026). Breaking the imitation game: Can LLMs fool humans and machines alike? *International Journal of Intelligent Systems*, 2026(1), 8117975. https://doi.org/10.1155/int/8117975.

Wang, N., & Liang, C. (2026). How to disclose? Strategic AI disclosure in crowdfunding. *arXiv*. https://doi.org/10.48550/arxiv.2602.15698.

U.S. Department of Justice, Nondiscrimination on the Basis of Disability by Public Accommodations, 28 C.F.R. pt. 36.

Weber-Wulff, D., Anohina-Naumeca, A., Bjelobaba, S., Foltýnek, T., Guerrero-Dib, J., Popoola, O., Šigut, P., & Waddington, L. (2023). Testing of detection tools for AI-generated text. *International Journal for Educational Integrity*, 19, art. 26. https://doi.org/10.1007/s40979-023-00146-z.

Weyer v. Twentieth Century Fox Film Corp., 198 F.3d 1104 (9th Cir. 2000).

Zhu, X., Jiang, F., & Volpone, S. D. (2025). The role of inclusive leadership in reducing disability accommodation request withholding. *Journal of Management*, 52(4), 1570–1601. https://doi.org/10.1177/01492063251314453.