

The Assistive Exemption:

Content Provenance Marking and the Compelled Disclosure of Disability

Travis Gilly*  August 2026 (v0.8)

Abstract

Article 50(2) of the EU Artificial Intelligence Act requires providers of generative systems to mark outputs in a machine-readable format so that they are detectable as artificially generated. The obligation carries an express carve-out: it does not apply where the system performs an assistive function for standard editing. The Commission's July 2026 Guidelines define that exception to cover grammar correction, spellchecking, minor stylistic polishing, machine translation, and, expressly, content that only transforms authentic human input through assistive technologies allowing persons with disabilities to communicate. They also confirm that the exception applies per operation rather than per system.

So the drafters did consider disabled users. They considered one kind. The carve-out reaches assistive technology that lets a person communicate without altering semantics; it does not reach assistive technology that helps a person compose or organize, and the same Guidelines place summarization and substantive rewriting back inside the marking duty. The line runs through the middle of the disabled population.

The first implementations went further still. In August 2026 Anthropic announced that its models mark generated text worldwide, and its own documentation concedes that a detected mark shows only that text may have been processed by the model, naming proofreading, translation, and summarization as uses it cannot distinguish. Two of those three are inside the Commission's exception. The mark attaches anyway, applies wherever the product is offered rather than to output used in the Union, and reaches personal use the Act excludes from scope.

The Code of Practice then decides who reads it. It concedes in operative text that watermark detection for free-form text has lower reliability and may produce misleading or low-confidence results, and it nonetheless guarantees access to a named class of expert users that includes educational and research organisations. It requires that a detection result be downloadable as a digitally signed artifact with a content hash and timestamp. It contains no provision entitling the person whose text was queried to notice, to the receipt, or to any route to contest the result.

This Article makes three claims. First, that accessibility enters this regime in one direction only: it constrains how a notice or a detection result is delivered to a reader, never who is marked, and the Guidelines state that Article 50 imposes no distinct accessibility requirements of its own. Second, that the resulting position is a compelled disclosure. The mark says nothing about disability; it says machine involvement, and where an institution restricts that involvement the only defense available to an accommodated author is the accommodation. A facially neutral rule answerable only by disclosing protected status is a disability related inquiry in substance,

*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.

and existing doctrine already says so. Third, that no doctrine reaches the injury when it occurs. The provider that creates the signal owes the marked person nothing; the institution that acts on it holds a guaranteed right of access and did not create it. Accommodation law is built for a dyad. Provenance marking makes a triad whose third element is a property of the artifact.

Keywords: EU AI Act; Article 50; content provenance; watermarking; assistive technology; disability disclosure; compelled disclosure; meaningful access; differential burden; academic integrity; AI detection; Section 504.

Contents

I	Introduction	4
II	The Marking Regime and What Was Built	6
	A Article 50(2) and Its Limits	6
	B What the Act Does Not Reach	7
	C What Anthropic Built.	8
	D Three Dimensions of Excess.	9
III	The Accessibility Asymmetry	10
IV	Who Gets to Read the Mark	11
V	Detection as Practiced	13
	A What Detectors Do to Whom	13
	B The Folk Taxonomy.	15
	C Why Marking Does Not Fix This, and Makes One Thing Worse	17
	D Who the Regime Says Will Read It	19
VI	The Disclosure Trap	20
	A The Unmarked Baseline	20
	B Pre-emption of a Disclosure Already Made	21
	C The Election	23
	D Protected Status as the Only Available Defense	24
	E The Differential Burden	25
	F Why Advance Registration Does Not Answer It.	26
VII	Where Existing Doctrine Runs Out	27
	A Against the Provider	27
	B Against the Deployer	29
	C The Gap	30
VIII	Repairs	31
IX	Conclusion	32

I. Introduction

A person with a disability writes a paragraph. The writing happens with help, because unaided it does not happen at all, or does not happen in a form anyone will read. The help is software. The paragraph is published.

Inside the paragraph, invisible to every human reader, is a statistical signature keyed to the software that helped produce it. The author did not put it there, cannot see it, cannot remove it, and was not asked. A defined class of institutions can run a check against it, at any time, for as long as the text exists. The author will not be told that a check occurred.

This Article is about that signature, about the sentence in European law that was meant to keep it off this paragraph, and about the two implementing documents that decide who gets to read it.

Article 50(2) of the Artificial Intelligence Act obliges providers of generative systems to mark their outputs in a machine-readable format and to make those outputs detectable as artificially generated or manipulated.¹ The obligation applies from 2 August 2026, with systems already on the market required to comply by 2 December 2026.² The same paragraph that creates the duty limits it. The obligation, the Act says, "shall not apply to the extent the AI systems perform an assistive function for standard editing or do not substantially alter the input data provided by the deployer or the semantics thereof."³

That sentence is the assistive exemption. The Commission has now told us what it means. Standard editing is "the process of preparing existing content for publication or distribution (e.g., small edits to improve readability and grammar, quality and format)," and it "does not involve

1. Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence, art. 50(2), 2024 O.J. (L 1689) [hereinafter AI Act].

2. European Commission, *Transparency Obligations Under Article 50 of the AI Act* (FAQ, last updated July 24, 2026), <https://digital-strategy.ec.europa.eu/en/faqs/transparency-obligations-under-article-50-ai-act> [hereinafter Commission FAQ].

3. AI Act, *supra* note 1, art. 50(2).

generating new content."⁴ The Guidelines list what benefits from the exception. Grammar correction and spellchecking. Minor stylistic polishing that does not change substance, meaning, style, or messaging. Machine translation of text. And, expressly, content "that only transforms authentic human input through assistive technologies allowing persons with disabilities to communicate."⁵

So the Commission did think about disabled users. It thought about one kind. The carve-out it wrote covers assistive technology that lets a person with a disability communicate without altering the semantics of what they meant. It does not address assistive technology that helps a person with a disability compose, organize, or complete what they meant, and the Guidelines place summarization and substantive rewriting squarely inside the marking obligation.⁶ The line the Commission drew runs through the middle of the disabled population, and nothing in the record suggests anyone noticed where it landed.

This Article argues three things.

First, that the assistive exemption is the most consequential provision in Article 50(2) and that the first implementations read it more narrowly than the Guidelines require, marking uniformly across the exempt and non-exempt categories and then conceding that the resulting signal cannot tell them apart. Part II sets out the regime and what one major provider built, and identifies three dimensions along which the implementation exceeds the mandate.

Second, that accessibility enters this regime in exactly one direction. Part III shows that in the Act, the Guidelines, and the Code of Practice alike, accessibility is a property of information delivered to the person receiving content or reading a detection result. It is never a constraint on who is marked or a right belonging to the person marked. The Code goes further: it names the

4. European Commission, *Guidelines on the Implementation of the Transparency Obligations for Certain AI Systems Under Article 50 of the AI Act* para. 90 (July 20, 2026), <https://digital-strategy.ec.europa.eu/en/library/guidelines-transparency-obligations-providers-and-deployers-ai-systems> [hereinafter Guidelines].

5. *Id.* (examples following para. 92), naming augmentative and alternative communication and customized neural voices.

6. *Id.* (examples of semantic changes requiring marking, listing AI-generated summaries and paraphrasing or rewriting beyond minor stylistic correction).

classes of expert users who may query text watermark detection, and educational and research organisations are on the list, while the person whose text is queried appears nowhere in the document.

Third, that the resulting injury is real, patterned, and orphaned. Parts IV and V establish what detection does in practice and why the mark converts publication into a compelled disclosure for one class of authors. Part VI reports honestly where existing doctrine reaches and where it stops: weakly against the provider, considerably better against the institutions that act on the signal, and nowhere at the moment the injury occurs. Part VII proposes four repairs available without new legislation and one that is not.

II. The Marking Regime and What Was Built

A. Article 50(2) and Its Limits

Article 50 imposes four duties on four facts. Providers of systems that interact directly with people must ensure people know they are interacting with a machine.⁷ Providers of generative systems must mark outputs so they are machine detectable.⁸ Deployers of emotion recognition and biometric categorisation systems must inform those exposed.⁹ Deployers must label deepfakes, and must label generated text published to inform the public on matters of public interest.¹⁰

Only the second attaches to the artifact rather than to a disclosure event. The others require someone to say something. Article 50(2) requires something to be true of a file.

The provision reaches providers wherever established, including outside the Union, "if the output of their AI system is used in the EU."¹¹ Recital 133 offers a menu of techniques: watermarks, metadata identifications, cryptographic provenance methods, logging, fingerprints,

7. AI Act, *supra* note 1, art. 50(1).

8. *Id.* art. 50(2).

9. *Id.* art. 50(3).

10. *Id.* art. 50(4).

11. Commission FAQ, *supra* note 2; *see* AI Act, *supra* note 1, art. 3(3) (defining provider).

or other appropriate techniques.¹² The Code of Practice on Transparency of AI-generated Content, published 10 June 2026 and assessed as adequate by the Commission and the AI Board, gives signatories a route to demonstrating compliance regardless of place of establishment.¹³

Four categories of output fall outside the duty entirely: short sequences of numbers, symbols, or letters; source code; machine to machine communication never exposed to humans; and closed loop industrial and product development environments.¹⁴ Then the assistive exemption, which is not a category of output at all but a description of what the system did.

Two features of the drafting matter and the Guidelines confirm both.

The exemption is a gradient rather than a switch. It applies "to the extent" the system performs the exempt function, and the Commission says plainly that where a system can be used both for generation and for non-substantial minor alteration, "the transparency obligation under Article 50(2) AI Act will not apply to the content altered in a minor non-substantial manner."¹⁵ That is express authority for marking that varies by operation.

And the test is functional. It asks what the system performed, not what the user is and not what the user said. Nothing requires the user to claim the exemption and nothing conditions it on a declaration.

B. What the Act Does Not Reach

Two further boundaries matter because implementations have crossed them.

The Act excludes personal, non-professional activity from the deployer obligations altogether. A natural person generating content in a personal capacity is outside scope; the same person becomes a deployer only where the activity yields regular economic benefit or is

12. AI Act, *supra* note 1, recital 133.

13. European Commission, *Code of Practice on Transparency of AI-Generated Content* (June 10, 2026), <https://digital-strategy.ec.europa.eu/en/policies/code-practice-ai-generated-content> [hereinafter Code of Practice]; Commission FAQ, *supra* note 2. Roughly 190 organisations had signed by the end of July 2026.

14. Commission FAQ, *supra* note 2.

15. Guidelines, *supra* note 4, para. 92.

otherwise part of a business, trade, occupational, or freelance activity.¹⁶ The line is drawn at commerce.

And the labelling duty for published text on matters of public interest disappears where the text has undergone human review or editorial control. The Commission defines human review as deliberate examination of the substance by a person with relevant knowledge and professional judgment, and names academic peer review as an example. Superficial, formal, or procedural checks, spell checking and grammatical correction, do not count.¹⁷

That definition is doing something the architecture may not have noticed. Article 50(4) treats spell checking and grammar correction as too thin to discharge a human's duty. Article 50(2), as construed by the Guidelines, treats the same operations as too thin to trigger a machine's duty. The Act is internally consistent that proofreading is not authorship.

C. What Anthropic Built

On 10 August 2026 Anthropic published support documentation describing how its Claude models mark generated content.¹⁸ The documentation states that a watermark is woven into generated text, that it survives copying and pasting, and that models launched on or after 2 August 2026 carry marking at launch, with earlier models covered by the transition period.¹⁹

The notice does not describe the mechanism. Public accounts have supplied one, drawing on the published literature on distribution biasing watermarks, in which a secret key partitions the vocabulary at each generation step and preferentially samples from the favored partition, leaving

16. Commission FAQ, *supra* note 2.

17. *Id.*

18. Anthropic, *How Claude Marks AI-Generated Content* (Claude Support, Aug. 2026), <https://support.claude.com/en/articles/16266773-how-claude-marks-ai-generated-content> [hereinafter Anthropic Marking Notice]. The notice states that Anthropic is a signatory to the Code of Practice as a provider of both models and systems. Contemporaneous reporting places the announcement on Monday, 10 August 2026, and describes coverage across the Claude Platform API, claude.ai, Claude Code, Claude Cowork, Claude Tag, and Claude models accessed through AWS, Google Cloud, and Microsoft Foundry, with digitally signed C2PA provenance metadata attached to supported image and vector files. *See, e.g., Anthropic Plans to Add an Invisible Mark to AI Text*, Fortune (Aug. 11, 2026); *Claude Now Watermarks Text Everywhere*, TechTimes (Aug. 11, 2026).

19. *Id.*

a statistical excess detectable by anyone holding the key.²⁰ That account is consistent with the notice and with the published state of the art. It is inference, and this Article does not rely on the mechanism being any particular one. The legally operative facts are that a mark exists, that it is machine detectable, that it is not human perceptible, and that the user cannot remove it.

Three features of the notice are load bearing.

First, the notice concedes the limits of what detection proves. A detected mark indicates that the text "may have been processed by Claude." The provider states that the model may not be the original author, and gives as examples users who employ the system to proofread, translate, and summarize their own writing.²¹ Two of those three are inside the Commission's exempt list; the third is expressly outside it.²² The provider has told us its signal cannot separate the exempt case from the non-exempt case.

Second, the notice states that marking applies wherever the product is offered, worldwide.²³

Third, the notice offers no opt out, no assistive mode, no regional election, and no mechanism by which a marked person learns that a check was run against their text.²⁴

D. Three Dimensions of Excess

Set the notice against the mandate as the Commission has construed it.

Subject matter. The Act exempts assistive editing and the Guidelines define it to include grammar correction, spellchecking, minor stylistic polishing, and translation. The provider's own examples of what a detected mark cannot rule out include two of those. The exemption is drafted as a gradient and the Commission has expressly endorsed per-operation application.

20. See John Kirchenbauer et al., *A Watermark for Large Language Models*, in Proceedings of the 40th International Conference on Machine Learning (2023); Sumanth Dathathri et al., *Scalable Watermarking for Identifying Large Language Model Outputs*, 634 Nature 818 (2024) (describing the tournament sampling approach deployed in Google's SynthID-Text).

21. Anthropic Marking Notice, *supra* note 17.

22. Guidelines, *supra* note 4 (listing grammar correction and AI-generated translations of text as exempt; listing AI-generated summaries of text as requiring marking).

23. Anthropic Marking Notice, *supra* note 17.

24. *Id.* The notice refers to forthcoming documentation for the detection service.

The implementation does not vary. It marks the exempt operation and the non-exempt operation identically, and then discloses that it cannot tell them apart.

Territory. The Act reaches providers outside the Union where the output is used in the Union. The implementation reaches every user everywhere. A subscriber in Illinois whose output never touches a European server, reader, or market receives a mark whose legal justification is an instrument that does not reach him. Whatever the operational reasons for uniformity, extending the mandate beyond its scope is the provider's choice, and the resulting burden should be assessed as a private product decision rather than as regulatory compliance.

Personal use. The Act excludes personal, non-professional activity. The implementation does not distinguish a court filing from a birthday note.

None of this is unlawful under the AI Act, which sets a floor and not a ceiling. A provider may mark more than it must. The point is analytic: the excess is a choice, it is defended in public as compliance, and the burden it creates does not fall evenly.

III. The Accessibility Asymmetry

There are accessibility provisions in this regime. Every one of them runs the same direction.

Article 50(5) requires that information provided under Article 50(1) to (4) reach natural persons clearly and distinguishably and conform to applicable accessibility requirements.²⁵

The Guidelines devote a section to it, direct providers and deployers to the Web Accessibility Directive and the European Accessibility Act, and then add a sentence that settles the question: "Article 50 AI Act does not impose distinct or additional accessibility requirements."²⁶

Accessibility is borrowed from elsewhere and applied to the delivery of a notice.

The Code of Practice carries the same orientation into detection. Measure 2.3 requires signatories to present detection results clearly and comprehensibly "to natural persons exposed to the content generated or manipulated by their AI system and who want to verify its origin," and

25. AI Act, *supra* note 1, art. 50(5); Guidelines, *supra* note 4, para. 141.

26. Guidelines, *supra* note 4, para. 144, citing Directive (EU) 2016/2102 and Directive (EU) 2019/882.

to ensure that the interfaces presenting those results are accessible to persons with disabilities under the European Accessibility Act and the Web Accessibility Directive, with reference to ETSI EN 301 549 and WCAG 2.1 Level AA.²⁷

Read that carefully. The accessibility obligation attaches to the screen on which a checker reads a result about someone else's writing. The regime has thought hard about a blind person running a detection query. It has not thought at all about a disabled person whose text is the object of one.

The Guidelines contain the sentence that makes this an asymmetry rather than an oversight. The marking and detection obligations, the Commission says, focus "on how the content has been created and its artificial origin, not on who created the content," and for the purpose of marking and detection "the information related to the creator of the content should not be processed."²⁸ That is a data protection instruction and it is correct on its own terms. It is also a description of a regime that has defined the author out of its own frame. Marking is about the how; the author is a who; therefore the author is not the regime's concern.

The difficulty is that in deployment the how becomes a proxy for the who. A signal that says this text was machine assisted, arriving at an institution with a rule against machine assistance, is answered only by the author explaining herself. The Commission's instruction not to process creator information does not prevent that. It only means the regime never planned for it.

IV. Who Gets to Read the Mark

The Code of Practice answers a question the support documentation does not, and the answer reorganizes the analysis.

27. Code of Practice, *supra* note 12, Measure 2.3.

28. *Id.* para. 94.

Signatories must make detection available to the audience that may be exposed to the content, and to the general public where the public may be exposed.²⁹ Then comes the carve-out for text. Signatories "may restrict access to detection mechanisms associated to watermarking techniques for free-form text to the extent that they have a lower level of reliability and robustness and that they may produce misleading or low-confidence results."³⁰

That is the drafters of the compliance instrument saying, in the operative text, that text watermark detection is less reliable and may mislead.

The sentence that follows is the one that matters here. Notwithstanding that unreliability, the results "may still provide valuable information to verified expert end-users with a legitimate need," and the Code names them: competent market surveillance authorities, other regulators, law enforcement authorities, media, fact-checkers, trusted flaggers, independent researchers, educational and research organisations, and civil society organisations. Signatories "will ensure that access for these expert end-users is granted subject to appropriate access controls and safeguards."³¹

Educational and research organisations are on the list. Universities are the single largest deployer of AI detection in the world, they have an existing enforcement apparatus pointed at exactly this question, and the Code guarantees them access to a signal it has just described as potentially misleading.

The Code also builds a receipt. Signatories must ensure that a detection result "can be downloaded upon request in a digitally signed format, including at least a hash of the content submitted for detection, a URL or other identifier of the detection solution, and a timestamp."³² A checker can obtain a cryptographically signed, timestamped artifact recording that a particular text was queried and what came back.

29. Code of Practice, *supra* note 12, Sub-measure 2.1.2.

30. *Id.*

31. *Id.*

32. *Id.*

Now search the Code for the person who wrote the text. There is no provision requiring that she be told a check occurred. None entitling her to the signed receipt, or to any receipt. None establishing a route to contest a result. None acknowledging that a false or misleading result has a subject who bears its consequences. The document runs thirty-eight pages and the marked person is not in it.³³

The regime therefore has this shape. A signal of acknowledged low reliability is embedded in text without the writer's consent. Access to read it is guaranteed to a named class that includes her university. A signed evidentiary artifact is generated for the reader. The writer is told nothing, at any stage, ever.

V. Detection as Practiced

The marking regime arrives into an enforcement infrastructure already built, already deployed at scale, and already documented as unreliable in a patterned way.

A. *What Detectors Do to Whom*

The empirical record on classifier based detection is unusually consistent for a field this young. A Stanford study tested seven widely used detectors against authentic essays and reported an average false positive rate of 61.3 percent on writing by non-native English speakers, against near zero for native speaker writing, with 19.8 percent of the non-native corpus unanimously misclassified by all seven tools.³⁴ A systematic evaluation of fourteen detectors found accuracy varying widely and neither accurate nor reliable performance across text conditions.³⁵ Other work found detectors performing poorly on mixed human and machine text, which is the most

33. This is stated after reading the Code end to end and searching its full text. The Code's accessibility provisions, Measures 2.3 and the deployer-side labelling measures, all run to persons exposed to content or reading a detection result.

34. Weixin Liang et al., *GPT Detectors Are Biased Against Non-Native English Writers*, 4 Patterns 100779 (2023).

35. Debora Weber-Wulff et al., *Testing of Detection Tools for AI-Generated Text*, 19 Int'l J. for Educ. Integrity art. 26 (2023).

common real world case.³⁶ Later work has narrowed but not closed the gap, and at least one study reports a commercial detector holding a one percent false positive rate on a non-native corpus.³⁷

The mechanism is understood. Detectors keyed to perplexity treat predictable word choice as machine signature. Writers who use more common vocabulary and simpler syntax produce lower perplexity prose. Second language writers do this. So do writers whose disabilities affect language production, working memory, or executive function, and so do writers using the tools that help them.

The literature has begun to say so. A 2024 corpus study of popular detection tools, reporting false positive rates well above the advertised figures, concluded from that study together with news reports and first person accounts that neurodivergent writers, alongside second language writers, are more likely to be caught by false positives, and argued that deployment of these tools is on that basis unethical.³⁸ That remains the clearest statement in the literature of the population this Article is about, and the evidentiary base under it is thinner than the claim deserves.

Institutions have acted on less. Vanderbilt University disabled Turnitin's AI detector in August 2023, reasoning in its own announcement that the vendor's claimed one percent false positive rate, applied to the 75,000 papers the university submitted in 2022, would have meant roughly 750 papers incorrectly labelled; the announcement also cited reported false accusations at other institutions, the documented bias against non-native English writers, and the vendor's refusal to explain how the tool reaches its determination.³⁹ Recent work analyzing thousands

36. Ahmed M. Elkhataat et al., *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).

37. Karim Hesham Shaker Ibrahim et al., *The Robustness of AI-Classifiers in the Face of AI-Assisted Plagiarism*, 15 Int'l J. Computer-Assisted Language Learning & Teaching 1 (2025). The disagreement with Liang is genuine and is reported here rather than resolved.

38. Whitney Gegg-Harrison & Claire Quartermann, *AI Detection's High False Positive Rates and the Psychological and Material Impacts on Students*, in *Advances in Educational Marketing, Administration, and Leadership* 199 (IGI Global 2024).

39. 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/>.

of posts across education communities found detector related discourse carrying significantly elevated negative emotion, largest among students, and characterized detectors as governance mechanisms imposing asymmetric burdens on those subject to enforcement rather than merely as inaccurate instruments.⁴⁰

B. The Folk Taxonomy

Underneath the detector market is a fact the market exists to conceal. People cannot do this.

The experimental record is consistent and it runs the wrong way for anyone who believes a trained reader can tell. A 2026 study of readers evaluating short fiction reported identification accuracy of 39.93 percent in one experiment, which is below chance, and 51.97 percent in another, which is indistinguishable from it; the same study found readers rating the machine written stories more highly than the human written ones.⁴¹ Accuracy degrades as models improve rather than improving with familiarity: evaluators distinguished an earlier generation model from human writing at 57.9 percent and a later one at 49.9 percent.⁴² And experience does not help. One study of teachers reported 45.1 percent accuracy among pre-service teachers and 37.8 percent among experienced ones.⁴³

Faced with a perceptual task they cannot perform, readers do what people always do. They substitute a rule they can apply. The rule that emerged is a list of surface features: the em dash, the list of three, certain vocabulary, and the antithetical construction that states a negation before an affirmation.

None of those is a machine artifact. Antithesis is among the oldest catalogued figures in the Western rhetorical tradition, treated as a named device in the classical handbooks and

40. Parth Gaba & Emiliano De Cristofaro, *Group-Differentiated Discourse on Generative AI in High School Education*, arXiv:2603.24972 (2026).

41. Reported in press coverage of a Villanova University study, August 2026. The underlying paper is identified here by its reported findings rather than by citation, and the citation is pending.

42. See Elizabeth Clark et al., *All That's "Human" Is Not Gold: Evaluating Human Evaluation of Generated Text*, in Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics (2021).

43. Study of teacher detection of AI-generated student essays; figures as reported in secondary summaries. Citation pending.

continuous in English prose from the King James translators through Dickens, Lincoln, and Kennedy. The em dash carries Dickinson, Melville, and Woolf. What the folk taxonomy has assembled is not a list of tells. It is a list of the ordinary equipment of formal English, now under suspicion because a system trained on formal English reproduces it at an even rate.

Two consequences follow and both bear on this Article.

The first is a selection effect that lands where the perplexity effect already landed. A heuristic that penalizes rule-faithful prose penalizes most heavily the writers who were drilled hardest on the rules and who apply them most literally. That description fits second language writers, and it fits writers whose cognition is characterized by literal rule application. The classifier flags them because their prose is predictable; the human reader flags them because their prose is correct. Two independent mechanisms, one population, and neither mechanism was designed with that population in view.

The second is a feedback loop. Writers strip the suspected features to avoid accusation. That prose enters the next generation of training data. The features thin out of the corpus, and a false heuristic becomes self-confirming while the language itself is impoverished in the process.⁴⁴

Provenance marking is offered as the answer to all of this, and in one narrow sense it is a real answer. A key is better than a guess. But it does not arrive into an empty field. It arrives into an enforcement culture the folk taxonomy built, staffed by readers who have already demonstrated that they will act on a surface feature, and it hands that culture a new signal that the compliance instrument itself calls unreliable.

44. The loop is described in commentary rather than established empirically. *See Guilty by Punctuation, Linguistically* (Substack) (2026), observing that the tell is the distributional monotony rather than any individual mark, and that the folk taxonomy of machine tells is riddled with false positives.

C. Why Marking Does Not Fix This, and Makes One Thing Worse

Provider side marking is a better instrument than classifier based detection. It starts from a key rather than from a guess, and on unmodified text it should produce far fewer false positives. That is the case for it and it is a real case.

It does not survive contact with two facts.

The first is fragility in one direction. Watermarks do not survive paraphrasing, translation, or heavy editing; the provider says so in its own notice,⁴⁵ the research literature demonstrated the point before deployment, showing that recursive paraphrasing collapses detection rates across watermarking, retrieval based detection, and neural classifiers alike,⁴⁶ and the Code of Practice concedes the point in its own operative text.⁴⁷ A user who wants to evade the mark can. The tooling is commercial and advertised.

The second is the asymmetry that follows. Marking catches the naive user and misses the deliberate one. The person who runs a draft through the model to fix its grammar and posts the result carries the mark. The person who runs generated text through a second model to launder it does not. Sort those two by intent and the instrument has it backwards. Sort them by disability and it is worse, because the first group contains the accommodated user by definition: the point of using the tool as an accommodation is that the output is used, not laundered.

There is a second route around the mark and it is simpler than paraphrase. A user may open a different product. Marking at the model level binds one provider's output, and as of the announcement no other frontier laboratory had deployed production text watermarking at comparable scale.⁴⁸ The first question a national broadcast segment raised on the day of the announcement was exactly that: what stops a student from moving to a competitor's tool.⁴⁹

45. Anthropic Marking Notice, *supra* note 17.

46. Vinu Sankar Sadasivan et al., *Can AI-Generated Text Be Reliably Detected?*, arXiv:2303.11156 (2023).

47. Code of Practice, *supra* note 12, Sub-measure 2.1.2 (lower reliability and robustness for free-form text; may produce misleading or low-confidence results).

48. OpenAI's own compliance statement in late July 2026 described text watermarking as difficult to deploy at scale; contemporaneous reporting characterized Anthropic as the first major laboratory to ship it across its products at once. See *Claude Now Watermarks Text Everywhere*, TechTimes (Aug. 11, 2026).

49. Broadcast transcript on file with the author. The segment is described in Part V.C below.

Substitution is cheap for a user of convenience and expensive for a user of accommodation. A person who has settled on one system as assistive technology has built process around it, and switching is not a free choice between interchangeable products. The mark therefore binds most tightly to the user least able to leave.

There is a finer grained version of substitution that costs almost nothing, and it is the one that matters. A user need not abandon the model he prefers. He can keep its output and route that output through a second model for rephrasing. He retains the quality he wanted and discards the provenance he did not.

This works because marking schemes do not interoperate. Detection is key specific: a detector initialized with one key finds only marks generated with that key, and the published open source release of the leading competing scheme is expressly not interoperable with that provider's production deployment.⁵⁰ There is no shared registry and no common reader. Provenance marking is therefore not one regime. It is a set of parallel regimes, each legible only to its own operator and to whoever that operator equips.

That is where the asymmetry becomes structural rather than incidental. Google's text watermarking has been deployed in its consumer model since 2024, and there is no public tool for reading it; verification requires the provider's key and the verification portal has remained in limited access.⁵¹ Anthropic, by contrast, announced on the day of its marking disclosure that it was building detection tooling for businesses, educators, and third parties.⁵²

The consequence is a ranking nobody designed. What a mark costs the person carrying it depends not on the strength of the mark but on how widely its provider distributes the ability to read it. The provider that distributes detection most broadly exposes its own users most. Compliance at the corporate level is converted into exposure at the individual level, and the individual bearing it is not the one who made the compliance decision.

50. On key specificity and the absence of cross-scheme detection, *see* Dathathri et al., *supra* note 19; *see also* the DeepMind evaluation's own statement that detection works only for services adopting the same scheme.

51. *See* Dathathri et al., *supra* note 19 (production deployment); contemporaneous accounts describe the detector portal as pilot or limited access and note the absence of any public consumer checker for that scheme's text output.

52. *See* the reporting collected *supra* note 17.

One caveat, stated because the literature does not support the stronger claim. Paraphrasing degrades a watermark rather than reliably erasing it, and on long passages a residual signal can survive.⁵³ The practical point is unaffected. A residual signal below a usable threshold, in a scheme whose key the checking institution does not hold, is not an enforcement signal. It is noise.

So the population that ends up carrying a readable mark is not the population using machine assistance most heavily, or most deceptively. It is the population that used one particular provider and then published what came out. The accommodated user is in that group by construction, and the route out of it, running the text through a second system to strip the provenance, is the one route that damages the thing the accommodation produced.

The regime therefore does not replace classifier based detection. It adds a second signal, with a different failure mode, to an enforcement culture that has shown it will act on the first one, and it hands that signal to the same institutions.

D. Who the Regime Says Will Read It

Three independent sources, published within days of one another, name the same recipient.

The provider stated that it is building verification tools to let businesses, educators, and third parties detect marked content.⁵⁴ Educators appear in the provider's own account of who the tool is for.

The Code of Practice names educational and research organisations among the expert end-users guaranteed access to detection for free-form text watermarking, in the same passage that concedes the technique's lower reliability.⁵⁵

And the first mainstream broadcast treatment went directly there. A national segment on the announcement opened with the anchor saying this was a conversation she had with her

53. See Sadasivan et al., *supra* note 43 (recursive paraphrasing collapses detection rates across schemes); the residual-signal qualification is acknowledged in the implementation literature for the leading competing scheme.

54. See the reporting collected *supra* note 17, describing the company's statement that detection tooling is under construction for businesses, educators, and third parties. The detection service had not shipped as of the announcement.

55. Code of Practice, *supra* note 12, Sub-measure 2.1.2.

children every day and that she could not wait to play the story at home, over footage of children in a school hallway, before turning to homework, tests, and book reports and to whether teachers can tell. The reporter identified the education sector as a party paying close attention, and framed the mark as making it easier for anyone to find out whether something was machine generated.⁵⁶

Set that last phrase against the compliance instrument. The Code restricts free-form text detection to a named class subject to access controls. The public was told anyone can check. Whichever description turns out to be operationally true, the gap between them is itself a finding: a regime whose access rules are narrow and whose public reception is universal will be experienced by the marked person as the second one.

The provider says educators. The Code says educational and research organisations. The first broadcast says homework. Nothing in this Part is speculation about where the signal will land.

VI. The Disclosure Trap

A. The Unmarked Baseline

Assistive technology has never been marked.

Speech recognition converts a disabled user's speech into text typographically identical to typed text. Screen readers, refreshable braille displays, word prediction, alternative input, and text to speech all mediate between a disabled person and a document, and none has ever embedded an identifier in the resulting file. Dictation software has been a recognized accommodation for decades and its output carries no signature.

The law has not merely tolerated this. It has ordered it. In 2011 the Ninth Circuit affirmed injunctions requiring a national testing body to let a legally blind applicant use particular screen reading and magnification software on licensing examinations, holding that the governing statute

56. Broadcast transcript on file with the author. The outlet and reporter are not identified with confidence from the transcript alone and are omitted rather than guessed; the segment is used here only as evidence of contemporaneous public framing, not as authority for what the provider said.

demands administration that "best ensure[s]" results reflect aptitude rather than disability, a standard the court accorded *Chevron* deference and placed above the reasonable accommodation standard the testing body urged.⁵⁷ The court further held that the statutory and regulatory lists of auxiliary aids are not exhaustive, reasoning that testing accommodations must keep pace with technology rather than freeze at the aids Congress could name in 1990.⁵⁸

Hold those two things together. A court compelled a testing entity to accept the output of assistive software, on reasoning that the category of assistive software must grow with the technology. Fifteen years later, the newest member of that category ships with a signature identifying its use, and the Code of Practice guarantees educational organisations access to read it.

The line an implementation must draw to justify marking one and not the other runs between software that transcribes a person's language and software that helps shape it. The Commission drew a version of that line when it exempted augmentative and alternative communication and customized neural voices on the ground that they do not alter semantics.⁵⁹ That is a coherent line for communication disabilities. It is not a line that reaches cognitive, attentional, or executive function disabilities, where the assistance is precisely organizational and where the Commission's own examples put summarization and substantive restructuring back inside the marking obligation.

B. Pre-emption of a Disclosure Already Made

The compelled disclosure account stated below assumes an author who would not otherwise disclose. A substantial number of authors in this class already do. For them the mark works differently, and the difference is not a lesser harm.

57. *Enyart v. Nat'l Conference of Bar Examiners, Inc.*, 630 F.3d 1153, 1161–62 (9th Cir. 2011); *see* 42 U.S.C. §12189; 28 C.F.R. §36.309(b)(1).

58. *Enyart*, 630 F.3d at 1164 (quoting committee language anticipating that technological advances would expand what entities must provide).

59. Guidelines, *supra* note 4 (examples following para. 92).

An author's disclosure is an account. It states what the system did, what the author did, and why the assistance is used, and it carries its own context because the author wrote it. It also has one structural property that governs everything else: it lives inside the document, and the document is the only place it can live.

Publishing convention limits it further. In law review format the disclosure occupies a star footnote on the first page and is at least encountered. In most other academic formats it goes in an acknowledgements section at the back, a placement that discharges an obligation to the record while reaching almost no reader. Author disclosure in scholarly publishing is already closer to a filing than to a communication, and this Article does not claim otherwise.

The point is that placement is not the mechanism of the harm, because a detection result does not travel in the document at all. It travels in a separate system, queried by a party performing a different task, at a moment the author does not select and is not informed of. The person running a check is not the person reading the acknowledgements. There is no position within a document at which a disclosure reaches that party, because that party is not reading the document. It is querying it.

The mark therefore does not compete with the author's disclosure. It routes around it. The finding is made in a system in which the author has no standing, on the basis of a signal the author cannot inspect, while the author's own statement on precisely that subject sits in a file the deciding party has no occasion to open. Where the author is later permitted to respond, the response is no longer a description of the work. It is a rebuttal to a finding already made.

The distinction is one of channel rather than content, and secondarily one of sequence. The mark adds no fact the disclosing author was withholding. It creates a route by which the question is answered without the author, and then relocates the author's own answer to the position of a defence. Nothing in the transparency rationale for provenance marking addresses this, because that rationale assumes the alternative to marking is concealment. For the disclosing author the alternative to marking is disclosure, and marking does not supplement it. It bypasses it.

Two consequences follow for the doctrine developed below. First, an institution cannot answer the differential burden by observing that a given author discloses anyway; the burden is not the disclosure but the loss of its position and authorship. Second, the remedies below that turn on notice to the marked person are responsive to this harm in a way that no improvement to disclosure practice can be. An author cannot write her way into a channel she has no access to. Notice is the only mechanism that puts her in it.

C. The Election

The mark does not say that its bearer is disabled. That is the difficulty, not the answer to it.

The mark says the text may have been processed by an artificial intelligence system. In a venue with no rule about such processing, it is inert. In a venue with a rule, it is an accusation arriving without process, obtained from a source the author cannot audit, generated by an instrument the compliance instrument itself calls potentially misleading. The accommodated author then has two options and no third.

Option one is to explain. The complete and truthful explanation is that the author uses the system as an accommodation for a disability. Producing it discloses protected status to a party that did not ask, may not be entitled to it, and owes no duty to keep it.

Option two is silence, which leaves standing the inference the rule was written to support.

The injury is neither the disclosure nor the adverse inference. The injury is the election. And the election is expensive in a way the literature has measured. Concealment of a stigmatized identity is itself an active, resource consuming process associated with worse psychological and physical outcomes; for people with non-apparent disabilities the disclosure decision is the one feature over which they retain control.⁶⁰ The disability studies literature is blunter about the cost:

60. Alecia M. Santuzzi et al., *Invisible Disabilities: Unique Challenges for Employees and Organizations*, 7 *Indus. & Org. Psychol.* 204 (2014).

disclosure is described as a form of forced intimacy, and the labor of substantiating a claim to accommodation has been likened to writing one's own autobiography at book length.⁶¹

That is the thing being taken. Not privacy in the abstract. The one decision the person still held.

D. Protected Status as the Only Available Defense

Employment discrimination law already contains the analytic move this Article needs, and it contains it for this reason.

The prohibition on disability related inquiries does not turn on whether an employer asks about disability. The operative standard, drawn from enforcement guidance and adopted by the courts, is whether a practice is likely to elicit information about a disability.⁶² Applying that standard, the Commission's guidance states that an employer may not ask at the pre-offer stage whether an applicant would need a reasonable accommodation, because such a question is "likely to elicit information about the existence of a disability because, generally, only an individual with a disability would require an accommodation." The First Circuit reproduced that reasoning in construing the provision.⁶³

That is the disclosure trap, stated by the government seventeen years before the technology existed. A question is prohibited not because of its words but because of who alone can answer it.

Two further features of the doctrine matter.

The class of protected plaintiffs is not limited to people with disabilities. The Second Circuit, agreeing with other circuits, held that a plaintiff need not prove an unknown disability to

61. Yvonne Wechuli, *Passing as Able-Bodyminded, Disabled, or Supercrip: Rethinking Impression Management Strategies Through a Disability Lens*, 48 Symbolic Interaction 293 (2024) (collecting the disclosure-as-forced-intimacy and accommodation-labor literature).

62. 42 U.S.C. §12112(d)(4)(A). On the guidance definition and its adoption, *see* Bobnar v. AstraZeneca Pharms. LP, 672 F. Supp. 3d 475 (N.D. Ohio 2023); Kroll v. White Lake Ambulance Auth., 691 F.3d 809, 815 (6th Cir. 2012) (treating EEOC enforcement guidance as very persuasive authority).

63. Andre Grenier v. Cyanamid Plastics, Inc., 70 F.3d 667, 674 (1st Cir. 1995) (quoting EEOC enforcement guidance of May 19, 1994). The quoted passage is Commission guidance rather than the court's independent holding, and is relied on here for the analytic proposition rather than as circuit authority.

challenge a medical inquiry, because the statutory text refers to employees rather than to qualified individuals with disabilities.⁶⁴ A detection regime that operates on everyone is challengeable by everyone it operates on.

And the trigger is low. The same court held that requiring even a general diagnosis suffices, because such a request "may tend to reveal" a disability or give rise to the perception of one.⁶⁵ An explanation that says only "I use an accommodation" clears that bar comfortably.

Against this, the employer's defense is business necessity, and it is not a light one. The asserted necessity must be vital rather than merely convenient, the inquiry must genuinely serve it, and the request must be no broader or more intrusive than necessary; where the policy operates on a class rather than an individual, the employer must justify how it drew the class.⁶⁶ An institution that runs detection across every submission, and treats a hit as presumptive misconduct rebuttable only by disclosure, has defined its class as everyone and its inquiry as maximally broad. Whether integrity enforcement is vital in the required sense, and whether an instrument the Code of Practice itself calls potentially misleading genuinely serves it, are questions that provision was built to ask.

E. The Differential Burden

Two authors, one rule, one instrument.

The first used the system because it was faster. Detected, she faces an integrity finding. Her loss is reputational and academic and it is the loss the rule was designed to impose.

The second used the system because without it the document does not exist. Detected, she faces the same finding unless she spends something the first author is never asked to spend. Her loss is her civil rights interest in controlling disclosure of her own medical status.

One facially neutral rule, applied identically, extracts two different things from two classes, and only one is protected. That is a within class differential burden on meaningful access,

64. *Conroy v. N.Y. State Dep't of Corr. Servs.*, 333 F.3d 88, 94–95 (2d Cir. 2003).

65. *Id.* at 95.

66. *Id.* at 97.

the structure disability doctrine has recognized since *Choate*: identical treatment can itself be the denial where the identity of the treatment is what produces the exclusion.⁶⁷

F. Why Advance Registration Does Not Answer It

The institutional answer is that accommodated users should register in advance, so a detection result can be checked against a file. Three problems.

Registration exists inside institutions. It does not exist for a journal, a conference program committee, a preprint server, a newsletter, an op-ed page, a grant panel, or a platform. Those are where independent scholars and disabled writers outside institutional employment actually publish, and none has an accommodations office.

Registration is itself the disclosure. Requiring it in advance, to preserve the ability to publish without adverse inference, moves the election earlier and makes it mandatory instead of contingent. The empirical literature describes the cost: most disability services offices require proof of status through medical or psychological testing, so obtaining a legal entitlement requires proving membership in a stigmatized group, and even once secured the access must be renegotiated with each instructor every term.⁶⁸

And registration is prospective and local while the mark is retroactive and permanent. An accommodation on file at one institution in one year does nothing when the same paragraph surfaces four years later at a venue that did not exist when the accommodation was granted, checked by a party on the Code's expert access list who has never heard of that office. Accommodation is scoped; the mark is not. Every future reader with access acquires, permanently, the power to reopen a disclosure decision the author already made.

That is the temporal dimension and it is the one most often missed. Control over disclosure is not only control over whether and to whom. It is control over when. A signal

67. *Alexander v. Choate*, 469 U.S. 287, 301 (1985); *see also* *Payan v. L.A. Cmty. Coll. Dist.*, 11 F.4th 729, 738–40 (9th Cir. 2021) (setting out the theories available under Title II and Section 504 and confirming that disparate impact claims survive *Sandoval*).

68. Erin A. Cech, *Engineering Ableism: The Exclusion and Devaluation of Engineering Students and Professionals with Physical Disabilities and Chronic and Mental Illness*, 112 J. Eng'g Educ. 462 (2023).

any authorized holder of the text may interrogate at any time, forever, without notice, is not a disclosure decision the author holds.

VII. Where Existing Doctrine Runs Out

This Part reports rather than advocates. The claims are weakest where the intuition is strongest.

A. *Against the Provider*

Direct claims against the model provider are the weakest available and should not be led with.

Public accommodation. A Title III claim would first have to survive the circuit division over whether a service with no physical locus is a place of public accommodation, and in the nexus circuits it would not.⁶⁹ Even in the broader circuits the theory would have to characterize the addition of a marking feature as a denial of full and equal enjoyment, a step no case has taken.

Consumer protection. The unfairness theory is more plausible and untested. The federal formulation reaches practices causing substantial injury not reasonably avoidable and not outweighed by countervailing benefits,⁷⁰ and while the federal statute affords no private right of action, state analogues do. Illinois directs courts to the Commission's interpretations by statute and asks whether the practice offends public policy, whether it is immoral, unethical, oppressive, or unscrupulous, and whether it causes substantial injury to consumers, holding expressly that all three need not be satisfied because a practice may be unfair by the degree to which it meets one.⁷¹ Non-avoidability is the strongest element on these facts. There is no opt out and no unmarked tier.

69. *Compare* Robles v. Domino's Pizza, LLC, 913 F.3d 898, 905 (9th Cir. 2019) (nexus to a physical place), *with* Carparts Distrib. Ctr., Inc. v. Automotive Wholesaler's Ass'n of New England, Inc., 37 F.3d 12, 19 (1st Cir. 1994) (holding that establishments of public accommodation are not limited to actual physical structures). *See* 42 U.S.C. §12182(a).

70. 15 U.S.C. §45(n).

71. *Robinson v. Toyota Motor Credit Corp.*, 201 Ill. 2d 403, 417–18, 775 N.E.2d 951, 960–61 (2002); 815 ILCS 505/2.

Contract. The strongest provider facing theory is neither property nor disability, and it does not depend on any disability fact. It is worth stating at length because it is the one theory on which the marked person is an ordinary plaintiff rather than a member of a class the regime never contemplated.

Begin by conceding assent. Assume the user clicked, assume the click bound him, and assume every formation defect a court might otherwise find is absent. The question that survives is what he assented to.

The provider assigns the user all of its right, title, and interest, if any, in outputs.⁷² It disclaims every warranty by name, including title and fitness for a particular purpose.⁷³ And it reserves the right to modify the service, specifying a notice period only for fee increases.⁷⁴ Those three clauses together produce an arrangement with an unusual shape. The user holds title to an artifact whose material characteristics the counterparty may alter unilaterally after formation, without notice, while warranting nothing about the thing assigned.

Three contract defects follow and each is independent of the others.

A promise the promisor may modify at will, including as to the properties of the very thing being conveyed, is illusory. The consideration flowing to the user is defined by a term the provider may rewrite. Where a party retains an unfettered right to alter its own performance, the promise supplies nothing the other party can enforce.

Changes delivered without notice do not bind the party who never saw them. A modification effective on posting, to a document the user has no occasion to reread, is not an agreement to which the user has agreed. The point is not that notice is a formality the provider skipped. It is that assent has to attach to something, and it cannot attach to a term the assenting party has never encountered.

And an arrangement whose terms and whose product characteristics reflow continuously lacks the definiteness a contract requires. Definiteness is not a technicality about drafting quality.

72. Anthropic, *Consumer Terms of Service* §4 (2026), <https://www.anthropic.com/legal/consumer-terms>.

73. *Id.* §11.

74. *Id.* §12.

It is the requirement that a court be able to say what the parties agreed to, and a court cannot do that where the subject matter of the assignment is a moving object.

There is a consumer protection dimension as well, and it is where the disability fact re-enters without being load bearing. The cost imposed by a continuously changing product is the cost of relearning it, and that cost is not evenly distributed. It falls hardest on users whose personal circumstances impair the executive function and working memory that relearning demands. An unconscionability analysis that asks who actually pays the price of a reserved right to amend will find the same population this Article has been describing, arriving from a different direction.

The provider's escape from each of these defects is a single clause, the reserved right to amend. That clause is separately contestable on assent grounds. This Article does not rely on that challenge. It grants assent and shows the agreement fails anyway.

None of these captures the harm. Ownership is not the injury; the user does own the marked text and no unmarked version was withheld. The injury is informational, and it happens later, in someone else's hands.

B. Against the Deployer

The stronger claims run against the institutions that act on detection results, and they are largely unexplored. The Code of Practice makes them stronger still, because it names those institutions and guarantees them the access.

Public entities and fund recipients. A public university, state agency, or federally funded institution that adopts detection and imposes consequences without a path by which an accommodated user can respond outside the disclosure election imposes a differential burden on meaningful access.⁷⁵ Remedies are constrained but not foreclosed: emotional distress

75. 29 U.S.C. §794(a); 28 C.F.R. pt. 35; *Choate*, 469 U.S. at 301.

damages are unavailable under the Spending Clause line, while compensatory damages for lost opportunities remain available as economic harm.⁷⁶

Employers. An employer that treats a detection result as presumptive misconduct, rebuttable in practice only by an accommodation explanation, has built a practice likely to elicit disability information. It never asks the prohibited question. It arranges the circumstances in which the answer is volunteered, and the case law supplies both the standard and the defense the employer would have to carry.⁷⁷

Testing and credentialing entities. Entities offering examinations and courses related to licensing, certification, or credentialing carry an independent duty to offer them in a place and manner accessible to individuals with disabilities, measured by the best ensure standard rather than by reasonableness.⁷⁸ An examination or credentialing regime that treats an assistive tool's signature as evidence of misconduct, with no adjudicative path that avoids disclosure, is a candidate for that provision, and the author has located no case applying it to an integrity or proctoring dispute.⁷⁹

C. The Gap

Line the parties up at the moment of injury and no one is holding it.

The provider created the signal. It has no relationship with the institution that read it, no obligation to tell the marked person a check occurred, and no duty to her at all. It can point, accurately, to a European instrument, a set of Guidelines, and a Code of Practice, and it can point to the Code's own text authorizing the access it granted.

The institution read the signal. It did not create it, cannot audit it, and did not choose its scope. It can point, accurately, to an integrity policy of general application that says nothing

76. *Cummings v. Premier Rehab Keller*, P.L.L.C., 596 U.S. 212 (2022); *Payan v. L.A. Cmty. Coll. Dist.*, 169 F.4th 971, 977–79 (9th Cir. 2026).

77. 42 U.S.C. §12112(d)(4)(A); *Conroy*, 333 F.3d at 94–95, 97.

78. 42 U.S.C. §12189; 28 C.F.R. §36.309(b)(1); *Enyart*, 630 F.3d at 1161–62.

79. The absence is stated as a search result, not as a conclusion of law. *Enyart* does not analyze whether the §12189 duty attaches independently of the place of public accommodation requirement, and this Article does not assume that it does.

about disability, and to a Code of Practice naming educational and research organisations as legitimate recipients.

The marked person is injured by the conjunction and has, at that moment, a strong claim against neither.

Accommodation law is architecturally a dyad. A person requests; an entity provides or refuses; the refusal is reviewable. The interactive process, the undue hardship defense, the reasonableness inquiry, all presuppose two parties in a relationship. Provenance marking introduces a third element that is not a party. It is a property of the artifact, created upstream by someone with no duty and read downstream by someone with a guaranteed right of access and no knowledge of the artifact's provenance. There is no interactive process with a watermark.

That gap is this Article's contribution and better lawyering under existing doctrine does not close it.

VIII. Repairs

Four repairs are available without new legislation, and each has a hook in a document that already exists.

Conditional marking. The Commission has already said that where a system performs both generation and non-substantial minor alteration, the marking obligation does not apply to the content altered in the minor manner.⁸⁰ A provider may therefore implement an assistive mode in which operations the Guidelines put inside the exception, grammar correction, spellchecking, minor stylistic polishing, and translation of user supplied text, are not marked. This requires no legal change. Note the boundary honestly: summarization and substantive rewriting are outside the exception on the Commission's own examples, and an assistive mode cannot reach them without an amendment.

80. Guidelines, *supra* note 4, para. 92.

Notice to the marked. The Code already requires that a detection result be downloadable in a digitally signed format carrying a content hash, a solution identifier, and a timestamp.⁸¹ The infrastructure for an auditable detection event exists and runs one direction. Extending the same receipt to the account whose output was queried costs the provider nothing it does not already build, and converts a covert signal into a symmetrical one.

Deployer duty before adverse inference. Institutions adopting detection should be required, by policy where not by law, to find misconduct only after a process that does not require the respondent to disclose protected status in order to prevail. A safe harbor in which a respondent asserts an accommodation without specifying it, and the finding is suspended pending review by a party bound to confidentiality, is available now. The business necessity framework already demands something close to this of employers, and the Code's own concession about reliability supplies the reason.

The unmarked alternative as a modification of policy. Where a covered entity's own policy generates the trap, providing an alternative route to evaluation that does not depend on the mark is a modification of policy, and the analysis proceeds under existing law.

The fifth repair needs a legislature. The assistive exemption should be restated as a right of the assisted person rather than as a limit on the provider's duty. As drafted it tells the provider what it need not do; nothing lets the person whose output is marked invoke it, and the Code that operationalizes the regime does not mention her. A duty with no correlative right is enforceable only by a market surveillance authority with no particular reason to look here.

IX. Conclusion

The sentence exists. Article 50(2) says the marking obligation does not apply to the extent a system performs an assistive function for standard editing, and the Commission has said that grammar correction, minor stylistic polishing, translation, and assistive technology that lets a disabled person communicate all fall inside it. Someone in the drafting understood that there is

81. Code of Practice, *supra* note 12, Sub-measure 2.1.2.

a difference between a machine that produces content and a machine that helps a person produce content.

The line they drew stops short. It reaches communication disabilities and not cognitive ones, and the Commission's own examples put summarization and restructuring back inside the marking duty. Then the first implementations went further and marked everything, everywhere, and published documentation conceding that the resulting signal cannot separate the cases the exception exists to separate. Then the Code of Practice conceded that the signal for text is unreliable and may mislead, and guaranteed access to it to a list of institutions that includes every university.

The people who will pay are not hard to identify. They already write with help. They already disclose it when they choose to. They now carry, in every paragraph they publish, a machine readable statement about their own writing that they did not make and cannot unmake, readable on demand by parties named in a compliance instrument, with no notice to them, ever. The regime calls this transparency. For them it is the inverse: a fact about their bodies, encoded in their prose, waiting for someone with access to ask.

☛ ☚

References

- Alexander v. Choate, 469 U.S. 287 (1985).
- Americans with Disabilities Act of 1990, 42 U.S.C. §§12101–12213.
- Andre Grenier v. Cyanamid Plastics, Inc., 70 F.3d 667 (1st Cir. 1995).
- Anthropic. (2026). *Consumer terms of service*. <https://www.anthropic.com/legal/consumer-terms>.
- Anthropic. (2026). *How Claude marks AI-generated content*. <https://support.claude.com/en/articles/16266773-how-claude-marks-ai-generated-content>.
- Bobnar v. AstraZeneca Pharmaceuticals LP, 672 F. Supp. 3d 475 (N.D. Ohio 2023).
- Carparts Distribution Center, Inc. v. Automotive Wholesaler’s Association of New England, Inc., 37 F.3d 12 (1st Cir. 1994).
- Cech, E. A. (2023). Engineering ableism: The exclusion and devaluation of engineering students and professionals with physical disabilities and chronic and mental illness. *J. Eng’g Educ.*, 112, 462. <https://doi.org/10.1002/jee.20522>.
- Conroy v. New York State Department of Correctional Services, 333 F.3d 88 (2d Cir. 2003).
- Cummings v. Premier Rehab Keller, P.L.L.C., 596 U.S. 212 (2022).
- Dathathri, S., et al. (2024). Scalable watermarking for identifying large language model outputs. *Nature*, 634, 818.
- Enyart v. National Conference of Bar Examiners, Inc., 630 F.3d 1153 (9th Cir. 2011).
- European Commission. (2026). *Guidelines on transparency obligations for providers and deployers of certain AI systems*. <https://digital-strategy.ec.europa.eu/en/policies/guidelines-transparency-ai-generated-content>.
- European Commission. (2026). *Transparency obligations under Article 50 of the AI Act*. <https://digital-strategy.ec.europa.eu/en/faqs/transparency-obligations-under-article-50-ai-act>.
- Federal Trade Commission Act, 15 U.S.C. §45.
- Gaba, P., & De Cristofaro, E. (2026). *Group-differentiated discourse on generative AI in high school education*. arXiv:2603.24972.
- Gegg-Harrison, W., & Quarterman, C. (2024). AI detection’s high false positive rates and the psychological and material impacts on students. In *Academic integrity in the age of artificial intelligence* (p. 199). IGI Global. <https://doi.org/10.4018/979-8-3693-0240-8.ch011>.
- Ibrahim, K. H. S., AlOtaibi, D., & Sibai, F. N. (2025). The robustness of AI-classifiers in the face of AI-assisted plagiarism. *Int’l J. Computer-Assisted Language Learning & Teaching*, 15, 1. <https://doi.org/10.4018/ijcallt.375428>.
- Illinois Consumer Fraud and Deceptive Business Practices Act, 815 ILCS 505.
- Kirchenbauer, J., et al. (2023). A watermark for large language models. In *Proceedings of the 40th International Conference on Machine Learning*.
- Kroll v. White Lake Ambulance Authority, 691 F.3d 809 (6th Cir. 2012).

- Liang, W., et al. (2023). GPT detectors are biased against non-native English writers. *Patterns*, 4, 100779. <https://doi.org/10.1016/j.patter.2023.100779>.
- Nondiscrimination on the Basis of Disability by Public Accommodations, 28 C.F.R. §36.309.
- Payan v. Los Angeles Community College District, 11 F.4th 729 (9th Cir. 2021).
- Payan v. Los Angeles Community College District, 169 F.4th 971 (9th Cir. 2026).
- Regulation (EU) 2024/1689 (Artificial Intelligence Act), 2024 O.J. (L 1689).
- Rehabilitation Act of 1973 §504, 29 U.S.C. §794.
- Robinson v. Toyota Motor Credit Corp., 201 Ill. 2d 403 (Ill. 2002).
- Robles v. Domino's Pizza, LLC, 913 F.3d 898 (9th Cir. 2019).
- Sadasivan, V. S., et al. (2023). *Can AI-generated text be reliably detected?* arXiv:2303.11156.
- Santuzzi, A. M., Waltz, P. R., Finkelstein, L. M., & Rupp, D. E. (2014). Invisible disabilities: Unique challenges for employees and organizations. *Indus. & Org. Psychol.*, 7, 204. <https://doi.org/10.1111/iops.12134>.
- Shekhar, A. R., et al. (2026). *DoPE: Decoy oriented perturbation encapsulation*. arXiv:2601.12505.
- Weber-Wulff, D., Anohina-Naumeca, A., Bjelobaba, S., Foltýnek, T., Guerrero-Dib, J., et al. (2023). Testing of detection tools for AI-generated text. *Int'l J. for Educ. Integrity*, 19, art. 26. <https://doi.org/10.1007/s40979-023-00146-z>.
- 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/>.
- Elkhatat, A. M., Elsaid, K., & Almeer, S. (2023). Evaluating the efficacy of AI content detection tools in differentiating between human and AI-generated text. *Int'l J. for Educ. Integrity*, 19, art. 17. <https://doi.org/10.1007/s40979-023-00140-5>.
- European Commission. (2026). *Code of practice on transparency of AI-generated content*. <https://digital-strategy.ec.europa.eu/en/policies/code-practice-ai-generated-content>.
- 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>.
- Wechuli, Y. (2024). Passing as able-bodiedminded, disabled, or supercrip: Rethinking impression management strategies through a disability lens. *Symbolic Interaction*, 48, 293. <https://doi.org/10.1002/symb.711>.