The Underwriter's Exhibit:

Employment Practices Insurance, Algorithmic Hiring, and the Disability Plaintiff's Proof

Travis Gilly*

Working paper, July 2026 (v0.5)

ABSTRACT

An employer runs an automated screen across thousands of applications, a disabled applicant is rejected, and she files a charge. The governance industry has a product ready for that moment: documentation. Its promise is that a well kept record, bias testing, human review, an audit trail, lets the organization show the decision was defensible. At the same moment that market matured, the insurance market reached the opposite conclusion about the same technology, and carriers began writing artificial intelligence out of the very policies that would pay an employment discrimination judgment, including, by named endorsement, the employment practices line. This Article identifies a loop between those two markets that the scholarship has not examined. Where the employment practices insurer still offers coverage of an algorithmic hiring claim rather than excluding it outright, it conditions that coverage on the production of specific artifacts, whether bias testing was performed, whether a human reviewed the output, whether the tool was validated. Those artifacts are the same ones that establish a disability plaintiff's case and defeat the employer's defense. The bias test that unlocks the policy is the disparate impact statistic. The human review that unlocks the policy is the adopted recommendation that carries agency liability under *Mobley v. Workday*, and a balanced overall outcome cannot cure a specific discriminatory barrier under *Connecticut v. Teal*. The validation the underwriter asks after is the business necessity showing that *Albemarle Paper Co. v. Moody* places on the employer and that an unexplainable model cannot supply. Where an assessment is designed to infer a mental or neurological condition, its administration before a conditional offer is a completed violation the moment it runs, as *Karraker v. Rent-A-Center* shows for a personality inventory. The demand for the record is not even unique to insurance: a city audit-publication mandate and the civil rights recordkeeping rules point at the same file. And the deployer's apparent escape, running the audit under privilege, fails, because the two acts the insurance regime compels, disclosure to the underwriter and reliance on the controls as a defense, are the two classic waivers of any self evaluative privilege, the underlying facts are not privileged under *Upjohn* regardless, and courts have already declined to shield discrimination audits. The Article locates the remedy the litigation actually seeks, an injunction to change the design, outside insurability altogether, and argues that in this domain governance documentation does not transfer risk. It converts risk into a discoverable record while the coverage that would have paid is withdrawn.

Keywords: employment practices liability insurance; algorithmic hiring; automated employment decision tools; Americans with Disabilities Act; Section 504; disparate impact; business necessity; pre-offer medical examination; *Albemarle Paper Co. v. Moody*; *Karraker v. Rent-A-Center*; *Mobley v. Workday*; *Connecticut v. Teal*; *Upjohn Co. v. United States*; self critical analysis privilege; insurance as governance; discovery; bias audit.

*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. Correspondence concerning this article should be addressed to Travis Gilly, Real Safety AI Foundation, Illinois, United States. Email: t.gilly@ai-literacy-labs.org. ORCID: 0009-0007-2954-6313.

CONTENTS

I. INTRODUCTION

Consider the sequence that this Article is about. An employer deploys an automated system to screen a large applicant pool. A qualified applicant with a disability is not advanced. She files a charge of discrimination. The employer has an answer that is by now familiar, and that a growing advisory market has trained it to give: the decision was not made by the machine, a recruiter and a hiring manager reviewed every recommendation, and here is the file that proves it. The file contains the job posting, the score, the human sign off, and the final decision. The organization is confident because it did what it was told to do. It kept the record.

That confidence rests on a premise worth naming, because two markets have taken opposite positions on it. The premise is that a sufficiently complete record can defend an automated employment decision. An entire category of governance advising is built on that premise. Its practitioners speak of explainability, of audit trails, of the questions to ask and the boxes to check, and they sell the resulting documentation as the thing that will carry the employer through an investigation.[1]

At roughly the same moment that market matured, the people whose business is pricing risk reached the opposite conclusion about the same technology. In late 2025, major carriers filed with state regulators to exclude artificial intelligence liabilities from standard corporate policies, and the standard form publisher followed by making a generative artificial intelligence exclusion available across the commercial general liability form.[2] The carriers were not raising premiums. They were declining to write the risk, which is what an insurer does with a loss it cannot model. As one insurance executive put it, the industry can absorb a single large loss from one company's deployment, but not a systemic, correlated event in which one upstream model failure produces thousands of losses at once.[3]

The two markets cannot both be right, and the gap between them is where this Article works. The governance market sells the record as protection. The insurance market is withdrawing the coverage that would pay if the record fails. And the specific mechanism by which the insurance market is withdrawing that coverage, described in Part II, turns out to interact with disability discrimination doctrine in a way that neither market appears to have noticed. The interaction is a loop, and the loop is the contribution.

Here is the loop stated plainly. Where an employment practices liability insurer still offers coverage for an algorithmic hiring claim rather than excluding it outright, the insurer increasingly conditions that coverage

[1]The governance advisory literature is large and largely promotional. For representative framing of documentation and human review as the mechanisms by which an organization demonstrates a defensible automated decision, see the practitioner sources collected *infra* notes 21–4. This Article treats that framing as the market position it is, and asks whether it holds against disability doctrine.

[2]Lee Harris & Cristina Criddle, *Insurers Retreat from AI Cover as Risk of Multibillion-Dollar Claims Mounts*, Fin. Times (Nov. 23, 2025), https://www.ft.com/content/abfe9741-f438-4ed6-a673-075ec177dc62. The reporting describes AIG, Great American, and W.R. Berkley seeking regulatory approval to carve artificial intelligence exposure out of corporate coverage, and quotes an AIG filing describing generative artificial intelligence as a wide-ranging technology whose claims will likely increase over time.

[3]The characterization of the feared loss as correlated and systemic, rather than a single large claim, is attributed to Aon's Kevin Kalinich in the Financial Times reporting: the industry can absorb a single large loss but not a thousand or ten thousand simultaneous losses. *See supra* note 2. The same concern appears in the peer-reviewed literature. *See* Daniel Bauer, Joan Schmit & Justin Sydnor, *The Economics of Emerging Insurance Technologies: Theory and Early Evidence*, 91 J. Risk & Ins. 809 (2024) (noting that correlated cyber and algorithmic loss events may render some existing coverage uninsurable in its current form).

on documentation: it asks, at renewal, whether the insured uses automated tools in employment decisions, whether it has conducted bias testing on those tools, and whether human review exists in the process. Without that documentation, the carrier narrows or denies.[4] Now set those same three artifacts against the elements of a disability claim. The bias testing the underwriter asks for is the disparate impact statistic that makes the plaintiff's prima facie case. The human review the underwriter asks for is the adopted recommendation that keeps the vendor's tool inside the liability chain as the employer's agent, a theory a federal court has already accepted.[5] And the validation the underwriter is really asking after, whether the tool actually predicts job performance, is the business necessity showing that the Supreme Court places on the employer and that an unexplainable model cannot supply.[6] The file that unlocks the policy is the file that proves the claim.

The demand for that file is not even unique to insurance, which is what makes the deployer's position acute rather than merely unlucky. A civil rights recordkeeping regime already requires employers to retain records of the adverse impact of their selection procedures.[7] And in at least one jurisdiction, a city ordinance requires a bias audit of an automated employment tool and requires that audit to be made publicly available, so the disparate impact statistic is not something a plaintiff must pry loose in discovery; the employer must publish it.[8] Three regimes point at the same record. Insurance is the newest and, because it is contractual and self interested, the most revealing.

The employer that sees this coming has an apparent escape, and Part V is devoted to why it fails. The escape is to run the bias audit under legal privilege so that its results never surface. That move does not work, for a reason that is structural rather than jurisdictional. The self critical analysis privilege is not reliably recognized to begin with, and federal courts have specifically declined to use it to shield an employer's internal audit from discovery in an employment case.[9] Even where some form of the privilege is available, it is waived by exactly the two acts that the insurance regime compels. Relying on an internal audit as a defense waives the privilege, and voluntarily disclosing the audit waives it.[10] And even a privileged legal analysis does not reach the fact at the center of the loop, because the attorney client privilege protects communications, not the underlying facts.[11] To satisfy the underwriter the employer must disclose. To defend the claim the employer must rely. Either act is a waiver, and the disparate impact is a fact in any event, so the employer

[4] See sources collected *infra* Part II.C. The pattern reported across carrier and broker commentary is an artificial intelligence questionnaire at renewal, with sublimits or denials where the insured cannot document bias testing and human review; several sources describe carriers requiring what one calls audit-grade governance evidence before renewing coverage that touches artificial intelligence exposure.

[5] *Mobley v. Workday, Inc.*, 740 F. Supp. 3d 796 (N.D. Cal. 2024).

[6] *Albemarle Paper Co. v. Moody*, 422 U.S. 405 (1975).

[7] 29 C.F.R. §1607.4(A), (B) (Uniform Guidelines on Employee Selection Procedures) (requiring covered employers to maintain records disclosing the impact of selection procedures on protected groups); *id.* §1607.5 (validation standards).

[8] N.Y.C. Local Law 144 of 2021 (Automated Employment Decision Tools), eff. July 5, 2023 (requiring a bias audit of an automated employment decision tool and public availability of a summary of results). *See* Katherine B. Forrest, *We Have No Idea What We Are Walking Into: AI and Ethical Considerations*, 1534 Annals N.Y. Acad. Sci. 19, 21 (2024) (describing Local Law 144 and the emerging state patchwork).

[9] *See Univ. of Pa. v. EEOC*, 493 U.S. 182 (1990); *Deel v. Bank of Am., N.A.*, 227 F.R.D. 456 (W.D. Va. 2005).

[10] *See Volpe v. US Airways, Inc.*, 184 F.R.D. 672 (M.D. Fla. 1998) (reliance on internal investigation as a defense waives any privilege over it); *Lawndale Restoration Ltd. P'ship v. Acordia of Ill., Inc.*, 853 N.E.2d 791 (Ill. App. Ct. 2006) (voluntary disclosure of an audit report waives the statutory self evaluative privilege).

[11] *Upjohn Co. v. United States*, 449 U.S. 383, 395–396 (1981).

cannot both unlock the coverage and keep the proof out of the plaintiff's hands.

A word on what this Article does not claim, because the surrounding literature is crowded and the contribution is narrow. It does not argue that algorithmic hiring tools discriminate against people with disabilities. That question has been examined at length, and the doctrinal hooks for reaching such tools are largely in place.[12] It does not argue that artificial intelligence is uninsurable, a proposition the insurance trade press and a developing scholarly literature have covered.[13] The unexamined question is the interaction: what happens to a disability defendant when the insurance market's chosen method of managing artificial intelligence risk is to demand precisely the documentation that disability doctrine treats as proof of liability. The answer is a closed loop with no clean exit, and it is worth a court's and an editor's attention because the deployer's dominant strategy under it is self defeating in both directions.

The Article proceeds in six further Parts. Part II sets the two markets side by side and isolates the line of coverage that actually matters, employment practices liability insurance, distinguishing it from the general liability exclusions that dominate the commentary and identifying the named endorsement that now attaches an absolute artificial intelligence exclusion to the employment practices coverage part. Part III explains why the disability claim does not run on the explainability axis that the governance market is built to serve, walking the three tracks along which an automated screen creates disability liability and showing that opacity indicts the disability defendant rather than shielding it. Part IV states the loop, maps each coverage condition onto the disability element it supplies, and shows that the demanded record is overdetermined by three converging regimes. Part V takes the privilege escape and shows why it collapses. Part VI turns to implications, situates the phenomenon in the literature on insurance as a private regulator, and states the limits of the argument honestly. Part VII concludes.

II. Two Markets, Opposite Verdicts

A. The Governance Market Sells the Record

The advisory market that has grown up around workplace artificial intelligence sells a promise of demonstrable diligence. The recurring move is to treat the risk as an evidentiary problem: if the organization documents its process, retains its audit trail, and preserves evidence of human involvement, it can reconstruct and defend any given decision after the fact. The framing is intuitive and, in the disparate treatment setting, not wrong. A claim that turns on intent can often be met with a record that shows a legitimate, documented, humanly supervised basis for the decision.

The trouble is that the framing is sold as general when it is domain specific. The evidentiary posture that

[12] On algorithmic discrimination in hiring and the tools available to reach it, see Solon Barocas & Andrew D. Selbst, *Big Data's Disparate Impact*, 104 Calif. L. Rev. 671 (2016); Pauline T. Kim, *Data-Driven Discrimination at Work*, 58 Wm. & Mary L. Rev. 857 (2016); Jeremias Adams-Prassl, Reuben Binns & Aislinn Kelly-Lyth, *Directly Discriminatory Algorithms*, 86 Mod. L. Rev. 144 (2023); Nicole Sheard, *Algorithm-Facilitated Discrimination: A Socio-Legal Study of the Use by Employers of Artificial Intelligence Hiring Systems*, 52 J.L. & Soc'y 269 (2025); Maarten Buyl et al., *Tackling Algorithmic Disability Discrimination in the Hiring Process*, in Proc. 2022 ACM Conf. on Fairness, Accountability, and Transparency 289 (2022). On the vendor as agent, see *Mobley*, 740 F. Supp. 3d 796.

[13] *See supra* notes 2, 3; sources collected *infra* Part II.B.

answers a disparate treatment claim does not answer a disability claim, because the disability claim does not ask who decided or whether the reasoning can be reconstructed. It asks whether the tool was permitted at all, whether an accommodation was provided, and whether the selection device can be validated. Those are not questions a richer file resolves in the employer's favor. As Part III shows, a complete and humanly supervised file can document a violation with precision, and an incomplete file can leave a violation uncured in a way no later reconstruction reaches.

B. The Insurance Market Is Pricing the Risk Out

While the governance market was selling the record, the insurance market was reaching the opposite verdict on the same technology. The reason the carriers give is consistent: generative model output is difficult to model actuarially, and the feared loss is a systemic, correlated event rather than a single large claim.[14]

The retreat has taken several forms. First, standard form exclusions. Effective January 1, 2026, the standard form publisher released endorsements excluding coverage under the commercial general liability form for injury arising out of generative artificial intelligence.[15] Second, regulatory clearance at scale for exclusions across general and management liability lines. Industry reporting indicates that several large carriers, including W.R. Berkley, Chubb, Travelers, Berkshire Hathaway, and Cincinnati Financial, obtained or sought state approval to add artificial intelligence exclusions across general liability, directors and officers, and errors and omissions lines, with regulators approving a large majority of the requests.[16] Third, and most important for this Article, bespoke absolute exclusions written into the management liability lines by named endorsement, discussed in Part II.C.

The replacement market is thin. Some carriers have begun to offer affirmative, standalone artificial intelligence coverage as a dedicated product, with limits that are modest relative to the aggregate exposure a widely deployed system can generate.[17]

This arc is not new to insurance, and the precedent is instructive. The sociolegal literature documents the same sequence in environmental liability: general liability insurers first used weak pollution exclusions that courts refused to enforce, then moved to stronger exclusions that courts did enforce, while a thin specialty

[14] See *supra* notes 2, 3.

[15] Ins. Servs. Office, Inc., *Exclusion — Generative Artificial Intelligence*, Form CG 40 47 (eff. Jan. 1, 2026) (excluding bodily injury, property damage, and personal and advertising injury under Coverage A and Coverage B); *Exclusion — Generative Artificial Intelligence (Coverage B Only)*, Form CG 40 48 (eff. Jan. 1, 2026); *Exclusion — Generative Artificial Intelligence (Products/Completed Operations)*, Form CG 35 08 (eff. Jan. 1, 2026). The forms are optional endorsements. Their content and January 1, 2026 effective date are described consistently across multiple carrier and law firm analyses, which report that carrier adoption is expected to approach near universal by the end of 2026.

[16] Industry reporting compiling carrier filings, citing PYMNTS, states that regulators approved more than 80 percent of the requests, with Florida, Connecticut, and Maryland among the leading approving states. The list of carriers seeking approval across general liability, directors and officers, and errors and omissions lines includes W.R. Berkley, Chubb, Travelers, Berkshire Hathaway, and Cincinnati Financial.

[17] Reporting describes affirmative, standalone artificial intelligence liability programs, including one backed by Lloyd's capacity (Armilla, with Chaucer and Axis) reporting limits around twenty-five million dollars, a program launched by Counterpart, and coverage introduced by HSB (a Munich Re company) for small and midsize enterprises. The point for this Article is only that the affirmative market is nascent and capped, so it does not neutralize the withdrawal of coverage from the standard lines.

market sold narrower coverage at higher prices with lower limits to buyers who had no choice.[18] Silent artificial intelligence exposure is traveling the same path that silent cyber traveled, and the endpoint the environmental and cyber histories predict is a market in which the standard lines exclude the risk and a narrow, expensive specialty market prices what remains.

C. The Line That Matters Is Employment Practices Liability, Not General Liability

A precision is necessary here, and getting it wrong would sink the argument. The exclusions that dominate the commentary are the commercial general liability endorsements, and the general liability form does not cover employment discrimination in the first place. It responds to bodily injury, property damage, and personal and advertising injury.[19] An employment discrimination claim, including a claim of disability discrimination in hiring, is the province of employment practices liability insurance, typically written within a management liability program alongside directors and officers and fiduciary coverage.[20] The general liability endorsements are therefore context, not the mechanism.

The mechanism is what is happening on the employment practices line, and it now has a form number. W.R. Berkley has filed an endorsement it styles an absolute artificial intelligence exclusion that, by its terms, amends the employment practices coverage part along with the directors and officers and fiduciary parts, and that bars any claim arising out of the actual or alleged use, deployment, or development of artificial intelligence.[21]

Where the carrier has not excluded artificial intelligence outright, it is conditioning employment practices coverage on documentation, and this is the development that produces the loop. Underwriters are adding artificial intelligence questions at renewal. A broker executive describes the now standard set as whether the insured uses artificial intelligence, whether it polices it, and whether it has protocols in place, and observes that professional liability underwriting is converging on the posture cyber underwriting adopted years earlier, in which documented controls are a precondition of coverage.[22] For the employment practices line specifically, reporting describes questionnaires asking whether automated tools are used in hiring,

[18] *See* Tom Baker & Anja Shortland, *The Government Behind Insurance Governance: Lessons for Ransomware*, 17 Regul. & Governance 1000 (2023) (tracing the environmental liability sequence from weak to enforced exclusions and the emergence of higher priced, narrower specialty coverage). The commentary describing the current artificial intelligence retreat expressly analogizes it to the earlier transition from silent cyber exposure to affirmative cyber endorsements and standalone cyber products.

[19] See *supra* note 15. The excluded coverages under the commercial general liability endorsements are bodily injury, property damage, and personal and advertising injury, none of which is the coverage that responds to an employment discrimination claim.

[20] Employment practices liability insurance responds to claims of discrimination, harassment, retaliation, and wrongful termination, and is commonly written as part of a management liability program. To say a disability hiring judgment goes unpaid because of a general liability endorsement would name the wrong policy.

[21] W.R. Berkley (Berkley Insurance Co.), Form PC 51380 00 (06-24), *Artificial Intelligence Exclusion (Absolute)*, amending the Directors & Officers, Employment Practices, and Fiduciary coverage parts of the Berkley management liability family, and defining artificial intelligence broadly as a machine based system that infers, from the input it receives, how to generate outputs. The form and its attachment to the employment practices coverage part are documented in multiple broker and governance analyses and in a major law firm's insurance recovery commentary describing Berkley's absolute artificial intelligence exclusion in specialty lines. Where such an exclusion is attached to an employment practices policy, an algorithmic hiring claim is simply uninsured.

[22] The standard renewal questions — whether the insured uses artificial intelligence, polices it, and has protocols in place — are attributed to Aon's Stan Sterna in trade reporting (April 2026), which also observes that professional liability underwriting is converging on the documented controls posture that cyber underwriting adopted years earlier. At least one professional liability application form contains a verbatim artificial intelligence question.

promotion, and termination, whether bias testing has been conducted, and whether human review exists, with carriers responding by sublimit or denial where the insured cannot document those controls.[23] That conditioning is the hinge on which the rest of the Article turns.

III. Why the Disability Claim Does Not Run on Explainability

The governance market is built to answer the question of who decided and why. The disability claim does not ask that question. This Part walks the three tracks along which an automated hiring screen creates disability liability, and shows that on each of them the record that reconstructs a decision is either beside the point or affirmatively harmful to the employer.

A. Three Tracks to Liability

1. The pre-offer medical examination bar

The first track is the one on which the timing of the violation is fixed and no later explanation reaches it. The Americans with Disabilities Act categorically forbids a covered employer from conducting a medical examination or making disability related inquiries of an applicant before a conditional offer of employment.[24] The reach of that bar into modern assessment tools is not hypothetical. In *Karraker v. Rent-A-Center*, the Seventh Circuit held that a widely used personality inventory was a medical examination under the Act because it was designed, at least in part, to reveal mental illness, and it held that the character of the instrument did not change with the manner of scoring: whether the employer read the results clinically or through a vocational protocol, and whether or not a psychologist interpreted them, the instrument remained a medical examination and its use in the pre-offer process violated the Act.[25] The line *Karraker* draws, between an instrument designed to identify a mental disorder and one that measures ordinary traits, is the line on which a large class of contemporary hiring tools sits. A gamified assessment or an inferential model that predicts a psychological or neurological condition falls on the medical examination side of that line, and administering it before a conditional offer completes the violation at the moment of administration. Consent does not cure it, because the statute withholds the examination at that stage regardless of any waiver. On this track the deployer's carefully maintained file does not walk the employer back through the gate; it documents the moment the employer walked through it.

---

[23] Practitioner reporting describes employment practices underwriting questionnaires that ask whether automated tools are used in employment decisions, whether bias testing has been conducted, and whether human review exists, and describes sublimits or denials absent that documentation. One widely circulated illustration posits a claim that would have been covered at five million dollars capped at five hundred thousand dollars for lack of documented bias testing and human review; that figure appears in broker commentary as an illustrative hypothetical rather than a reported claim outcome, and is used here only to describe the mechanism. Several sources describe carriers requiring audit-grade governance evidence before renewing coverage that touches artificial intelligence exposure, and at least one expressly identifies the employment practices line, citing *Mobley*, as where this exposure lands.

[24] 42 U.S.C. §12112(d)(2)(A). The prohibition is categorical at the pre-offer stage and contains no exception for applicant consent.

[25] *Karraker v. Rent-A-Center, Inc.*, 411 F.3d 831, 834–837 (7th Cir. 2005) (holding that the Minnesota Multiphasic Personality Inventory was a medical examination because designed in part to reveal mental illness; that a test designed to identify a mental disorder or impairment is a medical examination while one that measures personality traits such as honesty, preferences, and habits is not; and that neither the scoring protocol nor the absence of a psychologist's interpretation was dispositive). The court did not reach whether the instrument was job related and consistent with business necessity, because that argument was waived. *Id.* at 835.

2. Failure to accommodate

The second track requires neither intent nor a comparator. A failure to accommodate claim asks whether the employer provided the reasonable accommodation or accessible alternative that would give the applicant meaningful access to the selection process, and it does not depend on proof that the employer intended to discriminate or on the existence of a similarly situated nondisabled applicant who was treated better.[26] Where a disabled applicant is screened out by an automated tool before any accommodation is offered or any accessible alternative is made available, the violation is the failure at that step. A reconstructable file that shows the screen operating exactly as designed simply documents the step that was skipped. Reconstruction is not accommodation.

3. Disparate impact and business necessity

The third track is the one that inverts the governance market's premise. A selection device that produces a disparate impact on individuals with disabilities is unlawful unless the employer carries the burden of showing that the device is job related and consistent with business necessity, and the employer carries that burden by validation, that is, by showing through professionally acceptable methods that the device predicts or significantly correlates with the requirements of the job.[27] The burden is affirmative, statistical, and prospective. It is a showing about the tool, made in advance and validated to the job, not a narrative about a particular decision offered after the fact. This is the pivot: the employer must produce something the governance file does not contain and that an unexplainable model cannot generate.

A qualification is owed to the reader. The Supreme Court has not squarely held that Section 504 reaches all disparate impact. In *Choate* the Court assumed without deciding that Section 504 reaches at least some conduct with an unjustifiable disparate impact on people with disabilities, while rejecting the broader proposition that every disparate impact showing states a prima facie case.[28] The firmer ground is the meaningful access and reasonable accommodation obligation, which does not depend on resolving that question. The validation point is best deployed as a reinforcement of the meaningful access argument rather than as an independent load bearing claim, and this Article treats it that way.

---

[26] On the reasonable accommodation obligation and the requirement of meaningful access under Section 504, see *Alexander v. Choate*, 469 U.S. 287, 301 (1985) (an otherwise qualified individual with a disability must be provided meaningful access to the benefit the grantee offers, and reasonable accommodations in the program may be required to assure it). On qualification standards and selection criteria that screen out individuals with disabilities, see 42 U.S.C. §12112(b)(6); 29 U.S.C. §794.

[27] *Albemarle Paper Co. v. Moody*, 422 U.S. 405, 425–432 (1975) (an employer using a selection device with a disparate impact bears the burden of showing, by professionally acceptable methods, that the device is predictive of or significantly correlated with important elements of the job); *see also Griggs v. Duke Power Co.*, 401 U.S. 424, 431–432 (1971) (the effects test); 29 C.F.R. pt. 1607 (Uniform Guidelines on Employee Selection Procedures). The disparate impact framework and its allocation of the business necessity burden to the employer were codified in the Civil Rights Act of 1991. *See* 42 U.S.C. §2000e-2(k). *Albemarle*, *Griggs*, and the Uniform Guidelines arise under Title VII, and this Article relies on them for the validation principle that the disability statutes incorporate through the business necessity defense; the reasonable accommodation and meaningful access obligations are distinct and are supplied by *Choate* and the Act itself.

[28] *Choate*, 469 U.S. at 297–299 (rejecting the boundless notion that all disparate impact showings state prima facie cases under Section 504, while assuming without deciding that Section 504 reaches at least some conduct with an unjustifiable disparate impact on individuals with disabilities). The firmer ground is therefore the meaningful access and reasonable accommodation holding, on which the disparate impact validation point rides as a reinforcement rather than the sole load path.

B. Opacity Indicts the Disability Defendant

The governance market treats explainability as a virtue that protects the deployer. On the disparate impact track it is the opposite. The model that cannot be explained is the model that cannot be validated, and the device that cannot be validated fails the business necessity defense as a matter of the burden allocation just described. The same opacity that defeats a disparate treatment plaintiff, who must prove intent and cannot reach into the black box to find it, sinks a disability defendant, who must affirmatively validate the box and cannot. The observation is not unique to the disability setting; commentators analyzing black box automated decision systems under a direct discrimination frame have reached the parallel conclusion that it is near impossible for a regulated decision maker to justify a system whose operation it cannot explain.[29]

The point is sharpened by the developer community's own research. The stated reasons a model gives for its outputs are not a reliable account of the computation that produced them, so an explanation generated by or about the model after the fact does not establish that the model's selections were job related.[30] The employer cannot cure a validation failure by asking the model to narrate its decision, because the narration is not evidence of the computation, and validation under *Albemarle* requires evidence of the computation's relationship to the job.

C. Human Review Does Not Launder the Tool

The employer's most common answer, that a human reviewed every recommendation, does not rescue the decision, and on close inspection it aggravates the exposure. A federal court has held that a provider of artificial intelligence based screening tools can be liable as the employer's agent under the disability and other anti discrimination statutes, and that the automated nature of the tool does not distinguish it from a human agent for that purpose; the focus is on the function delegated, not the manner in which the agent performs it.[31] The recruiter who clicks approve on a recommendation the employer cannot validate has adopted a screen for which no one has an independent, job related basis, and the tool rides into the liability chain as the employer's agent regardless.

A further answer must be closed off, because it is the one a defense will reach for. Suppose the employer's overall hiring numbers are balanced across disabled and nondisabled applicants; can it point to that balanced bottom line to defend the specific screen that excluded this applicant? It cannot. The Supreme Court has held that a balanced bottom line neither defeats a plaintiff's prima facie case nor supplies the employer a

[29] *See* Adams-Prassl, Binns & Kelly-Lyth, *supra* note 12, at 173–174 (observing that biased black box automated decision systems will be near impossible to justify). The frame there is European direct discrimination law; the structural point, that inexplicability defeats justification, transfers to the business necessity defense.

[30] On the unreliability of model stated reasoning as an account of the underlying computation, see Miles Turpin et al., *Language Models Don't Always Say What They Think: Unfaithful Explanations in Chain-of-Thought Prompting*, arXiv:2305.04388 (2023); Tamera Lanham et al., *Measuring Faithfulness in Chain-of-Thought Reasoning*, arXiv:2307.13702 (2023). The doctrinal use here is narrow: a post hoc explanation, whether produced by a human reviewer or by the model, is not a validation study and does not carry the business necessity burden.

[31] *Mobley v. Workday, Inc.*, 740 F. Supp. 3d 796, 800–801, 808–809 (N.D. Cal. 2024) (delegation of traditional hiring functions to an automated agent is not distinguishable from delegation to a human agent for purposes of agency liability, and applicants may bring disparate impact claims). The decision resolved a motion to dismiss and accepted the agency theory at the pleading stage; it is not a merits determination, and the litigation remains pending.

defense, because the statutory obligation runs to the individual applicant and a balanced workforce cannot immunize an employer from liability for a specific discriminatory barrier.[32] Human review of an unvalidated output is not validation; it is a second actor ratifying the same unvalidated output, and pointing to a balanced downstream result does not answer the barrier the applicant actually hit.

IV. The Loop: The Coverage Artifact Is the Liability Exhibit

The two preceding Parts can now be joined. Part II established that the employment practices insurer, where it has not excluded artificial intelligence outright, conditions coverage on the production of specific documentation. Part III established the elements of a disability claim arising from an automated screen. The loop is the observation that the documentation the insurer demands and the proof the plaintiff needs are the same documentation.

A. What the Underwriter Demands Is What the Plaintiff Needs

Take the underwriter's questions one at a time and set each against the disability element it supplies.

The underwriter asks whether the insured conducted bias testing on the tool. Bias testing is the measurement of the tool's disparate impact across protected groups. A bias test that found a disparate impact on individuals with disabilities is the plaintiff's prima facie case, produced and retained by the defendant at the insurer's instruction. A bias test that was demanded and not performed leaves the insured without the documentation the carrier requires, which is the coverage problem, and without a validation record, which is the liability problem.

The underwriter asks whether human review exists in the process. Documented human review is the adopted recommendation. It answers the deflection that the machine made the decision by confirming that a human ratified the machine's output, and it anchors the agency theory under which the vendor's tool is the employer's agent.[33] The presence of human review, offered to the insurer as a control, is offered to the plaintiff as proof that a human with no independent basis adopted an unvalidated screen.

The underwriter asks, in substance, whether the tool works, whether it was validated and is job related. That is the exact question the business necessity defense poses, and the exact showing *Albemarle* places on the employer.[34] If the insured can answer it, the answer is the validation study, which is discoverable and which the plaintiff would otherwise have had to extract. If the insured cannot answer it, the tool fails business necessity and the coverage is impaired at the same time.

The pattern is symmetric and inescapable. Every artifact the insurer requires to grant or preserve coverage is an artifact that advances the disability claim. The file that opens the policy is the file that proves the wrong.

[32] *Connecticut v. Teal*, 457 U.S. 440, 442, 450–456 (1982) (the bottom line does not preclude a prima facie case or provide a defense; a racially balanced workforce cannot immunize an employer from liability for specific acts of discrimination). *Teal* arises under Title VII; the principle transfers through the disability statutes' incorporation of disparate impact analysis. The point here is that an aggregate, downstream, humanly supervised outcome does not answer a challenge to the specific automated barrier that screened out the applicant.

[33] See *Mobley*, 740 F. Supp. 3d at 808–809.

[34] *Albemarle*, 422 U.S. at 425.

This is not a coincidence of a particular policy form. It follows from the fact that both the insurer and the plaintiff are asking the same underlying question, whether the tool is a valid, accessible, supervised selection device, and that question has one answer that lives in one set of documents.

B. The Record Is Overdetermined by Three Converging Regimes

The insurance condition would be enough to create the loop by itself, but it does not stand alone, and the convergence is what makes the deployer's position acute. Three regimes now demand the same file.

The first is the civil rights recordkeeping regime. The Uniform Guidelines on Employee Selection Procedures already require covered employers to maintain records disclosing the adverse impact of their selection procedures on protected groups, and they set the validation standards a selection device must meet.[35] The second is a state and local audit regime that in places goes further and requires publication. A New York City ordinance requires a bias audit of an automated employment decision tool and requires a summary of the results to be made publicly available, and a growing patchwork of state statutes regulates automated hiring tools in adjacent ways.[36] The third is the insurance regime described in Part II, which conditions coverage on the same bias testing, human review, and validation documentation.

The three regimes point at one record. A deployer complying with all of them assembles, retains, in places publishes, and hands to its insurer the very documents that establish the disability plaintiff's case. The insurance condition is the newest of the three, and it is the most revealing, because it is contractual and self interested: the insurer is not a regulator pursuing the public interest but a counterparty protecting its own book, and it has independently concluded that it needs the same file the plaintiff needs. When the party trying to limit its own exposure demands the identical documentation as the party trying to establish liability, the documentation is not a defense of the decision. It is a description of it.

C. The Remedy the Litigation Seeks Was Never Insurable

There is a further asymmetry that the loop exposes. The remedy that a systemic disability challenge to an automated screen actually seeks is not primarily damages. It is injunctive relief that changes the design of the system for everyone it screens, the kind of classwide relief available under Federal Rule of Civil Procedure 23(b)(2).[37] No liability policy pays a court order to redesign a system. Injunctive relief is not a covered loss; it is a change to the insured's operations. So the remedy that matters in this domain was outside insurability before any artificial intelligence exclusion was written. The exclusions and the documentation conditions now withdraw the damages coverage on top of a remedy that insurance never reached. The employer is exposed to

[35] 29 C.F.R. §1607.4(A), (B); *id.* §1607.5; *id.* §1607.14 (validation standards). The recordkeeping and validation obligations predate the current controversy and attach to any selection procedure with an adverse impact.

[36] N.Y.C. Local Law 144 of 2021, eff. July 5, 2023; *see* Forrest, *supra* note 8, at 21 (describing Local Law 144's public availability requirement and cataloguing state measures, including the Illinois Artificial Intelligence Video Interview Act, 820 Ill. Comp. Stat. 42/1, and proposed legislation in several states). Where a bias audit must be published, the disparate impact statistic is disclosed by operation of law rather than extracted in discovery.

[37] On the availability of classwide injunctive relief where the party opposing the class has acted on grounds generally applicable to the class, see Fed. R. Civ. P. 23(b)(2). The point here is structural: the design change that a systemic challenge seeks is an equitable remedy, not a sum of money.

the operational remedy in full and, increasingly, to the damages remedy without a payer.

V. The Privilege Escape Hatch Fails

The sophisticated deployer will see the loop and reach for the obvious tool. Run the bias audit under legal privilege, the argument goes, so that the very testing the insurer wants documented never becomes an exhibit for the plaintiff. This Part explains why that move fails. It fails not because privilege is unavailable in the abstract, but because the structure of the insurance regime forces the deployer into the two acts that waive it, and because the material at the center of the loop is not the kind of material the privilege reaches in any event.

A. The Self-Critical Analysis Privilege Is Not a Reliable Shield

The privilege the deployer is reaching for is usually the self critical analysis privilege, also called the self evaluative privilege, which some courts have recognized to protect candid internal reviews from discovery on the theory that disclosure would deter socially useful self assessment.[38] Its recognition is not uniform, courts are split on whether it exists and in what contexts, and it has fared especially poorly where it is asserted to shield discrimination related materials. A federal court applying it to an employer's internal audit in an employment case declined to do so, doubting the privilege's validity, finding that the employer had not shown that discovery would deter such reviews, and noting that the concern was weakest where the audit was prompted by pending litigation; the same court observed that the Supreme Court counsels strongly against the recognition of new privileges.[39] The Supreme Court's refusal in *University of Pennsylvania* to create a new privilege for confidential self evaluative materials in a discrimination investigation, and its concern that such a privilege would arm employers seeking to delay enforcement, signals the direction a court is likely to take when the self critical analysis privilege is asserted to bury a bias audit.

B. The Insurance Regime Forces the Two Classic Waivers

The deeper problem is waiver, and it is structural. A self evaluative privilege, where it exists at all, is waived by voluntarily disclosing the protected material, and it is waived by relying on the protected material as a defense.[40] The insurance regime forces the deployer into both. To satisfy the underwriter and unlock or preserve coverage, the deployer must disclose the substance of the bias testing and human review, which is

[38]The privilege is traced to *Bredice v. Doctors Hospital, Inc.*, 50 F.R.D. 249 (D.D.C. 1970) (protecting confidential medical staff review materials to encourage candid self assessment).

[39]*Deel v. Bank of Am., N.A.*, 227 F.R.D. 456, 459–460 (W.D. Va. 2005) (declining to apply the self critical analysis privilege to an employer's internal payroll audit, doubting its validity, and citing *Jaffee v. Redmond*, 518 U.S. 1 (1996), for the reluctance to recognize new privileges). *See also Univ. of Pa. v. EEOC*, 493 U.S. 182, 189–193 (1990) (declining to create a common law privilege against disclosure of confidential peer review materials in a Title VII investigation, and warning against placing a potent weapon in the hands of employers who wish to frustrate enforcement).

[40]On waiver by relying on an internal investigation as an affirmative defense, see *Volpe v. US Airways, Inc.*, 184 F.R.D. 672, 673–674 (M.D. Fla. 1998) (holding that by relying on its internal investigation as a defense the defendant waived any privilege over the investigation materials, and that a plaintiff is entitled to test the bona fides of an investigation the defendant relies upon). On waiver by voluntary disclosure of an audit report, see *Lawndale Restoration Ltd. P'ship v. Acordia of Ill., Inc.*, 853 N.E.2d 791 (Ill. App. Ct. 2006) (disclosure of the audit to the state insurance department waived the privilege under the Illinois Insurance Compliance Self-Evaluative Privilege Act).

voluntary disclosure. And to defend the discrimination claim on the merits, whether by asserting business necessity, reasonable oversight, or good faith, the deployer must rely on that same testing and review, which is reliance as a defense. The two moves the insurance regime compels are the two moves that waive the privilege. The deployer cannot occupy the position it needs, privileged for the plaintiff but disclosed for the insurer, because the disclosure to the insurer is itself a waiver as against the plaintiff.

C. The Underlying Data Is Not Privileged in Any Event

Even if the deployer routes the audit through counsel and claims attorney work product or attorney client protection for the analysis, the protection does not reach the material that matters to the loop. The Supreme Court has held that the attorney client privilege protects communications, not the underlying facts; a party cannot immunize a fact from discovery by the expedient of communicating it to a lawyer.[41] The disparate impact of a selection tool is a fact about how the tool behaves, not a communication with counsel. It surfaces in discovery whether or not a lawyer supervised the test, and it is the plaintiff's prima facie case. Worse for the deployer, structuring the audit to look like a routine business or compliance exercise, which is what an audit run to satisfy an insurer's questionnaire will resemble, can defeat even the attorney client privilege, because a court may find that the participants were not made sufficiently aware that the exercise was for the purpose of obtaining legal advice.[42]

D. The Theoretical Escape Collapses

One narrow escape remains to be closed. Suppose the deployer runs the audit under privilege, declines to disclose its substance to the insurer, attesting only that testing occurred, and declines to rely on it as a defense. In principle that sequence avoids both waivers. In practice it forfeits the very things it was meant to preserve. The underwriter's questionnaire, as reported, demands substantive documentation of bias testing and human review, not a bare attestation, and the consequence of failing to supply it is the sublimit or denial described in Part II.[43] So withholding the substance from the insurer forfeits the coverage that was the point of the exercise. And declining to rely on the testing as a defense forfeits the business necessity and reasonable oversight arguments that were the employer's best answers on the merits. The pincer holds in both directions: to keep the proof from the plaintiff, the deployer must give up either the coverage or the defense.

E. The One Genuine Safe Harbor, and Why It Does Not Reach Here

Honesty requires noting the one place a self evaluative privilege has real statutory footing. A number of states have enacted insurance compliance self evaluative audit privileges that protect an insurer's own voluntary

[41] *Upjohn Co. v. United States*, 449 U.S. 383, 395–396 (1981) (the privilege protects disclosure of communications, not disclosure of the underlying facts by those who communicated with the attorney). The mental impressions of counsel may separately be protected as opinion work product, *id.* at 399–402, but that protection does not extend to the operational facts the audit measured.

[42] *See Deel*, 227 F.R.D. at 461–462 (holding that employee questionnaires in an internal audit were not protected by the attorney client privilege where the employer's own materials framed the exercise as a routine business review rather than clearly communicating a legal purpose, in contrast to *Upjohn*). An audit conducted to answer an underwriter's questionnaire is, on its face, a business and compliance exercise, which is the posture that defeated the privilege claim in *Deel*.

[43] See *supra* note 23.

compliance audits from discovery, sometimes providing that disclosure to a regulator does not waive the privilege as to other parties.[44] Those statutes do not reach the deployer's problem for three reasons. They protect the insurer's audits, not an employer's bias testing of a hiring tool. They are creatures of particular states rather than a federal rule. And they govern insurance regulatory compliance, not a federal disability discrimination claim brought by a rejected applicant, which is where the loop bites. The statutory safe harbor exists, but it is in a different room from the one the deployer is standing in.

VI. IMPLICATIONS

A. Insurance as a Private Regulator, and Documentation as Risk Conversion

The phenomenon this Article describes belongs to a recognized tradition. A sociolegal literature has long argued that insurers govern the conduct of those they insure, managing moral hazard through conditionality, incentives, limits, and exclusions, and in some settings acting as compliance managers who shape how organizations comply with law through policy language and risk management services.[45] The artificial intelligence questionnaire and the documentation conditioned sublimit are that governance function operating on algorithmic hiring. The employment practices insurer has become, in effect, a private regulator of workplace artificial intelligence, and the terms of its regulation are the production and retention of bias testing, human review, and validation records.

The consequence for the deployer is that governance documentation, sold as risk management, is not risk transfer in this domain. It is risk conversion. The record converts a diffuse, hard to prove exposure into a concrete, discoverable one, and it does so at the moment the coverage that would have paid the resulting judgment is being withdrawn or conditioned on producing that very record. The operative question a deployer should ask its governance advisor is not whether the file is complete. It is whether the advisor has explained that the file the insurer requires is the file the plaintiff will use, and that the coverage behind it is contracting. A deployer that documents diligently, insures through a policy carrying an artificial intelligence exclusion or a documentation conditioned sublimit, and litigates a systemic disability claim seeking injunctive relief may find that it has built the plaintiff's exhibit, retained the operational remedy in full, and kept little of the coverage.

B. A Perverse Equilibrium for Carriers and Regulators

The questionnaire conditioned sublimit design produces an incentive structure worth a regulator's attention. It rewards the insured that does not test, because untested is undocumented and undocumented avoids creating

[44] See, e.g., the Illinois Insurance Compliance Self-Evaluative Privilege Act, 215 ILCS 5/155.35, which protects an insurer's voluntary internal compliance audits and creates a limited nonwaiver, and comparable insurance compliance self evaluative privilege statutes in a number of other states. Even that statute is construed narrowly: in *Lawndale* the Illinois court held that disclosure of the audit to the state insurance department waived the privilege. 853 N.E.2d 791.

[45] *See* Shauhin A. Talesh, *Data Breach, Privacy, and Cyber Insurance: How Insurance Companies Act as "Compliance Managers" for Businesses*, 43 Law & Soc. Inquiry 417 (2018) (insurers act as compliance managers shaping organizational compliance through policy language, pricing, and risk management services); Baker & Shortland, *supra* note 18 (insurers manage moral hazard through conditionality, incentives, limits, and exclusions); Omri Ben-Shahar & Kyle D. Logue, *Outsourcing Regulation: How Insurance Reduces Moral Hazard*, 111 Mich. L. Rev. 197 (2012) (insurers as private regulators of safety).

the exhibit, at the cost of coverage and of the very safety testing the design was meant to encourage. And it penalizes the insured that does test and documents honestly, because the honest documentation both satisfies the carrier and equips the plaintiff. A coverage design that makes candid bias testing hazardous to the insured is in tension with the public interest in candid bias testing, and it mirrors the concern the Supreme Court voiced about privileges that deter self assessment, arriving at the same place by a private contractual route rather than an evidentiary one. The tension is sharpest where a jurisdiction has resolved it in the opposite direction by mandating that the audit be published, because there the law has already decided that the value of disclosure outweighs the deterrent effect, leaving the insurance condition to pile a coverage penalty on top of a disclosure the deployer cannot avoid.

C. A Predictable Discovery Map for Plaintiffs and Enforcement

For a plaintiff or an enforcement agency, the loop draws a discovery map. The renewal file, the artificial intelligence underwriting questionnaire and the insured's responses to it, and the bias testing, human review, and validation documentation the insured produced to obtain coverage are now predictable and high value targets, and they exist because the insurance market required them to exist. What discovery once had to pry loose, the insurance relationship now assembles and dates. Where a bias audit must also be published under local law, part of the prima facie case is available before a complaint is even filed.

D. Limits of the Argument

Several honest limits bound the claim. First, the exclusions are not universal, and strong arguments for coverage still exist where artificial intelligence was only associated with a claim; legacy occurrence based policies that predate the exclusions may still respond, and the addition of an exclusion to later policies can support an argument that earlier forms were broader.[46] The defensible claim is therefore that the coverage floor is contracting and that an employer that renewed employment practices coverage in 2026 with a new artificial intelligence exclusion is likely uninsured for an algorithmic hiring claim, not that no employer has any coverage. Second, the disparate impact reach of Section 504 is assumed rather than settled, so the validation point should ride on the meaningful access and reasonable accommodation obligations rather than stand alone.[47] Third, the leading agency decision is at the pleading stage and remains subject to further proceedings, so it should be cited for the theory it accepted rather than as a merits holding.[48] Fourth, the recordkeeping obligations have limits in the algorithmic context; the Uniform Guidelines contemplate sample based recordkeeping for very high volume, frequently administered selection procedures, and an employer may not need to preserve the exact model or comparison dataset, which narrows what discovery can reach.[49] Fifth, the sharpest empirical support for the employment practices questionnaire mechanism comes from broker and consultant reporting rather than a filed form or a reported denial; the named Berkley endorsement

[46] On the argument that the introduction of an artificial intelligence exclusion in later policies supports broader coverage under earlier occurrence based policies, see practitioner analysis of policy vintage and coverage under successive forms.

[47] *See Choate*, 469 U.S. at 297–299.

[48] *See Mobley*, 740 F. Supp. 3d 796.

[49] *See* 29 C.F.R. §1607.4(A) (contemplating records where the number of applicants is large and selection procedures are administered frequently).

and the standard form general liability exclusions are the firmest of the currently available anchors, and a documented sublimit or a produced employment practices questionnaire would strengthen the mechanism further. Sixth, both the affirmative artificial intelligence coverage market and the state statutory privileges vary, and a full treatment would map that variation. None of these limits touches the core observation, which is structural: where coverage is conditioned on documentation, the documentation is the plaintiff's proof.

VII. Conclusion

The organization in the opening scenario believed it had done the safe thing. It kept the record. The advisory market told it that a complete, humanly supervised file would let it defend the decision, and the insurer told it that coverage depended on producing that file. Disability doctrine tells a different story. The record cannot defend the decision, because on the medical examination track the violation was complete before the record began, on the accommodation track the record documents the step that was skipped, and on the disparate impact track the record is either the missing validation or the disparate impact itself. And the record cannot be hidden, because the two acts that unlock the coverage, disclosure to the insurer and reliance on the controls, are the two acts that waive any privilege that might have buried it, and the disparate impact is a fact in any event.

The result is a closed loop. The remedy the litigation seeks was never insurable, because a court order to redesign a system is not a covered loss. The damages coverage that remained is now being excluded or conditioned. And the condition on what remains is the production of the documentation that proves the claim, documentation that a civil rights recordkeeping rule and, in places, a public audit mandate were already demanding. In this corner of the market, the record is not a defense of the decision. The record is the indictment, assembled at the insurer's request and dated by the renewal.