UNIVERSITY OF PITTSBURGH LAW REVIEW
Vol. 87 ● Spring 2026

HARMLESS AI ERROR

Abdi Aidid

ISSN 0041-9915 (print) 1942-8405 (online) ● DOI 10.5195/lawreview.2026.1167
http://lawreview.law.pitt.edu

This work is licensed under a Creative Commons Attribution-Noncommercial-No Derivative Works 3.0 United States License.

This journal is published by Pitt Open Library Publishing.

HARMLESS AI ERROR

Abdi Aidid*

ABSTRACT

The proliferation of artificial intelligence ("AI") in criminal proceedings has brought the technology's propensity to err into sharp focus. Whether AI is more or less accurate than human judgment is a favorite subject of recent scholarly debate, but what is certain is that AI methods generate particular kinds of errors that are distinct from those produced by humans. Indeed, the data science techniques that underlie many AI applications can involve processes so distinct from human reasoning and investigation—for instance, large language model ("LLM")'s practice of converting words to numbers and back to words—that the resulting errors can be difficult to anticipate or altogether detect.

This poses a special challenge for constitutional criminal procedure, and particularly the harmless error doctrine. Harmless error, which has been described by Landes and Posner as "probably the most cited rule in modern criminal appeals," provides that appellate courts can uphold criminal sentences even where trial courts made errors—including constitutional errors—so long as said errors would not be dispositive of the outcome. As AI-generated errors may not align with traditional notions of harmlessness, they complicate courts' ability to assess the impact on defendants' rights. Moreover, the technological valence of these errors may obscure constitutional violations. Courts thus need more particularized standards for evaluating the harms associated with AI error.

This Article takes the first step towards addressing this challenge by proposing a procedural adjustment: the creation of an "AI error threshold" to distinguish errors amenable to traditional harmless error analysis from those demanding heightened scrutiny akin to structural constitutional errors. By recognizing arbitrariness as a doctrinal fault line, courts can better uphold defendants' rights amidst technological evolution, ensuring fair adjudication in the AI age.

---

* Assistant Professor and Canada Research Chair in Artificial Intelligence and Access to Justice, Henry N.R. Jackman Faculty of Law, University of Toronto.

ISSN 0041-9915 (print) 1942-8405 (online) ● DOI 10.5195/lawreview.2026.1167
http://lawreview.law.pitt.edu

INTRODUCTION

Artificial intelligence's ("AI") propensity to err has been a public fixation, to put it mildly.[1] Several surveys about AI trustworthiness have found its relative "unreliability"—often code for its capacity to make mistakes—to be a leading reason for AI skepticism.[2] Popular discussions of AI errors received a meaningful boost in November of 2022 when the large language model ChatGPT became available for public use.[3] Its developers, the Microsoft-backed outfit OpenAI, boasted that ChatGPT acquired one million users in its first five days.[4] By December of 2024, ChatGPT had 300 million active users—*per week*.[5] For all the interest in the application and its competitors,[6] news media and the public at large were also seized

[1] A search on Google News for articles reveals the following sampled articles in just the first three weeks of February of 2025: Sara Merken, *AI 'Hallucinations' in Court Papers Spell Trouble for Lawyers*, REUTERS (Feb. 18, 2025, 3:55 PM), https://www.reuters.com/technology/artificial-intelligence/ai-hallucinations-court-papers-spell-trouble-lawyers-2025-02-18/ [https://perma.cc/A76L-N8YS]; Nicola Jones, *AI Hallucinations Can't Be Stopped—But These Techniques Can Limit Their Damage*, NATURE (Jan. 21, 2025), https://www.nature.com/articles/d41586-025-00068-5 [https://perma.cc/7X8S-JLC6]; Ben Fritz, *Why Do AI Chatbots Have Such a Hard Time Admitting 'I Don't Know'?*, WALL ST. J. (Feb. 11, 2025, 5:30 AM), https://www.wsj.com/tech/ai/ai-halluciation-answers-i-dont-know-738bde07 [https://perma.cc/366M-ZNZU]; Belle Lin, *Why Amazon is Betting on 'Automated Reasoning' to Reduce AI's Hallucinations*, WALL ST. J. (Feb. 5, 2025, 7:00 AM); https://www.wsj.com/articles/why-amazon-is-betting-on-automated-reasoning-to-reduce-ais-hallucinations-b838849e [https://perma.cc/Z8L9-7T5W].

[2] *See, e.g.*, Eleanor Busby, *Two in Three Teachers Say AI is 'Too Unreliable' to Assess Pupils' Work—Survey*, INDEP. (Mar. 19, 2024, 20:01), https://www.the-independent.com/news/uk/politics/chatgpt-department-for-education-university-of-exeter-dfe-yougov-b2515361.html [https://perma.cc/G5CD-FRJK]; *Public's Confidence in Its Ability to Evaluate AI-Generated Text Cause for Concern, Says Com Researcher*, BOS. UNIV. COLL. OF COMMC'N (Mar. 25, 2024), https://www.bu.edu/com/articles/publics-confidence-in-its-ability-to-evaluate-ai-generated-text-cause-for-concern-says-com-researcher/ [https://perma.cc/LRV5-ACK5].

[3] Bernard Marr, *A Short History of ChatGPT: How We Got to Where We Are Today*, FORBES (May 19, 2023, 1:14 AM), https://www.forbes.com/sites/bernardmarr/2023/05/19/a-short-history-of-chatgpt-how-we-got-to-where-we-are-today/ [https://perma.cc/U26M-2DYV].

[4] Cheyenne DeVon, *On ChatGPT's One-Year Anniversary, It Has More Than 1.7 Billion Users—Here's What It May Do Next*, CNBC (Nov. 30, 2023, 5:03 PM), https://www.cnbc.com/2023/11/30/chatgpts-one-year-anniversary-how-the-viral-ai-chatbot-has-changed.html [https://perma.cc/64XL-HSV4].

[5] Emma Roth, *ChatGPT Now Has Over 300 Million Weekly Users*, THE VERGE (Dec. 4, 2024, 11:39 AM), https://www.theverge.com/2024/12/4/24313097/chatgpt-300-million-weekly-users [https://perma.cc/FX7C-68YY].

[6] It is easy to think of ChatGPT as the only publicly available large language model, but Google's Gemini and Anthropic's Claude, as examples, also boast sizeable user bases. *See* Miles Kruppa, *Google's Resolution for 2025: Catch Up to ChatGPT*, WALL ST. J. (Jan. 16, 2025, 5:30 AM), https://www.wsj.com/tech/ai/google-gemini-2025-chatgpt-openai-b6eb595d [https://perma.cc/MAY8-EVRN]; Kyle Wiggers & Maxwell Zeff, *Claude: Everything You Need to Know About Anthropic's AI*, TECHCRUNCH (Feb. 25, 2025, 5:23 PM), https://techcrunch.com/2024/10/19/claude-everything-you-need-to-know-about-anthropics-ai/ [https://perma.cc/9R3F-T2JU].

ISSN 0041-9915 (print) 1942-8405 (online) ● DOI 10.5195/lawreview.2026.1167
http://lawreview.law.pitt.edu

by the phenomenon of *hallucinations*—a new term in the public lexicon denoting the AI's tendency to provide false or misleading information.[7]

Popular coverage of hallucinations has oscillated between earnest curiosity about its causes (no one is quite sure)[8] to stories speculating about what it might mean for future AI adoption (no one is quite sure about that, either).[9] But what makes the specter of these hallucinations such compelling public fodder is not simply that AI is sometimes mistaken, but rather that the AI is mistaken in *ways that we do not fully recognize*.[10] Indeed, the data science techniques that underlie many AI applications can involve processes so distinct from human reasoning and investigation—for instance, large language models' practice of converting words to numbers and back to words, also known as "text vectorization"[11]—that the resulting errors can be difficult to anticipate or altogether detect. Even AI applications that do not hallucinate, like machine learning-enabled algorithms, may err at levels just beyond our reach, such as in obscure technical processes or in complex computations.

The widespread application of AI to criminal justice and adjudication raises the stakes here. Today, AI is used to predict crime (and therefore direct the allocation of

[7] Sebastian Farquhar, Jannik Kossen, Lorenz Kuhn & Yarin Gal, *Detecting Hallucinations in Large Language Models Using Semantic Entropy*, 630 NATURE 625, 625 (2024) (footnote omitted) ("Hallucinations are often defined as LLMs generating 'content that is nonsensical or unfaithful to the provided source content' but they have come to include a vast array of failures of faithfulness and factuality.").

[8] Dan Milmo, *Google Chief Warns AI Could Be Harmful if Deployed Wrongly*, THE GUARDIAN (Apr. 17, 2023, 6:43 AM), https://www.theguardian.com/technology/2023/apr/17/google-chief-ai-harmful-sundar-pichai [https://perma.cc/4VA5-WH9S] (quoting Google Chief Executive Officer Sundar Pichai) ("There is an aspect of this which we call, all of us in the field call it as a 'black box.' You know, you don't fully understand. And you can't quite tell why it said this, or why it got wrong.").

[9] *See generally* Bhaskar Chakravorti, *AI's Trust Problem*, HARV. BUS. REV. (May 3, 2024), https://hbr.org/2024/05/ais-trust-problem [https://perma.cc/3TEQ-JBTT].

[10] Yujie Sun, Dongfang Sheng, Zihan Zhou & Yifei Wu, *AI Hallucination: Towards a Comprehensive Classification of Distorted Information in Artificial Intelligence-Generated Content*, 11 HUMANS. & SOC. SCIS. COMMC'NS 1278, 1279 (2024).

[11] Mayowa Akinwande, Oluwaseyi Adeliyi & Toyyibat Yussuph, *Decoding AI and Human Authorship: Nuances Revealed Through NLP and Statistical Analysis*, 13 INT'L J. ON CYBERNETICS & INFORMATICS 85, 90 (2024).

ISSN 0041-9915 (print) 1942-8405 (online) ● DOI 10.5195/lawreview.2026.1167
http://lawreview.law.pitt.edu

police resources),[12] to evaluate evidence,[13] to litigate cases,[14] and increasingly to impose sentences.[15] Some of the most high-impact applications have garnered both scholarly attention and public ire, such as tools aimed at crime prevention through predictive policing (e.g., forecasting where crime is most likely to occur) or risk assessment tools aimed at imposing "optimally-deterrent" criminal sentences or bail, parole, and release conditions.[16] Even outside of these purpose-built applications, participants in criminal adjudications will use AI-enabled technologies for more general tasks like drafting and transcription.[17] The possibility of error, then, means significant prospective impact on defendants' rights—both procedural rights and substantial rights.

This poses a special challenge for constitutional criminal procedure and, in particular, the harmless error doctrine. Harmless error, which has been described by William M. Landes and Richard A. Posner as "probably the most cited rule in modern criminal appeals," provides that appellate courts can uphold criminal sentences even where trial courts made errors—including constitutional errors—so

[12] *See, e.g.*, Ibrahim Raji & Damilola Bartholomew Sholademi, *Predictive Policing: The Role of AI in Crime Prevention*, 13 INT'L J. COMPUT. APPLICATIONS TECH. & RSCH. 66, 67 (2024) (examining law enforcement applications of AI as a method of crime prevention).

[13] *See, e.g.*, Justin Jouvenal, *A Secret Algorithm Is Transforming DNA Evidence. This Defendant Could Be the First to Scrutinize It*, WASH. POST (July 13, 2021), https://www.washingtonpost.com/local/legal-issues/trueallele-software-dna-courts/2021/07/12/66d27c44-6c9d-11eb-9f80-3d7646ce1bc0_story.html [https://perma.cc/8G6H-TWBE].

[14] *See generally* Yonathan A. Arbel, *Judicial Economy in the Age of AI*, NAT'L CIV. JUST. INST. 12–13 (July 20, 2024), https://ncji.org/wp-content/uploads/2024/06/2024-NCJI-Judges-Forum-Judicial-Economy-in-the-Age-of-AI-Arbel.pdf [https://perma.cc/4WJY-7JJG].

[15] *See generally* Saul Levmore & Frank Fagan, *Competing Algorithms for Law: Sentencing, Admissions, and Employment*, 88 U. CHI. L. REV. 367, 368 (2021).

[16] For examples of serious scholarship explaining and critiquing predictive policing and predictive sentencing, see Sandra G. Mayson, *Bias In, Bias Out*, 128 YALE L.J. 2218, 2224–25 (2019); Richard A. Berk, *Artificial Intelligence, Predictive Policing, and Risk Assessment for Law Enforcement*, 4 ANN. REV. CRIM. 209, 210 (2021).

[17] There have been some high-profile examples of late involving celebrity defendants Michael Cohen and Prakazrel Michel. *See* Jonathan Stempel, *Michael Cohen Will Not Face Sanctions After Generating Fake Cases with AI*, REUTERS (Mar. 20, 2024, 3:00 PM), https://www.reuters.com/legal/michael-cohen-wont-face-sanctions-after-generating-fake-cases-with-ai-2024-03-20/ [https://perma.cc/96W9-YFFZ]; Lindsay Whitehurst, *Judge Rejects Claims that Generative AI Tanked Political Conspiracy Case Against Fugees Rapper Pras*, ASSOC. PRESS (Aug. 30, 2024, 9:06 PM), https://apnews.com/article/fugees-pras-michel-trial-conspiracy-987f6fe908d84a030f0f2018a64d4866 [https://perma.cc/T2AL-EK8A].

ISSN 0041-9915 (print) 1942-8405 (online) ● DOI 10.5195/lawreview.2026.1167
http://lawreview.law.pitt.edu

long as said errors would not be dispositive of the outcome.[18] As AI-generated errors may not align with traditional notions of harmlessness, they complicate courts' ability to assess the impact on defendants' rights. Moreover, the technological valence of these errors may obscure constitutional violations. Courts thus need more particularized standards for evaluating the harms associated with AI error. At the same time, the harmless error doctrine suggests some helpful lessons for how to absorb AI error into criminal adjudications; to the extent that it is a remedial doctrine, harmless error suggests that there is a path towards an error-tolerance that nevertheless purports to protect substantial rights.

This Article takes the first step towards making sense of AI error in the context of criminal adjudication, with a view to the future of the harmless error doctrine. In Section I, I consider how criminal courts currently absorb error in adjudication, canvassing the constitutional harmless error doctrine. After considering the doctrine's recent evolution, I observe that the normative and jurisprudential justifications for the doctrine—namely that error is inevitable and retrials are too costly—are themselves challenged by the advent of technologized criminal proceedings. In Section II, I discuss how AI is currently being used in criminal adjudication. Rather than take all AI applications as equally problematic from the standpoint of error, I offer a typology of AI applications—(1) predictive (2) generative, and (3) evaluative—that produce distinct types of errors. In other words, the impetus for using AI in criminal trials (e.g., speed, systematic error detection, etc.) undermines the case for affirming convictions that feature harmless error. In Section III, I align the various types of AI error with the harmless error doctrine, suggesting lessons for future criminal adjudication and criminal-procedural analysis.

I. THE SPECTER OF CONSTITUTIONAL ERROR

A. The Contours of Constitutional Harmless Error

Harmless error has been described recently by Epps as the "most frequently invoked doctrine in all criminal appeals."[19] In its simplest rendering, the doctrine allows appellate judges to uphold convictions in cases wherein errors at the trial level

[18] William M. Landes & Richard A. Posner, *Harmless Error*, 30 J. LEGAL STUD. 161, 161 (2001); *see also* Daniel Epps, *Harmless Errors and Substantial Rights*, 131 HARV. L. REV. 2117, 2119 & n.1 (2018) ("Harmless error is almost certainly the most frequently invoked doctrine in all criminal appeals.").

[19] Epps, *supra* note 18, at 2119.

ISSN 0041-9915 (print) 1942-8405 (online) ● DOI 10.5195/lawreview.2026.1167
http://lawreview.law.pitt.edu

would not have affected the ultimate outcome.[20] In practice, this means that the prosecution will argue that even if an error occurred below, reversal is nevertheless too extreme a remedy. For most of the twentieth century, harmless error review was exclusively concerned with garden-variety trial errors; in other words, procedural defects that tended not to implicate substantive rights.[21] In its landmark 1967 ruling in *Chapman v. California*, however, the Supreme Court extended the scope of harmless error review to also include *constitutional* error, effectively "impos[ing] on appellate courts an obligation to vacate or reverse criminal judgments marred by constitutional error unless the government demonstrates that the error was harmless beyond a reasonable doubt."[22] Note that by and large, harmless constitutional error governs errors of determination—errors that stem from judicial decisions made during the course of a trial or in its ultimate disposition.[23] These are not errors committed by parties, witnesses, or other actors in the legal process.[24] To the extent that the mistakes of others—such as a prosecutor engaging in unconstitutional questioning—are relevant, they only rise to the level of constitutional error when a court makes an incorrect determination that allows the violation to stand.[25] In this sense, harmless constitutional error is inextricably tied to judicial misjudgments; it does not extend to every misstep that may occur in a proceeding, but is instead concerned with the legal propriety of the court's own rulings.[26]

Courts have placed certain constitutional errors—deemed "structural" constitutional errors—beyond the scope of the doctrine and thus "subject to

[20] *See* Roger A. Fairfax, Jr., *Harmless Constitutional Error and the Institutional Significance of the Jury*, 76 FORDHAM L. REV. 2027, 2029 (2008) ("The doctrine of harmless error, which saves a flawed criminal conviction from reversal, generally permits a conviction to stand where the reviewing court believes the defect in the proceeding was harmless beyond a reasonable doubt.").

[21] *See* Charles J. Ogletree, Jr., Arizona v. Fulminante*: The Harm of Applying Harmless Error to Coerced Confessions*, 105 HARV. L. REV. 152, 157–58 (1991) (noting that *Chapman v. California* extended harmless error analysis to constitutional errors); *see also* Fairfax, *supra* note 20, at 2035 ("Indeed, the notion that an error implicating a federal constitutional right could be harmless was not even taken seriously until the 1960s, when a significant expansion in the rights of criminal defendants was underway.").

[22] John M. Greabe, *The Riddle of Harmless Error Revisited*, 54 HOUS. L. REV. 59, 59 (2016).

[23] *See, e.g.*, Chapman v. California, 386 U.S. 18 (1967); Arizona v. Fulminante, 499 U.S. 279 (1991).

[24] *See, e.g.*, *Chapman*, 386 U.S. at 25; *Fulminante*, 499 U.S. at 283–84.

[25] *See, e.g.*, United States v. Hastings, 461 U.S. 499, 499 (1983).

[26] *See, e.g.*, *Chapman*, 386 U.S. at 23–24 ("illegally admitting highly prejudicial evidence or comments," "[errors] in admitting plainly relevant evidence").

ISSN 0041-9915 (print) 1942-8405 (online) ● DOI 10.5195/lawreview.2026.1167
http://lawreview.law.pitt.edu

automatic reversal."[27] The leading case, *Arizona v. Fulminante*, ostensibly clarified *Chapman* as it announced this rule.[28] As Roger A. Fairfax, Jr. explains, convictions involving non-structural constitutional errors can lead to a situation wherein "the appellate court may both acknowledge the conviction was tainted with constitutional error and, at the same time, affirm the conviction."[29] The line between what constitutes a structural error and what constitutes an ordinary trial error that is therefore entitled to harmless error review is unclear. John M. Greabe describes the uncertainty and some recognized examples of structural error:

> The Court bifurcated the universe of constitutional error and suggested that most errors fall into a default category of "trial" errors—"error[s] which occur[] during the presentation of the case to the jury, and which may therefore be quantitatively assessed in the context of other evidence presented in order to determine whether [their] admission was harmless." But the Court also acknowledged a second category of "structural defects"—"defects in the constitution of the trial mechanism, which defy analysis by harmless-error standards." More recently, in *United States v. Gonzalez-Lopez*, the Court explained that structural defects "bear[] directly on the framework within which the trial proceeds." It then provided a list of errors that cause such structural defects: errors having the effect of denying the right to counsel, self-representation, a public trial, and an appropriate reasonable-doubt instruction.[30]

Indeed, where the line of demarcation between trial errors and structural errors actually sits is so unclear that the only sure bets are those cases that have been

---

[27] *See* Fairfax, *supra* note 20, at 2029 ("A criminal conviction infected by structural error is immune to harmless error review and, therefore, is subject to automatic reversal. Conversely, if the constitutional error of which the criminal defendant complains is not deemed structural, the appellate court may both acknowledge the conviction was tainted with constitutional error and, at the same time, affirm the conviction."); *see also* Carol S. Steiker, *Counter-Revolution in Constitutional Criminal Procedure? Two Audiences, Two Answers*, 94 MICH. L. REV. 2466, 2528 (1996) (footnote omitted) (citing *Chapman*, 386 U.S. at 23) ("[T]he *Chapman* Court noted that some constitutional errors are 'so basic to a fair trial that their infraction can never be treated as harmless error' and gave as examples the admission of a coerced confession, the denial of the right to counsel, and the denial of the right to trial before an impartial judge.").

[28] *See Fulminante*, 499 U.S. at 280 (citations omitted) ("The admission of [an involuntary] confession [in *Chapman* was] a 'trial error,' which occurs during a case's presentation to the trier of fact and may therefore be quantitatively assessed in the context of other evidence presented in order to determine whether its admission is harmless beyond a reasonable doubt. A trial error differs markedly from violations that are structural defects in the constitution of the trial mechanism and thus defy analysis by harmless-error standards.").

[29] Fairfax, *supra* note 20, at 2029; *see also Fulminante*, 499 U.S. at 279.

[30] Greabe, *supra* note 22, at 75 (alterations in original) (citations omitted).

ISSN 0041-9915 (print) 1942-8405 (online) ● DOI 10.5195/lawreview.2026.1167
http://lawreview.law.pitt.edu

previously recognized as structural errors. For instance: (1) a *complete* denial of the Sixth Amendment right to counsel;[31] (2) a related denial of the Sixth Amendment right to self-represent;[32] (3) racially discriminatory jury selection;[33] (4) a "constitutionally deficient reasonable-doubt instruction" violation of the Fifth and Fourteenth Amendments;[34] (5) suppressing one's right to a public trial pursuant to the First and Sixth Amendments;[35] or (6) a failure to guarantee the right to an "impartial" tribunal pursuant to the Fifth and Fourteenth Amendments.[36] To the extent that there have been examples that exist outside of the six aforementioned categories, they have not been particularly instructive of what the boundary between trial error and structural error might be. For instance, in the 2018 case *McCoy v. Louisiana*, the Supreme Court reversed a state appellate court finding against ineffective assistance of counsel in a case that involved a defense attorney admitting a client's guilt despite the client's desperate objections.[37] The Supreme Court, in finding that there *was* ineffective assistance of counsel, explained that violating a client's right to autonomy here "blocks the defendant's right to make a fundamental

[31] The right emanates from the oft-cited *Gideon v. Wainright*, 372 U.S. 335 (1963). But the Court has recognized that denial of this right can constitute structural error in *United States v. Gonzalez-Lopez*, 548 U.S. 140 (2006). *See id.* at 148–49 (citations omitted) ("The second class of constitutional error we called 'structural defects.' These 'defy analysis by "harmless-error" standards' because they 'affec[t] the framework within which the trial proceeds,' and are not 'simply an error in the trial process itself.' Such errors include the denial of counsel, [and] the denial of the right of self-representation . . . .").

[32] *See, e.g.*, McKaskle v. Wiggins, 465 U.S. 168, 198 n.6 (1984) (White, J., dissenting) ("As is the case when the trial court completely denies a defendant's right of self-representation, application of the result-oriented harmless-error standard to cases like this one, where the defendant was allowed to proceed pro se but the conduct of his appointed standby counsel inhibited his ability to do so, would result in the denigration of the right.").

[33] *See, e.g.*, Vasquez v. Hillery, 474 U.S. 254, 263–64 (1986) ("Like these fundamental flaws, which never have been thought harmless, discrimination in the grand jury undermines the structural integrity of the criminal tribunal itself, and is not amenable to harmless-error review.").

[34] Sullivan v. Louisiana, 508 U.S. 275, 285 (1993) (Rehnquist, J., concurring).

[35] *See, e.g.*, Presley v. Georgia, 558 U.S. 209 (2010) (reversing trial court for decision to close courtroom to public during voir dire); *see also* Waller v. Georgia, 467 U.S. 39, 49 n.9 (1984) (quoting People v. Jones, 47 N.Y.2d 409, 416 (1979) ("'The harmless error rule is no way to gauge the great, though intangible, societal loss that flows' from closing courthouse doors.").

[36] *See* Tumey v. Ohio, 273 U.S. 510, 535 (1927) (holding that a trial by a judge who had a pecuniary interest in the outcome and a motive to convict because of his political role was a denial of due process in violation of the Fourth and Fourteenth Amendments); *see also* Turner v. Louisiana, 379 U.S. 466, 471–74 (1965) (holding that a close and continuous association between a key witness and the jury deprived the parties of a right to an impartial jury as required by the Due Process Clause of the Fourteenth Amendment).

[37] McCoy v. Louisiana, 584 U.S 414, 414 (2018).

ISSN 0041-9915 (print) 1942-8405 (online) ● DOI 10.5195/lawreview.2026.1167
http://lawreview.law.pitt.edu

choice about his own defense."[38] Crucially, the Court found additional rationales for categorizing it as "structural," namely that it was the kind of rights-denial that would have squarely affected the jury's view of the defendant, given that it was quite literally an admission of guilt.[39] Like *Fulminante*, the Court's effort at clarifying sowed yet more confusion: In finding that the Sixth Amendment right to counsel *and* the right to self-representation were both implicated,[40] the Court did not clarify which of the two constituted structural constitutional error. As a result, it is not entirely clear whether reversal was mandated by the pre-fixed categories discussed earlier or if the Court was signaling its openness to structural constitutional errors beyond the pre-fixed categories.

Despite being relatively commonplace, harmless constitutional error remains a deeply contested concept.[41] Indeed, there is a rather substantial body of literature devoted to analyzing its myriad conceptual confusions. At least three distinct scholarly debates merit attention here. The first concerns the doctrine's ill-defined relationship to the harmless error rules in 28 U.S.C. Section 2111 and Federal Rule of Criminal Procedure 52(a).[42] Since shortly after *Chapman*, scholars and judges alike have remarked about the judicial doctrine's uneasy fit with these statutory

[38] *Id.* at 416.

[39] *See id.* at 427–28 (footnote omitted) (citation omitted) ("[A] counsel's admission of a client's guilt over the client's express objection is error structural in kind. . . . And the effects of the admission would be immeasurable, because a jury would almost certainly be swayed by a lawyer's concession of his client's guilt. McCoy must therefore be accorded a new trial without any need first to show prejudice."). For a helpful treatment of *McCoy*, see generally Nina Varsava, Judith Foo, Elizabeth Villarreal & David Walchak, *Allocating Authority Between Lawyers and Their Clients After* McCoy v. Louisiana, 23 NEW CRIM. L. REV. 170 (2020).

[40] *McCoy*, 584 U.S. at 417.

[41] *See* David R. Dow & James Rytting, *Can Constitutional Error Be Harmless?*, 2000 UTAH L. REV. 483, 483 (2001) ("[H]armless error doctrine is conceptually incoherent. We argue that in view of this incoherence, constitutional errors ought not to be subject to harmless error review. In short, *Chapman* should be overruled."). One of the more helpful debates in the literature is between Daniel Epps and John Greabe, who disagree about a particular proposal that Epps suggests, which is to reconceive of harmless error. *Compare* Epps, *supra* note 18, at 2119 (arguing that the harmless error rule is better understood as a substantive inquiry rather than a remedial one), *with* John M. Greabe, *Criminal Procedure Rights and Harmless Error: A Response to Professor Epps*, 118 COLUM. L. REV. ONLINE 118, 118–19 (2018) (pointing out the flaws in Epps's approach to reconceptualizing the harmless error rule). Elsewhere, scholars have described the doctrine as "incoherent."

[42] Greabe gives a helpful summary of the confusion: "The juridical status of the *Chapman* principle is an enigma, despite the efforts of Traynor and others to shine an illuminating light upon it. Although *Chapman* states that the principle derives from federal law, the Supreme Court has never specified its precise source." Greabe, *supra* note 22, at 61 (footnote omitted).

ISSN 0041-9915 (print) 1942-8405 (online) ● DOI 10.5195/lawreview.2026.1167
http://lawreview.law.pitt.edu

siblings.[43] Justice Traynor took the view that the doctrine as articulated in *Chapman* seemed to occupy the same space as Section 2111.[44] In practice, however, there are some apparent differences. As an initial matter, Section 2111 makes a more subtle command than the harmless error doctrine; it asks courts to "ignore" errors that do not implicate substantive—which I understand as "non-procedural"[45]—rights, but it does not require that they altogether discard judgments where said rights have been violated. In this sense, the harmless error doctrine makes a more robust guarantee. With respect to 52(a), the differences are even more pronounced. As federal procedural rules, 52(a) and Section 2111 were not explicitly meant to govern state court trials, whereas cases can originate in state courts and ultimately be subjected to *Chapman*'s constitutional harmless error review when those cases "go federal," so to speak.[46]

Second, and relatedly, there remains considerable disagreement as to the essential nature of the doctrine: Is it principally remedial, or is it a more substantive inquiry into constitutional rights? Scholars like J.S. Marks have argued the former, explaining that the doctrine emerged chiefly to correct a runaway trend towards form-over-substance, whereby non-critical procedural irregularities could potentially undermine otherwise accurate judgments.[47] Since *Chapman*, scholars

[43] *See, e.g.*, *id.* at 65 ("[T]he Supreme Court should jettison the *Chapman* principle and abandon the structural defect/trial error dichotomy in favor of a simplified, unitary, and transcontextual harmless-error test—reconceived as an elaboration of 28 U.S.C. § 2111."); ROGER J. TRAYNOR, THE RIDDLE OF HARMLESS ERROR 16–18, 48–49 (1970). *See generally* 28 U.S.C. § 2111 (1988); FED. R. CRIM. P. 52(a).

[44] TRAYNOR, *supra* note 43, at 42 ("The most plausible explanation is that the Court lost sight of the controlling federal statute. It is hardly plausible that the Court did not deem [the federal harmless-error statute] 'appropriate congressional action.'"). For a critique of Traynor on this point, see Craig Goldblatt, *Harmless Error as Constitutional Common Law: Congress's Power to Reverse* Arizona v. Fulminante, 60 U. CHI. L. REV. 985, 999 (1993) ("Traynor glosses over the interesting interpretive question presented by § 2111. The applicability of § 2111 to state deprivations of federal constitutional rights was by no means obvious and is far beyond what Congress likely intended when the statute was enacted. First, like the contemporary state harmless error statutes, § 2111 was designed to govern violations of the rules of evidence, not constitutional errors. In fact, until *Chapman* itself, it was not even clear that constitutional violations could ever be treated as harmless.").

[45] John Hart Ely, *The Irrepressible Myth of* Erie, 87 HARV. L. REV. 693, 725 (1974) ("Or the protection of the process may proceed at wholesale, as by keeping the size of the docket at a level consistent with giving those cases that are heard the attention they deserve. The most helpful way, it seems to me, of defining a substantive rule—or more particularly a substantive right, which is what the Act refers to—is as a right granted for one or more nonprocedural reasons, for some purpose or purposes not having to do with the fairness or efficiency of the litigation process.").

[46] *See* Goldblatt, *supra* note 44, at 999 ("[Section] 2111 was intended to govern the procedures in federal appellate courts, not to set out a harmless error standard to bind the states.").

[47] Jason S. Marks, *Harmless Constitutional Error: Fundamental Fairness and Constitutional Integrity*, 8 CRIM. JUST. 2, 3 (1993).

ISSN 0041-9915 (print) 1942-8405 (online) ● DOI 10.5195/lawreview.2026.1167
http://lawreview.law.pitt.edu

have largely rehearsed the view that harmless constitutional error ought to be considered a remedial doctrine.[48] The logic behind this is straightforward: Courts conducting harmless error review are not asking whether a constitutional violation occurred but are rather assessing whether the occurrence of a *past* constitutional violation was harmful. At the same time, some compelling analyses of harmless error characterize it as "constitutional common law," a term used by Henry Paul Monaghan to describe the "substructure of substantive, procedural, and remedial rules" that "draw[] their inspiration and authority from," but are not strictly compelled by, the Constitution.[49] In this sense, conceiving of harmless error as constitutional common law means implying a cause of action that enables courts to "exercise their jurisdiction in a fashion thought appropriate for the vindication of federal constitutional rights."[50] Here, harmless error operates more like *Bivens* than, say, a Fifth Amendment analysis.[51]

A third and perhaps most significant conceptual confusion of the harmless error doctrine relates to its apparent uneasy fit with rule-of-law values. A subset of the literature is intensely critical of harmless error. These critiques tend to coalesce around two related concerns: (1) that harmless error analyses deemphasize constitutional violations; and (2) that, in practice, harmless error review enables a systematic devaluation of constitutional rights. David R. Dow and James Rytting make the most forceful case, arguing that:

> When a court applies harmless error analysis in the constitutional context, that court is concluding that a constitutional right—a fundamental legal principle—

[48] *See* Greabe, *supra* note 41, at 119 (noting that there is a "conventional understanding of harmless error review as a *remedial* inquiry").

[49] Henry P. Monaghan, *The Supreme Court, 1974 Term—Foreword: Constitutional Common Law*, 89 HARV. L. REV. 1, 2–3 (1975) ("[I]t might be more readily recognized that a surprising amount of what passes as authoritative constitutional 'interpretation' is best understood as something of a quite different order—a substructure of substantive, procedural, and remedial rules drawing their inspiration and authority from, but not required by, various constitutional provisions; in short, a constitutional common law subject to amendment, modification, or even reversal by Congress.").

[50] Daniel J. Meltzer, *Harmless Error and Constitutional Remedies*, 61 U. CHI. L. REV. 1, 31 (1994).

[51] Epps, *supra* note 18, at 2149–50 ("Understanding harmless constitutional error as a form of constitutional common law means thinking of it as akin to the implied cause of action recognized in *Bivens v. Six Unknown Named Agents*."); *see also* Meltzer, *supra* note 50, at 29–30 (footnote omitted) ("I would like to draw particular attention to the link between *Chapman* and the *Bivens* line of cases. An appellant attacking an adverse judgment is seeking affirmative relief against a past wrong, not unlike a plaintiff suing the government for damages in a constitutional tort action. Thus, it seems highly plausible to analogize rules regulating the provision of relief to the appellant to rules authorizing the provision of damages relief in *Bivens* actions."); Richard H. Fallon, Jr. & Daniel J. Meltzer, *New Law, Non-Retroactivity, and Constitutional Remedies*, 104 HARV. L. REV. 1731, 1731, 1772, 1819 (1991).

ISSN 0041-9915 (print) 1942-8405 (online) ● DOI 10.5195/lawreview.2026.1167
http://lawreview.law.pitt.edu

> was violated by the state. Yet it disregards the violation of that principle, at least sometimes. . . . [A] court applying harmless error doctrine is thereby acting particularistically . . . . [I]t is saying that notwithstanding the violation of a principle, the right result was attained.[52]

In practice, cases involving constitutional violations can sometimes produce jarring results. For instance, in *United States v. Escamilla*, the Fifth Circuit considered an officer's warrantless search of a defendant's phone after an arrest.[53] The search yielded phone records, including 196 phone calls and twenty-nine text messages, linking the defendant to a vehicle carrying contraband.[54] The government did not argue that there was any exception to the Fourth Amendment's prohibition on unreasonable searches, but it nevertheless relied on the evidence at trial.[55] Unsurprisingly, the Fifth Circuit found that the search was unconstitutional and that any resulting evidence should have been suppressed by the district court.[56] Still, the Fifth Circuit deemed the error a harmless one. As it happened, the police officers had additional evidence of the same calls and text messages from another device, leading the Fifth Circuit to conclude that the improperly obtained phone records were superfluous *even if* those improperly obtained phone records led to the independent corroborating evidence.[57]

Escamilla is rather typical of constitutional harmless error cases and reveals two important characteristics of the doctrine. First, *Escamilla* exemplifies courts' relatively generous application of harmless error. Scholars and judges alike have lamented the seemingly low threshold to trigger harmless error review. Perhaps the most forceful declaration came from Judge Harry T. Edwards, the then-Chief Judge of the United States Court of Appeals for the District of Columbia Circuit, who wrote, "[a]ppellate judges merely apply a 'drop' of harmless error, and the coerced

---

[52] Dow & Rytting, *supra* note 41, at 533. For other related critiques, *see, e.g.*, James Edward Wicht III, *There Is No Such Thing as a Harmless Constitutional Error: Returning to a Rule of Automatic Reversal*, 12 BYU J. PUB. L. 73, 74 (1997) (footnote omitted) ("While the original harmless error rule provided a necessary remedy to a legitimate jurisprudential problem, the rule has evolved well beyond merely addressing the small technical defects for which it was originally intended. So much so that under current Supreme Court case law, even the wrongful admission of a coerced confession may be deemed a harmless error. This explosion of the harmless error doctrine goes beyond the original purpose of the rule and has taken Constitutional law as applied to criminal defendants down a road better left untraveled.").

[53] United States v. Escamilla, 852 F.3d 474, 477–78 (5th Cir. 2017).

[54] *Id.* at 479, 487.

[55] *Id.* at 485–87.

[56] *Id.*

[57] *Id.*

ISSN 0041-9915 (print) 1942-8405 (online) ● DOI 10.5195/lawreview.2026.1167
http://lawreview.law.pitt.edu

confession, warrantless search, erroneous jury instruction, faulty exclusion of evidence, unfair restriction on cross-examination, and a host of other errors simply vanish as though they never had occurred."[58]

Even a cursory glance at the data suggests some support for Judge Edwards's argument. In the thirteen years immediately following the Supreme Court's ruling in *Chapman*, Goldberg found that harmless error was an issue in determining "possibly as high as ten percent of all criminal appeals."[59] Landes and Posner's study found that in cases involving the doctrine, courts found that an error was harmless some eighty-seven percent of the time.[60] Importantly, courts did not always evaluate the error; Landes and Posner further observe that in forty-two percent of the cases, courts did not find it necessary to determine whether there was actual error, instead concluding that "even if there was an error," that prospective error was harmless.[61] The result is that a substantial subset of the case law universe does not give any meaningful guidance as to what constitutes harmless error.

Second, and relatedly, the fact of the error in *Escamilla* being constitutional in nature did not *necessarily* afford it any special attention. The Fifth Circuit's analysis

---

[58] Harry T. Edwards, *To Err Is Human, but Not Always Harmless: When Should Legal Error Be Tolerated?*, 70 N.Y.U. L. REV. 1167, 1170 (1995); *see also* James E. Boren & Michael A. Fiser, *Fear of a Paper Tiger: Enforcing Louisiana's Procedural and Statutory Rules in the Wake of Harmless Error Analysis*, 64 LA. L. REV. 5, 6 (2003) (affirming Judge Harry T. Edwards's critique and stating that "harmless error enables courts to uphold convictions in trials riddled with constitutional and statutory blunders"); Lisa Kern Griffin, *Criminal Adjudication, Error Correction, and Hindsight Blind Spots*, 73 WASH. & LEE L. REV. 165, 199–204 (2016) (explaining that "hindsight bias" has unfairly raised the standard for harmless error review).

[59] Steven H. Goldberg, *Harmless Error: Constitutional Sneak Thief*, 71 J. CRIM. L. & CRIMINOLOGY 421, 421 & n.3 (1980). Goldberg's analysis was accompanied by the following curious footnote 3, which I reproduce in full here: "The empirical research to establish the exact relationship between harmless constitutional error determinations and all criminal appeals is probably not worth the effort. The available data and a couple of assumptions make a fair argument for the likelihood of the 10 percent estimate. While the Shepard's citations undoubtedly include some cases of harmful error and some multiple citations in the same case, a number of cases are determined by a finding of harmless constitutional error without any mention of *Chapman*. Harrington v. California, 395 U.S. 250 (1969), provided a less rigorous test than *Chapman* and is often cited in preference to *Chapman*. Similarly, many state courts cite the state authority for harmless constitutional error rather than *Chapman*. It is not unreasonable to assume that the cases which do not cite *Chapman* for the harmless error proposition are at least as numerous as those that do and do not find the error harmless. Something in excess of two-thirds of the *Shepard*'s citations to *Chapman* are in state cases." *Id.*

[60] Landes & Posner, *supra* note 18, at 184 ("In 87 percent of the cases, the errors were held to be harmless—in 45 percent the appellate court found errors but held that they were harmless and in another 42 percent the court concluded that even if there was an error (which the court did not decide) it was harmless.").

[61] *Id.*

ISSN 0041-9915 (print) 1942-8405 (online) ● DOI 10.5195/lawreview.2026.1167
http://lawreview.law.pitt.edu

of the error was categorical; that is, whether it was an error mattered more than the quality or nature of the error. In fact, courts will rarely contend with the gravity of the constitutional violation as part of the harmless error review but instead will consider its prospective impact on the outcome. Put another way, there is no doctrinal reason why an extreme denial of a constitutional right should not be considered harmless if it was not dispositive of the outcome, unless it is structural in nature. In *Escamilla*, the fact that the evidence was obtained through an improper search pursuant to the Fourth Amendment did not alter what Fallon and Meltzer call the "balancing calculus."[62]

B. The Accuracy/Efficiency Trade-off vs. Error Tolerance

One underrated rationale for the harmless error doctrine relates to its "information-forcing" function. In an early version of this argument, Green posited that the doctrine of harmless error can incentivize greater judicial candor.[63] Judges, who may otherwise feel considerable pressure to affirm criminal convictions, could be more prone to acknowledge errors when the stakes are moderated by the harmless error standard. Essentially, this doctrine mitigates the fear that any error, regardless of its impact, could unravel an entire verdict, thus encouraging judges to more openly recognize and address mistakes. Consequently, this transparency theoretically reduces the informational costs associated with evaluating the overall fairness and accuracy of a trial, promoting a more honest and efficient judicial process.

This, however, is largely a byproduct of the doctrine's prevailing policy rationales: efficiency and finality.[64] Here, courts conceive of efficiency and finality in the specific sense of avoiding wasteful litigation. In other words, courts consider the harmless error doctrine against their inherent role as gatekeepers of the judicial economy. There is at least an implicit motivation that some modicum of non-

[62] Fallon & Meltzer, *supra* note 51, at 1772 ("The modern view that constitutional errors can be harmless does not rest upon a sure denial of prejudice, but emerges instead from a balancing calculus familiar to the law of remedies: if the risk of prejudice, though not nonexistent, is small, the burdens of retrial are not warranted.").

[63] Stuart P. Green, *The Challenge of Harmless Error*, 59 LA. L. REV. 1101, 1103 (1999) ("To require a judge to reverse in every case in which she finds error might well create a perverse incentive against finding error in the first place. In theory, the harmless error doctrine should free up appellate judges to consider the merits of the alleged violation on its own terms, before considering the ultimate disposition of the case. Even if a judge ultimately decides to affirm the conviction, she can preserve the integrity of the common law by writing a decision that fully explains the nature of the violation that occurred.").

[64] United States v. Hasting, 461 U.S. 499, 509 (1983) ("The goal [of harmless error review] is 'to conserve judicial resources by enabling appellate courts to cleanse the judicial process of prejudicial error without becoming mired in harmless error.'").

ISSN 0041-9915 (print) 1942-8405 (online) ● DOI 10.5195/lawreview.2026.1167
http://lawreview.law.pitt.edu

prejudicial error is preferable to being inundated with costly retrials that are highly likely to confirm the same outcome.

So, what's *really* going on here? One framework to evaluate the doctrine of harmless error's aims is to consider it in light of a common heuristic in legal analysis: the accuracy/efficiency trade-off.[65] Cooper, Levy & De Sa recognize the presence of the accuracy/efficiency trade-off in other areas of law:

> For two examples, the standard for preliminary injunctive relief in the United States explicitly considers whether irreparable injury will occur because of the passage of time, if relief is not granted before the (often lengthy) full resolution of a case. Federal Rule of Evidence 403 allows for the exclusion of relevant evidence [sic] its probative value is substantially outweighed by a danger . . . undue delay, wasting time, or needlessly presenting cumulative evidence.[66]

In the simplest rendering, the accuracy/efficiency trade-off recognizes that systems have objectives that are in tension.[67] In engineering, it might be between how "*correct* computations are and *how long* it takes to compute them."[68] In the context of harmless error, efficiency is a more expansive and variable notion. Here, think less of efficiency as speed and more of efficiency in the way the class action procedure is justified in civil adjudication—as a way not only to improve speed-to-resolution, but also to prevent courts from being inundated with claims and frustrating the docket.[69] Indeed, it resembles a "kitchen sink" analysis: At various

[65] Helpfully, given this topic, much of the clearest writing about accuracy and efficiency has been in the context of evaluating algorithms. *See, e.g.*, Hongche Liu, Tsai-Hong Hong, Martin Herman, Ted Camus & Rama Chellappa, *Accuracy vs. Efficiency Trade-Offs in Optical Flow Algorithms*, 72 COMPUT. VISION & IMAGE UNDERSTANDING 271 (1998).

[66] A. Feder Cooper, Karen Levy & Christopher De Sa, *Regulating Accuracy-Efficiency Trade-Offs in Distributed Machine Learning Systems*, SSRN (Mar. 10, 2021), https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3650497 [https://perma.cc/TLA3-YAE9].

[67] For a general understanding of the treatment of the accuracy/efficiency trade-off in a non-legal context, see generally Lennart Dabelow & Masahito Ueda, *Three Learning Stages and Accuracy-Efficiency Tradeoff of Restricted Boltzmann Machines*, NATURE COMMC'NS, Sept. 17, 2022, at 1.

[68] Cooper et al., *supra* note 66, at 1.

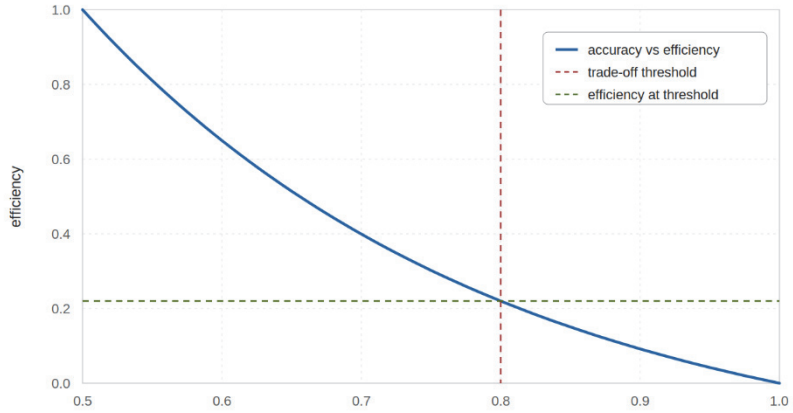
[69] *See generally* Hillel J. Bavli, *Aggregating for Accuracy: A Closer Look at Sampling and Accuracy in Class Action Litigation*, 14 LAW, PROBABILITY & RISK 67, 67 (2015) (noting efficiency benefits of certain class action techniques like sampling in claim aggregation).

ISSN 0041-9915 (print) 1942-8405 (online) ● DOI 10.5195/lawreview.2026.1167
http://lawreview.law.pitt.edu

times, any cost-raising considerations were considered germane to the analysis.[70] At bottom, however, the accuracy/efficiency trade-off involves the tension between the relative correctness of a judgment (which the doctrine suggests courts need not be absolutely committed to), and the adjudicative resources.[71]

Figure 1 below demonstrates a sketch of an accuracy/efficiency trade-off for illustrative purposes only:

Figure 1. Accuracy vs efficiency trade-off



The x-axis represents rising accuracy levels using arbitrary intervals. The y-axis represents rising levels of efficiency; in the context of criminal procedure, imagining adjudication that optimally conserves resources, disposes of matters relatively quickly, and is otherwise not wasteful. The blue curve reveals the interaction; note a decline in efficiency as accuracy rises. In this model, the most accurate outcomes take longer and demand the most resources. For instance, a

[70] *See* Brecht v. Abrahamson, 507 U.S. 619, 633–38 (1993) (explaining that the "additional deterrent effect, if any," of applying *Chapman* on federal habeas review is "outweighed by the costs of that application," which "undermines the States' interest in finality and infringes upon their sovereignty over criminal matters," is "at odds with habeas' purpose of affording relief only to those grievously wronged," and imposes significant "social costs," including delayed retrials, "the erosion of memory," and "the dispersion of witnesses").

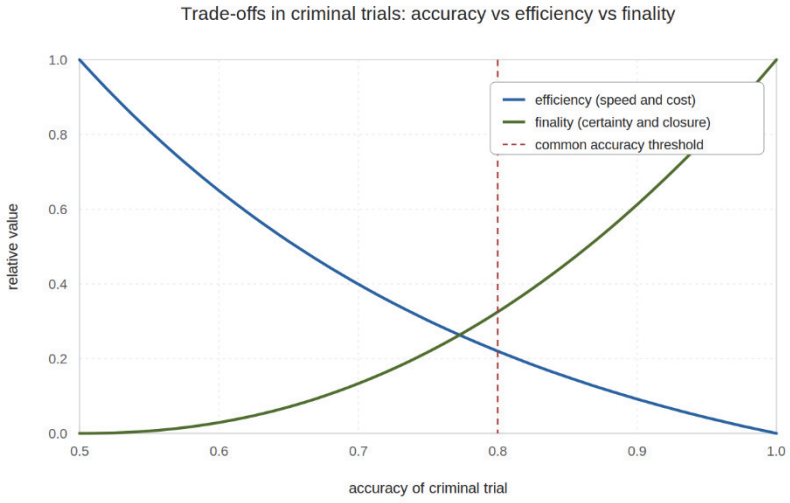
[71] Darryl K. Brown observes the way that harmless error creates incentives: "Here again we see that the harmless error standard, whatever other criticisms it merits, reinforces the right incentives for litigants for error prevention in the trial stage. It deprives parties of any reward for sandbagging insignificant errors at trial, and it grants a remedy only for harmful errors that the injured party is already highly motivated to prevent at trial. After all, for an error to be harmful, by hypothesis it must have been *effective* in (wrongly) improving the odds for the party that employed it—precisely the kind of error that the injured party would be eager to prevent at trial." Darryl K. Brown, *Does It Matter Who Objects? Rethinking the Burden to Prevent Errors in Criminal Process*, 98 TEX. L. REV. 101, 134–35 (2020) (footnote omitted).

ISSN 0041-9915 (print) 1942-8405 (online) ● DOI 10.5195/lawreview.2026.1167
http://lawreview.law.pitt.edu

judicial system that is absolutely committed to reversing *all* judgments where *any* substantial rights are implicated would have y-axis inefficiency: Some procedures would be duplicated, witnesses would be re-examined, and parties would expend resources to prepare for a second trial or appeal. The vertical red dash line marks a hypothetical balance in accuracy and efficiency—in other words, where courts deem a case *accurate enough*. Cases beyond this line would be the modal "harmless error review" cases, i.e., the cases wherein courts deem that constitutional rights were substantially guaranteed and residual errors are insufficient to justify the resource-intensive exercise of an intervention. Note that beyond the red line and below the horizontal green dash line, minor accuracy gains require considerably more resources than before the trade-off threshold. This mirrors the harmless error logic; part of the rationale for not overturning cases where errors are not dispositive of outcomes is that any incremental improvement in the accuracy of the adjudication comes at an outsized cost.

Figure 2 below considers the accuracy/efficiency trade-off once more, but this time with the finality of a given litigation in mind. Note that this is for illustrative purposes only (i.e., the inputs are arbitrary). Finality demands closer attention because the doctrine often attends to it either as a component part of its efficiency analysis or as a separate consideration.

Figure 2



Here, the x-axis once more represents the accuracy of a given adjudication. The y-axis this time graphs both the relative value of efficiency and finality. Observe the blue curve, which shows once more that trials can be more efficient when there is less demand for accuracy. Note the new green curve; here, the stylized accuracy/efficiency trade-off model suggests that low accuracy cases are less final.

ISSN 0041-9915 (print) 1942-8405 (online) ● DOI 10.5195/lawreview.2026.1167
http://lawreview.law.pitt.edu

This stands to reason as low accuracy cases would be more vulnerable to appeal; by contrast, high accuracy cases are more likely final because of their relative correctness.[72] The key insight follows. Note that at the red dash line, which once again represents an arbitrary common accuracy threshold, courts have achieved an acceptable balance of efficiency and accuracy. Once that line is exceeded in favor of more accuracy, efficiency becomes considerably more difficult to achieve, although expending those substantial incremental resources has a significant effect on a given case's finality.

The stylized model represented in Figures 1 and 2 above is only partially indicative of the dynamic underlying constitutional harmless error. Why? In practice, the harmless error doctrine depends on categorical distinctions that the trade-off does not fully capture. Recall that harmless error tolerates certain kinds of trial errors as long as they are non-dispositive, but altogether exempts structural constitutional errors from its analysis and mandates automatic reversal. This carve-out means that accuracy and efficiency where structural error is introduced is non-continuous. I suggest here that a modified but more appropriate framework is one of error tolerance; some errors are harmless, others are corrected for on appeal, whereas a subset is not tolerated at all. Here, I use the term *error tolerance* to refer more to the practice of absorbing benign error into legal system;[73] a more accurate rendering, then, moves beyond the stylized smooth curve of trade-offs in favor of a threshold-based rendering. Consider Figure 3 below.

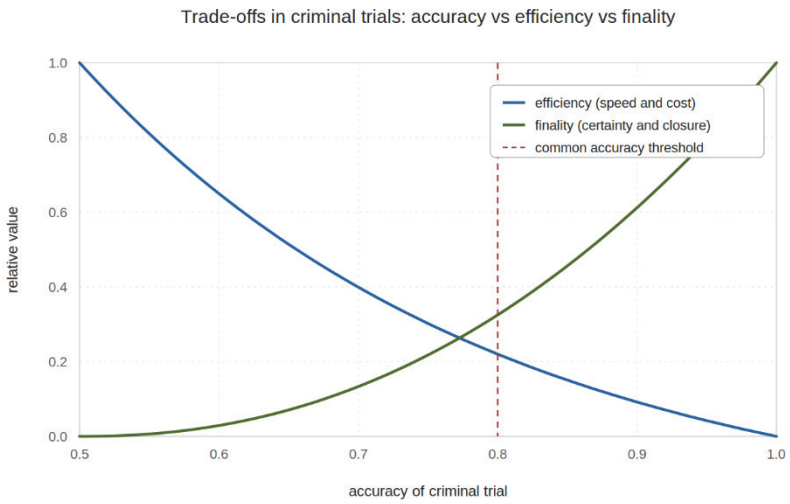
---

[72] *See generally* Brandon L. Garrett, *Accuracy in Sentencing*, 87 S. CAL. L. REV. 499 (2014). One of Garrett's most interesting insights is that inaccuracy can create a certain kind of duplication in the adjudicative process because "judges now have more discretion to sentence, and therefore an outright error might result in an appellate court requiring the judge to resentence by exercising discretion again, which may result in the same sentence." *Id.* at 503. For a more general treatment of the costs associated with accuracy in adjudication, see Louis Kaplow, *The Value of Accuracy in Adjudication: An Economic Analysis*, 23 J. LEGAL STUD. 307 (1994).

[73] The term "error tolerance" is influenced by its use in mathematics to broadly the same effect. *See, e.g.*, Regina Sandra Burachik & B.F. Svaiter, *A Relative Error Tolerance for a Family of Generalized Proximal Method Points*, 26 MATHEMATICS OPERATIONS RSCH. 816 (2001); Randy F. Hall, *Error Tolerance Formulas Related to Classical Algebra and Geometry*, 105 MATHEMATICS TCHR. 460 (2012). However, I would be remiss not to mention one of the most interesting treatments of the term, which comes in Joseph E. David's treatment of "toleration" within Jewish thought. David describes "tolerable error" in the context of Talmudic approaches as follows: (1) a "substantive approach" that deems all deviations "from the fundamentals" as illegitimate, (2) a "conventional approach" wherein "the tolerable–intolerable boundary is determined by cases involving contravention of religious ideas or behaviour that all" should share, and (3) a "scholastic approach" wherein "tolerable error is related to the broad rabbinic project of rearticulating the notion of divine law and identifying it with the teachings of the sages." The rich and broad use of the concept is an advantage. Joseph E. David, *The Notion of Tolerable Error from the Mishnah to Maimonides*, *in* TOLERATION WITHIN JUDAISM 57, 59–60 (2013).

ISSN 0041-9915 (print) 1942-8405 (online) ● DOI 10.5195/lawreview.2026.1167
http://lawreview.law.pitt.edu

Figure 3



Here, the y-axis represents an arbitrary scale of error tolerance. The x-axis represents accuracy (as in Figures 1 and 2 above). The orange line reveals the key insight that at a certain point along the accuracy scale, errors become outcome-determinative (what I term here the "dispositivity threshold"). The purple line represents error tolerance as a function of accuracy. Simply put, trials can tolerate errors up until the dispositivity threshold, at which point courts become intolerant. But note the purple line, which graphs error tolerance. In the leftmost section of the graph, the purple line is flat, denoting that errors are tolerated before the dispositivity threshold, meaning convictions will generally stand. In the middle section, where harmless error review is applied, note that the purple line drops somewhat sharply: cases here *may* be reversed depending on whether the errors are determinative of the outcomes. In other words, the doctrine's tensions are evaluated in between the orange and red lines. A much sharper decline occurs once the error passes the red dashed line, indicating that it is a structural constitutional error. Here, the court becomes intolerant of errors and counsels automatic reversal under *Chapman*'s rubric.[74]

[74] *See* Chapman v. California, 386 U.S. 18, 23 (1967) ("[S]ome constitutional rights [are] so basic to a fair trial that their infraction can never be treated as harmless error . . . ."); *see also* Meehan Rasch, *California's Dueling Harmless Error Standards: Approaches to Federal Constitutional Error in Civil Proceedings and Establishing the Proper Test for Dependency*, 35 W. ST. U. L. REV. 101, 110–11 (2008) ("[T]he *Chapman* Court concluded that the ideal approach to analyzing the harmlessness of identified federal constitutional error, while stopping short of reversal per se, should promote 'an intention not to treat as harmless those constitutional errors that "affect substantial rights" of a party.' To this end, the Court announced a more protective 'harmless-constitutional error rule . . . .").

ISSN 0041-9915 (print) 1942-8405 (online) ● DOI 10.5195/lawreview.2026.1167
http://lawreview.law.pitt.edu

The stylized model in Figure 3 shows that courts employ a flexible approach to reversal when an error is prospectively outcome-determinative but not in one of the pre-fixed structural error categories. So why describe this as categorical error tolerance and not a function of the interaction between accuracy and efficiency and finality? Consider Figure 3 once more. If this was a simple measure of accuracy on one hand and efficiency on the other, observers would reasonably anticipate a more gradual error-tolerance decline where there is more accuracy. In other words, courts would progressively demand more accuracy. Likewise, if this were a simple measure of accuracy and finality, observers would reasonably expect a gradual drop in error tolerance. Instead, Figure 3 is a more accurate rendering: Indeed, the doctrine does not counsel courts to *gradually* relax their preference for finality based on the nature of the constitutional violation, but to categorically override that preference in some pre-determined types of cases.[75]

Put differently, harmless error does not involve a relentless balancing of interests, nor does each case necessitate an individualized weighing of competing priorities, requiring courts to continuously assess whether procedural rigor must yield to considerations of efficiency. In contrast, the superior error-tolerance framework operates on a more categorical basis: The threshold does all the work. In this model, courts establish a definitive line between acceptable and unacceptable errors, creating a bright-line rule that simplifies adjudicative determinations. Some errors fall above the line, demanding correction, while others remain below, deemed insufficient to warrant reversal. This threshold-based approach reduces the need for ongoing judicial recalibration and instead imposes a structured, rule-based mechanism for assessing legal errors. While this framework enhances predictability and procedural clarity, it risks oversimplifying the inherently complex nature of legal adjudication, potentially overlooking case-specific nuances that might otherwise influence determinations of fairness and justice.

II. CRIMINAL ADJUDICATION AND AI ERROR

What does the fact of the harmless error doctrine and its categorical error tolerance approach tell us about how AI ought to be absorbed in criminal adjudication? As an initial matter, the technologization of criminal adjudication is hardly new. In fact, there is reason to believe that the type of technological processes

[75] The x-axis here is labeled "accuracy" to maintain continuity with Figures 1 and 2, but the reader will note that the variable is doing different work in this figure. In Figures 1 and 2, accuracy is a property of the adjudication; here, movement along the x-axis tracks the gravity of the prospective error. Numeric intervals have accordingly been omitted, as the precise position of either threshold along an accuracy scale is not what the harmless-error doctrine fixes, as what matters is the categorical distinction among the three zones.

ISSN 0041-9915 (print) 1942-8405 (online) ● DOI 10.5195/lawreview.2026.1167
http://lawreview.law.pitt.edu

that underlie artificial intelligence—for instance, machine-led analyses of large volumes of data, automated computation, synthesis, etc.—have been more substantively at play in criminal trials than in other segments of the legal profession.[76] There are at least two explanations. First, law enforcement has been at the frontier of AI adoption.[77] Government agencies in general are quick to explore AI solutions, particularly in austerity contexts or where there is a purported labor bottleneck, such as in processing applications or responding to inquiries.[78] Yet the peculiar political economy of policing warrants some attention here. Departmental budgets in most American locales are a year-to-year proposition,[79] and AI adoption proposes the dual benefit of cost savings while also entitling departments to a piece of a growing pie of AI-specific funding.[80]

Apart from this, some of the enthusiasm for AI in policing has to do with how AI techniques dovetail with a long and broader movement toward "data-driven" policing.[81] Indeed, policing in general has taken a turn—albeit a fraught one—toward *some* empirical approaches for a century.[82] The result is a set of practices that might

---

[76] *See* Danielle Kehl, Priscilla Guo & Samuel Kessler, *Algorithms in the Criminal Justice System: Assessing the Use of Risk Assessments in Sentencing*, BERKMAN KLEIN CTR. FOR INTERNET & SOC'Y, HARV. L. SCH. 3, 9 (2017), https://dash.harvard.edu/entities/publication/73120379-010a-6bd4-e053-0100007fdf3b [https://perma.cc/MWQ2-Y8SN] (noting that "[c]rime prediction has been a feature of the United States criminal justice system since the early 1920s," but today's algorithmic tools are part of a fourth generation of risk assessment).

[77] NICOLE EZEH, AMBER WIDGERY & CHELSEA CANADA, ARTIFICIAL INTELLIGENCE IN LAW ENFORCEMENT: THE FEDERAL AND STATE LANDSCAPE 1 (2025).

[78] Clare Martorana, *AI in Action: 5 Essential Findings from the 2024 Federal AI Use Case Inventory*, U.S. CHIEF INFO. OFFICERS COUNCIL (Jan. 15, 2025), https://www.cio.gov/ai-in-action/ [https://perma.cc/C34H-MTRQ].

[79] *See* Stephen Rushin & Roger Michalski, *Police Funding*, 72 FLA. L. REV. 277 (2020) (discussing the mechanics of police funding).

[80] James Redden, Brian Aagaard & Travis Taniguchi, *Artificial Intelligence Applications in Law Enforcement: An Overview of Artificial Intelligence Applications and Considerations for State, Local, and Tribal Law Enforcement*, U.S. DEP'T OF JUST., NAT'L INST. JUST. 4 (Aug. 2020), https://cjtec.org/files/5f5f94aa4c69b [https://perma.cc/M7CU-C4CL].

[81] Joh describes the concept as capturing "both AI systems today and earlier efforts dating back to the 1990s that simply emphasize the increasing reliance of police decision-making on statistics." Elizabeth E. Joh, *Ethical AI in American Policing*, 3 NOTRE DAME J. ON EMERGING TECH. 261, 266 (2022).

[82] *See also The Evolution of Law Enforcement Data*, BENCHMARK ANALYTICS (Apr. 27, 2020), https://www.benchmarkanalytics.com/blog/the-evolution-of-law-enforcement-data/ [https://perma.cc/E944-VQBD]. Frank Pasquale notes that there are two waves of discourse about algorithms: "the first wave of algorithmic accountability focuses on improving existing systems," while "a second wave of research has asked whether they should be used at all—and, if so, who gets to govern them." Frank

ISSN 0041-9915 (print) 1942-8405 (online) ● DOI 10.5195/lawreview.2026.1167
http://lawreview.law.pitt.edu

be unfamiliar to the police officers of old, but has also ushered in a robust epistemic community of researchers, policy analysts, and professional associations that debate—if not promote—the use of more scientific approaches to law enforcement. Accommodating AI is reasonably seen as an extension of these approaches; in fact, AI purports to do some of the computational tasks that are involved in data-driven policing (e.g., prediction, surveillance, mapping, etc.) more effectively and more cheaply.[83] The "product-market fit," to borrow a technologist's well-worn term, is rather apparent.

There is also growing evidence of earnest enthusiasm on the part of law enforcement professionals.[84] Recent surveys of law enforcement officials have suggested less compunction about AI than initially presumed; one recent global survey, for instance, found that seventy-nine percent of law enforcement personnel "agree that AI contributes to more effective investigations and makes their job easier."[85] The result is that where lawyers and law firms might perseverate about whether to use generative AI tools for comparatively low-stakes endeavors like case summaries and memorandum drafting,[86] police have embraced AI tools without as much reservation. The wider technology industry seems to have taken note. Familiar

Pasquale, *The Second Wave of Algorithmic Accountability*, LAW & POL. ECON. PROJECT (Nov. 25, 2019), https://lpeproject.org/blog/the-second-wave-of-algorithmic-accountability/ [https://perma.cc/5R44-D7JP].

[83] Christopher Rigano, *Using Artificial Intelligence to Address Criminal Justice Needs*, 2019 NAT'L INST. JUST. J. 37, 38–44. For a discussion of AI's purported cost savings at the level of given tasks, see Daron Acemoglu, *The Simple Macroeconomics of AI*, 39 ECON. POL'Y 3, 26–28 (2024).

[84] *See, e.g.*, Tom McKay, *Here's How Federal Law Enforcement Officials Are Actually Using AI*, IT BREW (June 13, 2024), https://www.itbrew.com/stories/2024/06/13/here-s-how-federal-law-enforcement-officials-are-actually-using-ai [https://perma.cc/5UUG-Y38S] ("Federal law enforcement officials are already using AI tools, and they're eager to pilot more, despite pesky challenges like admissibility, privacy, and sustainability. That's according to federal officials who spoke at General Dynamics Information Technology's (GDIT) AI conference in DC on June 4, where they explained how agencies like the DOJ, the FBI, and the State Department approach implementing AI technologies.").

[85] *2025 Industry Trends Survey—Insights into AI and Digital Investigations*, CELLEBRITE (Nov. 5, 2024), https://cellebrite.com/en/2025-industry-trends-survey-insights-into-ai-and-digital-investigations/ [https://perma.cc/7MMX-PGYN].

[86] There are endless examples of lawyerly and scholarly skepticism about AI in the field. *See, e.g.*, Simon Stern, *Introduction: Artificial Intelligence, Technology, and the Law*, 68 U. TORONTO L.J. 1 (2018); Frank Pasquale, *The Resilient Fragility of Law*, *Foreword* to IS LAW COMPUTABLE?: CRITICAL PERSPECTIVES ON LAW AND ARTIFICIAL INTELLIGENCE (Simon Deakin & Christopher Markou eds., Hart Publishing 2021); *see also* Nicholas Mignanelli, *The Legal Tech Bro Blues: Generative AI, Legal Indeterminacy, and the Future of Legal Research and Writing*, 8 GEO. L. TECH. REV. 298 (2024).

ISSN 0041-9915 (print) 1942-8405 (online) ● DOI 10.5195/lawreview.2026.1167
http://lawreview.law.pitt.edu

names like Microsoft[87] and IBM[88] have developed products and applications specifically for law enforcement, while more specialized outfits like Clearview AI,[89] Palantir,[90] and Axon[91] are all used by law enforcement. As a result, critical personnel in criminal adjudication have had to contend with such significant considerations as the integrity of AI evidence, accuracy and biases in prediction, the technology's capacity to evaluate data, etc.[92]

In addition to law enforcement's early and steady embrace of AI, a *second* reason for the proliferation of AI in criminal adjudication relates to a long effort on the part of AI enthusiasts to disrupt "lawyering," so-to-speak. Much scholarly ink has been spilled over whether the legal profession can indeed be "automated,"[93] but

---

[87] *See* Chad Wallace, *Upgrade Law Enforcement Response with AI Solutions from Microsoft*, MICROSOFT: INDUS. BLOG (Sept. 21, 2023), https://www.microsoft.com/en-us/industry/blog/government/2023/09/21/upgrade-law-enforcement-response-with-ai-solutions-from-microsoft/ [https://perma.cc/D5EE-AJ5Y] (discussing the new technologies Microsoft is developing to help law enforcement).

[88] *See* Cristina Caballé Fuguet & Markus Straub, *How Emerging Technologies Can Transform Law Enforcement and Public Safety*, IBM: INST. FOR BUS. VALUE (Mar. 13, 2023), https://www.ibm.com/thought-leadership/institute-business-value/en-us/blog/emerging-technologies-law-enforcement-safety [https://perma.cc/Y3NM-9HVQ] ("IBM and the Munich Innovation Labs have been working on a series of innovations to leverage AI to help enhance the role of the law enforcement and public safety personnel while also ensuring its responsible use.").

[89] *See* Suzanne Smalley, *Law Enforcement Searches of Clearview AI Facial Recognition Doubled in Past Year*, THE RECORD (June 27, 2024), https://therecord.media/clearview-ai-police-searches-doubled-2023 [https://perma.cc/DAS8-WMEX] ("The company's technology is used by federal law enforcement agencies and police departments nationwide.").

[90] *See* Ali Ahmed, *Is Palantir Technologies (PLTR) the Best Prison and Law Enforcement Stock to Invest In?*, YAHOO FIN. (Nov. 29, 2024), https://finance.yahoo.com/news/palantir-technologies-pltr-best-prison-084158977.html [https://perma.cc/Q22Z-6N89] (discussing how Palantir is widely used by defense and law enforcement sectors).

[91] James Morales, *How a Pivot to AI Helped Axon Enterprise Stock Climb 144% in a Year*, CCN (Jan. 25, 2025, 4:10 PM), https://www.msn.com/en-us/technology/tech-companies/how-a-pivot-to-ai-helped-axon-enterprise-stock-climb-144-in-a-year/ar-AA1xQT4d? [https://perma.cc/T985-MB3R] (noting that Axon develops technology for law enforcement). For an academic treatment of Axon's flagship law enforcement technology called "Draft One," see Andrew Guthrie Ferguson, *Generative Suspicion and the Risks of AI-Assisted Police Reports*, 120 NW. U. L. REV. 299, 314 (2025).

[92] Of course, at least some of the impetus behind law enforcement's embrace of AI relates to the fact that AI is also being exploited by those who seek to break the law. *See* Bart Custers, *AI in Criminal Law: An Overview of AI Applications in Substantive and Procedural Criminal Law*, *in* LAW AND ARTIFICIAL INTELLIGENCE: REGULATING AI AND APPLYING AI IN LEGAL PRACTICE, 35 INFO. TECH. & L. SERIES 205 (Bart Custers & Eduard Fosch-Villaronga eds., 2022) (describing use of AI tools by both criminals and individuals in law enforcement).

[93] *See, e.g.*, Amin Ebrahimi Afrouzi, *John Robots, Thurgood Martian, and the Syntax Monster: A New Argument Against AI Judges*, 37 CANADIAN J.L. & JURIS. 369 (2024); Abdi Aidid, *Toward an Ethical*

ISSN 0041-9915 (print) 1942-8405 (online) ● DOI 10.5195/lawreview.2026.1167
http://lawreview.law.pitt.edu

technologists have nevertheless developed tools for use in a growing number of legal contexts.[94] Though many of these are not specific to criminal lawyering, they have cross-cutting influence: for instance, machine-learning enabled prediction of case outcomes[95] or generative AI for legal drafting. Economic reasons for targeting the legal profession aside, law has been seen by scholars as an area that might lend itself especially well to AI-enabled applications.

Constitutional concerns have long been present in discussions about the use of AI in legal settings.[96] Indeed, scholars have been particularly attentive to AI's purported unreliability when surfacing prospective substantive rights issues. Taken together, concerns have tended to focus on several constitutional criminal-procedural rights, namely the Fifth and Fourteenth Amendments' Due Process Clauses.[97] Here, it is worth identifying the particular context in which AI will have a heightened tendency to implicate substantive rights in criminal investigation and adjudication.

Human-Computer Division of Labor in Law Practice, 92 FORDHAM L. REV. 1797 (2024); Dana Remus & Frank Levy, *Can Robots Be Lawyers?: Computers, Lawyers, and the Practice of Law*, 30 GEO. J. LEGAL ETHICS 501 (2017).

[94] *See* Sarah A. Sutherland, *AI Use Skyrocketing at North American Law Firms*, CANADIAN BAR ASS'N NAT'L MAG. (Nov. 4, 2024), https://nationalmagazine.ca/en-ca/articles/legal-market/legal-tech/2024/ai-use-skyrocketing-at-north-american-law-firms [https://perma.cc/2EWU-T3GV] ("79 per cent of North American legal professionals reported incorporating AI tools into their daily work this year . . . ."); David Halliwell & Mike McGlinchey, *Artificial Intelligence Taking Growing Role in Legal Sector*, PINSENT MASONS (Nov. 8, 2024, 4:49 PM), https://www.pinsentmasons.com/out-law/analysis/artificial-intelligence-growing-role-legal-sector [https://perma.cc/8HZQ-U4C2] ("An impressive 82% of lawyers are now either using or planning to integrate AI into their practices, up from 39% last year, according to recent LexisNexis research."); David Horrigan, *New Research Study Predicts Continued Growth for Generative AI in Legal*, LAW.COM (Nov. 26, 2024, 11:01 AM), https://www.law.com/legaltechnews/2024/11/26/new-research-study-predicts-continued-growth-for-generative-ai-in-legal/ [https://perma.cc/R65L-V37R] ("A new study . . . predicts continued growth in the use of generative AI by legal professionals, with 69% of study respondents reporting they expect their use to increase over the next two years.").

[95] *See, e.g.*, Benjamin Alarie & Abdi Aidid, *Predicting Economic Substance Cases with Machine Learning*, 2020 J. TAX. PRAC. & PROC. 35, 35–36 (2020) (demonstrating how algorithms can correctly predict outcomes in certain tax cases).

[96] *See, e.g.*, Thomas B. Nachbar, *Algorithmic Fairness, Algorithmic Discrimination*, 48 FLA. STATE U. L. REV. 509 (2021).

[97] *See* Brandon L. Garrett & Cynthia Rudin, *The Right to a Glass Box: Rethinking the Use of Artificial Intelligence in Criminal Justice*, 109 CORNELL L. REV. 561, 606 (2024) (focusing on the Sixth Amendment and the Due Process Clauses in the Fourteenth and Fifth Amendments); Andrew Guthrie Ferguson, *Facial Recognition and the Fourth Amendment*, 105 MINN. L. REV. 1105 (2021); Brandon L. Garrett & Cynthia Rudin, *Testing AI* (Duke L. Sch. Publ. L. & Legal Theory Series, Working Paper No. 2024-62, 2024), https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4948789 [https://perma.cc/8QAC-EY6A]; Ferguson, *supra* note 91 (manuscript at 37); Sonia K. Katyal, *The Paradox of Source Code Secrecy*, 104 CORNELL L. REV. 1183, 1250–79 (2019).

ISSN 0041-9915 (print) 1942-8405 (online) ● DOI 10.5195/lawreview.2026.1167
http://lawreview.law.pitt.edu

As Garrett et al. observe, there is a meaningful difference between AI being used "to search for a missing person feared to have been kidnapped" compared to "using AI to identify a culprit from a surveillance video."[98] The latter AI application is considerably more likely to generate evidence that can then be used to restrain a prospective defendant. In other words, AI surveillance will always implicate a prospective defendant, whereas using AI for certain public safety ends may only incidentally do so.

III. A TYPOLOGY OF AI ERROR IN CRIMINAL JUSTICE

With this considered, AI tools in criminal adjudication generally belong to one of three types: predictive, evaluative, and generative.[99] I elaborate on these below, but predictive AI generally refers to the use of data science techniques such as machine learning algorithms to assess historical data for patterns and to create forecasts, often to support decision-making.[100] Generative AI might have some predictive components, but is used here generally to denote AI that learns from data to create net new information or novel content.[101] Evaluative AI is method-agnostic, but refers chiefly to its purpose: to assess, analyze, and make probabilistic determinations;[102] for example, whether DNA samples in criminal trials belong to a given defendant. Note that for the purposes of this Article, I focus on the technologies which are relevant to a defendant's prospective conviction. This means excluding the kind of AI which might be adopted more readily, but generates few of the same

[98] Brandon L. Garrett, Alicia Carriquiry, Karen Kafadar, Robin Mejia, Cynthia Rudin, Nicholas Scurich & Hal Stern, *AI and Criminal Procedure Rights: A Response to the National Institute of Justice Request for Input*, WILSON CTR. FOR SCI. & JUST. 8 (May 28, 2024), https://wcsj.law.duke.edu/publications/ai-and-criminal-procedure-rights-a-response-to-the-national-institute-of-justice-request-for-input/ [https://perma.cc/ZG5R-V5YN].

[99] These three categories—predictive, evaluative, and generative—are not the only ways to classify AI. However, for the purposes of this article, I have chosen these cross-cutting categories to better illustrate the broader implications of AI in legal practice. This framework allows for a clearer discussion of AI's multifaceted role while acknowledging that other categorizations may be equally valid in different contexts.

[100] Tim Mucci, *What Is Predictive AI?*, IBM: THINK, https://www.ibm.com/think/topics/predictive-ai [https://perma.cc/82YX-K8D3] (last visited Oct. 30, 2025).

[101] *See* Michael D. Murray, *Generative Artificial Intelligence—Where Did it Come From? How Does it Work?* (Univ. of Ky., Rosenberg Coll. of L., Working Paper, 2024), https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4909859 [https://perma.cc/9BEL-66PX] (giving a more general description of generative AI and how it works outside of the legal market). *See generally* Paul Douglas Callister, *Generative AI and Finding the Law*, 117 L. LIBR. J. 1 (2025) (providing a detailed overview of generative AI and how it can be used in the legal field).

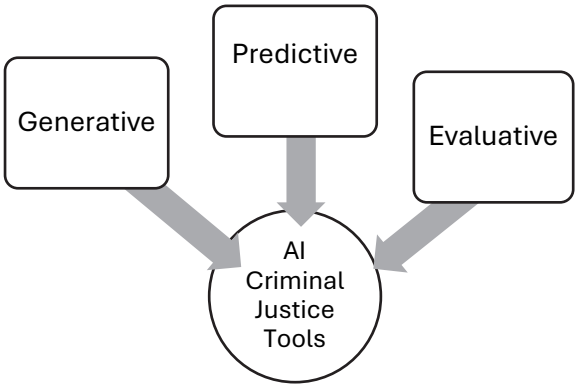
[102] Tim Miller, *Explainable AI Is Dead, Long Live Explainable AI!: Hypothesis-Driven Decision Support* 1, 1 (Sch. of Computing & Info. Sys., Univ. of Melb., Working Paper, 2023), https://arxiv.org/abs/2302.12389 [https://perma.cc/6U38-3W4N].

ISSN 0041-9915 (print) 1942-8405 (online) ● DOI 10.5195/lawreview.2026.1167
http://lawreview.law.pitt.edu

ethical, legal, or policy concerns. For instance, prosecutors, defense counsel, judges, and police officers may use AI-enabled productivity tools that might affect the criminal defendant in a background sense, but do not bear on a defendant's substantive rights.

As I explain, the nature of the errors made by predictive AI, generative AI and evaluative AI varies in terms of arbitrariness. Predictive AI, being cabined to what exists in the data, tends to make errors that are data-induced and relate to how the underlying information is collected, structured, and analyzed.[103] Generative AI, in its higher-order creative aspirations, will tend to err in ways that are more arbitrary in the sense that they are untethered from the underlying information. Evaluative AI's errors are the least arbitrary because they exist in simple, binary states of success or failure, given that they make largely classification determinations on the basis of user-supplied inputs. Figure 4 below charts the three types according to the *arbitrariness of error*.

Figure 4



A. Predictive AI Error

Predictive AI tools are those that are meant to forecast outcomes, typically by ingesting significant volumes of data and deploying algorithms to synthesize that data. Prior to the advent of ChatGPT and the resulting public fixation with generative AI, predictive AI tools predominated the public discourse, largely due to their use in particular high-stakes contexts. Algorithmic decision-making tools ("ADM") are the

[103] Saying that errors are "cabined to the data" simply means that the data is the ultimate source for the issues, though issues can be generated from how that data is tested as well. *See* Catherine Stinson, *Algorithms Are Not Neutral*, 2 AI & ETHICS 763, 763 (2022) ("The main concern here is where in the workflow from data collection and algorithm design to testing and implementation the causes of discriminatory outcomes can be found. Because those causes are not restricted to the data preparation phases, de-biasing data sets is not always a viable strategy for mitigating algorithmic bias.").

ISSN 0041-9915 (print) 1942-8405 (online) ● DOI 10.5195/lawreview.2026.1167
http://lawreview.law.pitt.edu

canonical example of predictive AI, wherein institutions will perform computational analyses on their own data to "automatically complete tasks and/or make decisions," such as a credit card company using an algorithm to determine creditworthiness or a social service agency using the same to determine a citizen's eligibility for a social program.[104]

Most current AI criminal justice applications fall into this category. Three particular applications have captured much of the public attention: (1) predictive policing, (2) predictive identification, and (3) post-arrest risk assessment. Predictive policing algorithms analyze patterns to forecast where and when—and sometimes more troublingly, by *whom*—crime is likely to occur.[105] The applications take stock of large volumes of historical crime data, which are reconciled against other indicators, such as social and demographic information and geography. The purpose is to enable purportedly higher-resolution policing and typically takes one of four forms: analytical techniques for "predicting crimes," for "predicting offenders," for "predicting perpetrators' identities," and for "predicting victims of crime."[106]

[104] Christoph F. Breidbach, *Responsible Algorithmic Decision-Making*, ORGANIZATIONAL DYNAMICS, Jan. 28, 2024, at 2; *see also id.* ("Algorithms can make decisions because of the vast volumes of electronic data that have become available. Often described as *big data*, organizations can collect, store, and process increasingly large and complex operations, customer, and other business data—often even in real-time. In fact, you may want to think of big data as the 'digital footprints' that result from increasingly complex technology-enabled human-to-human, human-to-machine, or machine-to-machine interactions."). For a good discussion of some of the broader challenges of using ADM, see Kiel Brennan-Marquez & Vincent Chiao, *Algorithmic Decision-Making When Humans Disagree on Ends*, 24 NEW CRIM. L. REV. 275 (2021).

[105] The term "predictive policing" is sometimes used interchangeably with related terms like pre-emptive policing, platform policing, and precision policing. For treatments of "predictive policing," see Elle Pearson, Rikke Bjerg Jensen & Peter Adey, *Pred-Pol-Pov: Visibility, Data Flows, and the Predictive Policing of Poverty*, 22 SURVEILLANCE & SOC'Y 120 (2024); Andrew Guthrie Ferguson, *Predictive Policing and Reasonable Suspicion*, 62 EMORY L.J. 259 (2012); Albert Meijer & Martijn Wessels, *Predictive Policing: Review of Benefits and Drawbacks*, 42 INT'L J. PUB. ADMIN. 1031 (2019); Craig D. Uchida, *Predictive Policing*, *in* ENCYCLOPEDIA OF CRIMINOLOGY AND CRIMINAL JUSTICE 3871 (Gerben Bruinsma & David Weisburd eds., 2014). For treatments of pre-emptive policing, see Mark Andrejevic, Lina Dencik & Emiliano Treré, *From Pre-Emption to Slowness: Assessing the Contrasting Temporalities of Data-Driven Predictive Policing*, 22 NEW MEDIA & SOC'Y 1528 (2020); Simon Egbert & Susanne Krasmann, *Predictive Policing: Not Yet, but Soon Preemptive?*, 30 POLICING & SOC'Y 905 (2020). For treatments of platform policing and precision policing, see Thomas Linder, *Surveillance Capitalism and Platform Policing: The Surveillant Assemblage-as-a-Service*, 17 SURVEILLANCE & SOC'Y 76, 78 (2019).

[106] WALTER L. PERRY, BRIAN MCINNIS, CARTER C. PRICE, SUSAN C. SMITH & JOHN S. HOLLYWOOD, PREDICTIVE POLICING: THE ROLE OF CRIME FORECASTING IN LAW ENFORCEMENT OPERATIONS xiv (2013) (articulating a "taxonomy of predictive methods").

ISSN 0041-9915 (print) 1942-8405 (online) ● DOI 10.5195/lawreview.2026.1167
http://lawreview.law.pitt.edu

Importantly, the tools are meant to elucidate where an increased likelihood of crime is present, not foretell any specific crime.[107]

Here, it is important to once again emphasize precisely how longstanding these methods are. Andreas M. Olligschlaeger developed a model, "predictive crime mapping," in 1997 that purported to forecast the locations of drug markets.[108] The basic idea behind the Olligschlaeger model was to train a neural network—i.e., AI—with historical crime data and to engage in an ongoing process of adjusting the weights of any input data, through an error-minimization technique called backpropagation.[109] These inputs "included 35 months of weapon-related, robbery, and assault-related calls for service; the relative proportions of residential and commercial properties [in a given area]; and a seasonally index."[110] The model would then be layered atop a geographic information system ("GIS")-assisted grid.[111] In contexts where there are rich data for productive inputs, such as voluminous arrest records or information about local properties, Olligschlaeger's predictive approach even outperformed leading-edge data science techniques of the day like multiple regression leading indicator models.[112]

Today's predictive policing approaches improve on these early models. Just as significantly, models have moved from being the near-exclusive domain of academic researchers to taking up residence in a world of start-up companies and software vendors. Take, as one example, the City of Philadelphia's experiment with predictive policing.[113] The project was "place-based"—i.e., aimed at identifying where a crime was likely to occur, rather than who was likely to commit it—and made use of an

[107] *Id.* at 127.

[108] Elizabeth R. Groff & Nancy G. La Vigne, *Forecasting the Future of Predictive Crime Mapping*, 13 CRIME PREVENTION STUD. 29, 46 (2002) ("One of the earliest efforts to do predictive crime mapping was that of Olligschlaeger (1997), who employed a 'feed-forward network with backpropogation' to predict areas where future drug markets will emerge."). *See generally* Andreas M. Olligschlaeger, *Artificial Neural Networks and Crime Mapping*, *in* 8 CRIME MAPPING AND CRIME PREVENTION 313, 313, 315–17 (David Weisburd & Tom McEwan eds., 1997).

[109] Olligschlaeger, *supra* note 108, at 325.

[110] Groff & La Vigne, *supra* note 108, at 46.

[111] *Id.*

[112] WILPEN GORR & ANDREAS OLLIGSCHLAEGER, CRIME HOT SPOT FORECASTING: MODELING AND COMPARATIVE EVALUATION¸ FINAL PROJECT REPORT 16 (2002); Groff & La Vigne, *supra* note 108, at 47 ("In more recent work, Olligschlaeger and Gorr (2001) found that neural networks outperform multiple regression leading indicator models when the set of leading indicators is rich and numerous.").

[113] *See generally* Jerry H. Ratcliffe, Ralph B. Taylor, Amber Perenzin Askey, Kevin Thomas, John Grasso, Kevin J. Bethel, Ryan Fisher & Josh Koelnlein, *The Philadelphia Predictive Policing Experiment*, 17 J. EXPERIMENTAL CRIMINOLOGY 15 (2021).

ISSN 0041-9915 (print) 1942-8405 (online) ● DOI 10.5195/lawreview.2026.1167
http://lawreview.law.pitt.edu

algorithm developed by an independent software firm Azavea, Inc., a described believer in "the power of geospatial technology to improve communities and the planet."[114] Evaluators described the algorithm as using "machine learning techniques that incorporate components of," among other techniques, "risk terrain modeling" ("RTM").[115] Kennedy et al. describe RTM as follows:

> RTM assigns a value signifying the presence, absence or intensity of each risk factor at every place throughout a given geography. Each factor is represented by a separate terrain (risk map layer) of the same geography. When all map layers are combined in a GIS [Geographic Information System], they produce a composite map—a risk terrain map—where every place throughout the geography is assigned a composite risk value that accounts for all factors associated with the particular crime outcome. The higher the risk value the greater the likelihood of a crime event occurring at that location.[116]

A rather substantial literature is dedicated to explaining and evaluating RTM,[117] but the simplest non-technical explanation is that data about historic crime patterns is mapped spatially and reconciled against attributes of the built environment. For a more detailed understanding, imagine that analysts had information—obtained through qualitative and quantitative research, literature reviews, etc.—about what factors relate to the incidence of a given crime. Now imagine that you could place each one of those factors on a map to identify where that factor is geographically present, but rather than simply map the presence of that factor, you can map its "presence, absence or intensity" at every location.[118] Now imagine you could make separate maps for each factor and later use GIS to combine all of the layers. The result would be a "composite map—a risk terrain map—where every place throughout the geography is assigned a composite risk value that accounts for all

[114] Jessica Cahail, *Addressing Environmental Challenges with the AWS Cloud*, AWS: PUB. SECTOR BLOG (Oct. 13, 2020), https://aws.amazon.com/blogs/publicsector/addressing-environmental-challenges-aws-cloud/ [https://perma.cc/N8XS-398A].

[115] Ratcliffe et al., *supra* note 113, at 21.

[116] Leslie W. Kennedy, Joel M. Caplan & Eric Piza, *Risk Clusters, Hotspots, and Spatial Intelligence: Risk Terrain Modeling as an Algorithm for Police Resource Allocation Strategies*, 27 J. QUANTITATIVE CRIMINOLOGY 339, 343 (2011).

[117] *See, e.g.*, Joel M. Caplan, Leslie W. Kennedy & Joel Miller, *Risk Terrain Modeling: Brokering Criminological Theory and GIS Methods for Crime Forecasting*, 28 JUST. Q. 360 (2011). For discussion of other methods, see Martin A. Andresen, *Cory P. Haberman, Shane D. Johnson & Wouter Steenbeek, Advances in Place-Based Methods: Editors' Introduction*, 37 J. QUANTITATIVE CRIMINOLOGY 327 (2021).

[118] Kennedy et al., *supra* note 116, at 343.

ISSN 0041-9915 (print) 1942-8405 (online) ● DOI 10.5195/lawreview.2026.1167
http://lawreview.law.pitt.edu

factors associated with the particular crime outcome."[119] The basic conceit is that being attentive to features of the landscape yields higher-resolution crime insights and equips law enforcement with the ability to identify "hotspots" and direct resources accordingly.[120]

Of course, human beings do the same thing. Individuals might avoid dimly lit stairways or back alleys because of an intuition that there is an elevated risk of violent crime. The most rational among us might only become motivated to avoid such places because of stories or personal experiences with violence in places that feature the same characteristics. Therein lies the key insight: Predictive tools lay reasonable claim to a superior ability to do what we are all trying to do, anyway. Return to the RTM example above. Where an individual might avoid the dim stairway because of an uneasy feeling or anecdotal evidence, the RTM replaces the uneasy feeling with a probabilistic determination and the anecdotal evidence with a synthesis of vastly more information. Moreover, the RTM is engaged in ongoing Bayesian updating; as it is fed more information in the way of geolocated crime statistics, it departs from previously-held views sooner, whereas individuals might continue to intuit on the basis of outmoded priors.[121]

Consider this in the context of criminal procedure's well-worn "reasonable suspicion" standard for legitimate arrests. The predictive component of the standard is perhaps most obvious in the context of the anticipatory search warrant, wherein a magistrate must "determine (1) that it is *now probable* that (2) contraband, evidence of a crime, or a fugitive *will be* on the described premises (3) when the warrant is executed."[122] Courts have been clear that this is fundamentally a predictive exercise: In *United States v. Grubbs*, the Supreme Court shared the oft-cited maxim that a magistrate's "determination that there is probable cause . . . amounts to a prediction . . . ."[123] As Ferguson explains, the accuracy of the prediction matters:

> It is always possible that the contraband will be gone, but there is a prediction that police will find it. The prediction usually includes a temporal element because information can grow stale. In addition, it is usually particularized to a specific

[119] *Id.*

[120] *Id.* at 341, 343.

[121] Zoe Marchment & Paul Gill, *Systematic Review and Meta-Analysis of Risk-Terrain Modeling (RTM) as a Spatial Forecasting Method*, CRIME SCI., June 16, 2021, at 2 (discussing Bayesian methods and information updating in certain RTM approaches).

[122] United States v. Grubbs, 547 U.S. 90, 96 (2006).

[123] *Id.* at 95.

ISSN 0041-9915 (print) 1942-8405 (online) ● DOI 10.5195/lawreview.2026.1167
http://lawreview.law.pitt.edu

> area or person to be searched. The controlling standard of probable cause, as the name suggests, turns on probabilities. Predicting those probable outcomes rests on predictive guesses about whether the evidence or person sought will be at a particular location at a particular time.[124]

Here, the procedure reveals that magistrates are fundamentally looking for meaningful data. In *United States v. Garcia*, the United States Court of Appeals for the Second Circuit explained that affidavits furnished to obtain anticipatory warrants have to show "not only that the agent believes a delivery of contraband is going to occur, but also *how* he has obtained this belief" and "how reliable his sources are."[125] The claim that a predictive policing model makes here is that, like the anticipatory search warrant, it is based on information (but *more* information), it makes predictions as to the likelihood of a crime (but *better* predictions), and provides an accounting of *how* it obtained the belief (but a *more* rigorous one).

What does this all mean for predictive AI's vulnerability to error? One important distinction between predictive AI and generative AI relates to how the former does not create net new information; instead, predictive AI excavates information that is already present in large historical data sets. By contrast, generative AI—which I discuss more extensively below—will produce distinct (i.e., novel) creations based on what it has learned from the data. Predictive AI instead works within the four corners of the data universe to infer rather than create. This fundamental limitation means that predictive AI's errors are not the result of spontaneous miscalculations or creative distortions, as might be the case with generative models. Instead, its inaccuracies typically stem from two primary sources: (1) deficiencies inherent in the underlying data and (2) suboptimal methodological choices in data processing and model selection. In other words, Predictive AI exhibits parallel failure points to human cognition: its mistakes are largely the result of the data it is trained on (akin to our sources of information) and the algorithms used to interpret that data (akin to our cognitive processing power).

With respect to this first category, deficiencies inherent in the data, these range from innocuous concerns like poor quality data, to substantive concerns like gaps in data, to more critical concerns like data encoding major historical inequities. Indeed, the critical literature thus far has rightly focused on AI's propensity to amplify harm

[124] Ferguson, *supra* note 105, at 286.

[125] United States v. Garcia, 882 F.2d 699, 703 (2d Cir. 1989).

ISSN 0041-9915 (print) 1942-8405 (online) ● DOI 10.5195/lawreview.2026.1167
http://lawreview.law.pitt.edu

by projecting negative histories forward.[126] And while the problem of, say, a history of racism is not a "data" problem, it becomes an issue when mediated through predictive AI in part because the technology reads negative histories *as* data. Put differently, there is a tendency of predictive AI to prioritize statistical correlations without an understanding of causal or contextual factors.[127] An RTM predictive policing algorithm that counsels in favor of investing law enforcement resources in a predominantly black housing complex does so in part because it identifies some correlative relationships in the topline statistics.[128] However, because predictive AI lacks the epistemic capacity to interpret historical injustices, it does not recognize that this pattern may be the result of over-policing rather than an inherent feature of the location itself. Unlike a human decision-maker who, at least in theory, has access to broader historical, sociological, and normative understandings, the algorithm is constrained by the data that has been provided to it and the specific techniques used to process that data.[129] In this way, methodological choices become complicit in the

[126] *See, e.g.*, Douglas Yeung, Inez Khan, Nidhi Karla & Osonde A. Osoba, *Identifying Systemic Bias in the Acquisition of Machine Learning Decision Aids for Law Enforcement Applications*, RAND CORP. (2021), https://www.rand.org/pubs/perspectives/PEA862-1.html [https://perma.cc/M2EM-K68T]; Mayson, *supra* note 16, at 2218 ("Algorithmic risk assessment has revealed the inequality inherent in all prediction, forcing us to confront a problem much larger than the challenges of a new technology."); Erin Collins, *Punishing Risk*, 107 GEO. L.J. 57 (2018); Teresa Scassa, *A Reasonable Apprehension of AI Bias: Lessons from* R. v. RDS (Ottawa Fac. of L., Working Paper No. 2024-01), https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4763287 [https://perma.cc/8JDR-S3CE]; Anupam Chander, *The Racist Algorithm?*, 115 MICH. L. REV. 1023 (2017).

[127] Abraham Itzhak Weinberg, Cristiano Premebida & Diego Resende Faria, *Causality from Bottom to Top: A Survey*, MACH. LEARNING, Sept. 21, 2025, at 27 ("AI and ML today rely primarily on statistical correlations in data to infer causation, whereas humans understand causation through intuitive theories about how the world works based on mechanisms, context, and common sense knowledge. Humans also view causation dynamically and consider uncertainties like hidden or confounding variables.").

[128] Anthony A. Braga, Brandon Turchan, Andrew V. Papachristos & David M. Hureau, *Hot Spots Policing of Small Geographic Areas Effects on Crime*, CAMPBELL SYSTEMATIC REVS., Sept. 8, 2019, at 2 ("Focusing police efforts at high activity crime places is more likely to produce a diffusion of crime prevention benefits into areas adjacent to targeted hot spots than crime displacement."); *see also* Anthony A. Braga, Andrew V. Papachristos & David M. Hureau, *The Effects of Hot Spots Policing on Crime: An Updated Systematic Review and Meta-Analysis*, 31 JUST. Q. 633 (2014).

[129] *See* Gabbrielle M. Johnson, *Algorithmic Bias: On the Implicit Biases of Social Technology*, 198 SYNTHESE 9941, 9945 (2021) (describing algorithmic bias from training data); *see also* Runshan Fu, Yan Huang & Param Vir Singh, *AI and Algorithmic Bias: Source, Detection, Mitigation and Implications*, PUSHING THE BOUNDARIES: FRONTIERS IN IMPACTFUL OR/OM RSCH., Nov. 2020, at 1, 19 (explaining the dangers of encoded bias via training data).

ISSN 0041-9915 (print) 1942-8405 (online) ● DOI 10.5195/lawreview.2026.1167
http://lawreview.law.pitt.edu

perpetuation of structural biases, as they dictate how raw statistical inputs are transformed into actionable intelligence.[130]

Given the second category, suboptimal methodological choices in data processing and model selection, errors in predictive AI extend beyond the dataset itself to the ways in which that data is structured, interpreted, and operationalized. Even if the dataset were perfectly representative—a theoretical ideal rarely, if ever, realized—the methodologies employed to extract patterns from that data introduce further layers of complexity and potential distortion. The particular modeling technique, the weighting of different variables, the optimization functions used, and the assumptions embedded in statistical and machine-learning frameworks all influence the final outputs in ways that are not always transparent, even to those designing the system.

The fact that predictive AI's core aspiration is to surface information from existing data sets suggests that its errors are largely approachable. What does this mean? Predictive errors tend not to stem from fluctuations or the spontaneous generation of incorrect outputs. Instead, predictive errors emerge from ascertainable sources—flawed data, biased feature selection processes, weaknesses in model selection, etc. In sum, errors are correctable—in principle. Where generative AI errors are more unpredictable due to models' creative synthesis of information, predictive AI mistakes often manifest as misapplications of statistical patterns or misinterpretation of historical data trends.

One of the key advantages of this characteristic is that predictive AI errors can often be mitigated without necessitating a complete overhaul of the underlying model. Rather than discarding an entire predictive system upon identifying inaccuracies, targeted interventions can refine and improve its performance. These interventions might include curating higher-quality training data, adjusting weighting parameters to counteract bias, or implementing fairness-aware algorithms designed to correct for known distortions. The modular and iterative nature of machine learning models allows for adjustments at various stages of the pipeline, ensuring that errors can be addressed without unseating the entire predictive framework. Furthermore, the potential for correction is enhanced by the ability to conduct external audits and implement feedback loops. For instance, if a risk assessment model in criminal justice is found to disproportionately assign higher risk scores to certain demographic groups, researchers and policymakers can introduce

[130] For an example of an editorial that takes this view, see Will Douglas Heaven, *Predictive Policing Algorithms Are Racist. They Need to Be Dismantled.*, MIT TECH. REV. (July 17, 2020), https://www.technologyreview.com/2020/07/17/1005396/predictive-policing-algorithms-racist-dismantled-machine-learning-bias-criminal-justice/ [https://perma.cc/MRJ9-CAMW].

ISSN 0041-9915 (print) 1942-8405 (online) ● DOI 10.5195/lawreview.2026.1167
http://lawreview.law.pitt.edu

constraints, recalibrate the model, or employ post-hoc statistical adjustments to ensure fairer outcomes.

Perhaps more importantly, predictive error has a familiar template for the ways we diagnose and correct human reasoning errors and provide useful analogues for addressing AI errors. For instance, if a human makes a faulty risk assessment based on incomplete or biased data, the solution is often to introduce new information or adjust the way they weigh existing evidence. Similarly, when predictive AI produces biased or inaccurate outputs, the solution often involves refining the training data, adjusting weighting mechanisms, or recalibrating the algorithm to prevent distortions. The fact that AI is operating within a computational framework—where variables, parameters, and optimization functions can be adjusted—makes these corrections even more direct and systematic than they often are in human reasoning.

B. Generative AI Error

Compare predictive AI tools to the second category of criminal justice applications: generative AI tools. Generative AI is, at this point, moderately well-understood: It describes tools that ingest data primarily to *learn* and then perform creative acts that establish net new information thereafter. Generative AI is not widely used in criminal investigations or adjudications for anything approximating decision-making yet. Instead, it is primarily used for assistance in drafting and transcription. A key insight: unlike predictive AI tools which often demand bespoke algorithms, some forms of generative AI—like large language models—are widely available for use, albeit often in a general (i.e., not criminal law-specific) form, such as ChatGPT. As a result, generative AI tools can be introduced in criminal investigation and adjudication by virtually any actor. Indeed, it is not just institutional actors like courts, police officers or prosecutors that may introduce generative AI, it can just as easily be defendants, defense lawyers, investigators, or, in the particular case of deepfake evidence, occasional third parties. The evidentiary implications of this shift away from exclusively human-generated information are beyond the scope of this Article but have significant error-producing potential. To the extent that a piece of deep-faked video footage, a convincing-but-hallucinated passage from an officer's AI-drafted report, or an altered voice recording dupe of an officer, prosecutor or judge, that evidence can inure against a defendant's benefit at the expense of their substantial—and indeed constitutional—rights.

Here, a quick return to the topic of hallucinations is instructive. Hallucinations point to a problem that is exclusive to artificial intelligence: it errs at the level of technical process, which is often difficult to detect using our traditional (or analog) evaluative tools. In the rather substantial literature dedicated to human error, researchers will note that in conventional human reasoning processes, mistakes can be identified through cross-referencing sources, applying logic, or appealing to expertise. When a person makes a factual error, another person can typically

ISSN 0041-9915 (print) 1942-8405 (online) ● DOI 10.5195/lawreview.2026.1167
http://lawreview.law.pitt.edu

recognize and correct it based on shared knowledge or established methods of verification. AI hallucinations, however, do not follow this pattern. They emerge from the underlying mechanics of large language models, which generate responses based on probabilistic associations rather than a grounded understanding of truth or falsity.[131] More specifically, the error-producing activity is a largely unapproachable technical protocol.

As an illustrative example, consider the most basic use of generative AI in criminal adjudication: drafting and transcription. Some of the earliest coverage of AI's misapplication in the legal profession related to examples of lawyers that misapprehended tools like ChatGPT and submitted hallucinated material to courts. Perhaps coverage was motivated by professional *schadenfreude*, but hallucinated submissions pose a significant constitutional criminal-procedural problem, particularly to the extent that they are bases for ineffective assistance of counsel claims. One notable example involved the rapper Prakazrel Michel, better known as Pras, who was found guilty of various criminal conspiracy and campaign finance charges in the United States District Court for the District of Columbia.[132] Shortly after his conviction, Michel sought a new trial alleging that his lawyer used an AI platform called EyeLevel.AI to draft the trial's closing arguments.[133] According to his new defense team, Michel's trial lawyer "used an experimental artificial intelligence (AI) program to draft the closing argument, ignoring the best arguments and conflating the charged schemes, and he then publicly boasted that the AI program 'turned hours or days of legal work into seconds.'"[134] Among other things, the AI-drafted closing misattributed lyrics to Michel.[135]

EyeLevel is powered by a large language model and so operates in a manner conceptually similar to OpenAI's ChatGPT.[136] LLMs will often use techniques called "text vectorization," which involves converting text into mathematical

[131] *See* Dang Anh-Hoang, Vu Tran & Le-Minh Nguyen, *Survey and Analysis of Hallucinations in Large Language Models: Attribution to Prompting Strategies or Model Behavior*, FRONTIERS IN ARTIFICIAL INTELLIGENCE, Sept. 30, 2025, at 2.

[132] Holmes Lybrand, *Ex-Fugees Rapper Pras Michel Found Guilty in Scheme to Help China Influence US Government*, CNN (Apr. 26, 2023, 5:35 PM), https://www.cnn.com/2023/04/26/politics/pras-michel-trial-verdict/index.html [https://perma.cc/PJ24-ENQR].

[133] Will Shanklin, *Fugees Rapper Pras Accuses His Lawyer of Using AI in Closing Arguments*, ENGADGET (Oct. 19, 2023, 2:53 PM), https://www.engadget.com/fugees-rapper-pras-accuses-his-lawyer-of-using-ai-in-closing-arguments-185311864.html [https://perma.cc/6UQG-UZCN].

[134] *Id.*

[135] *Id.*

[136] *About Us*, EYELEVEL, https://www.eyelevel.ai/company/about-company [https://perma.cc/M783-PEYZ].

ISSN 0041-9915 (print) 1942-8405 (online) ● DOI 10.5195/lawreview.2026.1167
http://lawreview.law.pitt.edu

notation, mapping those notations in a theoretical "word space" and converting it back to words at the output level.[137] Put another way, text is turned into numbers and back into words. This process lends itself to more productive manipulation. Imagine a word that is frequently used in criminal adjudication like "charge." Converting the word from text to vector allows us to assign a distinct notation to each permutation of charge, so that its use in the context of a criminal accusation (arbitrarily -0.334232) will carry a different value than its use to denote a debit (arbitrarily -0.45224) and different from its use to denote an aggressive forward maneuver (arbitrarily -0.65322).

Here, the text vectorization process is so distinct from human reason that it produces errors that are ostensibly random and certainly arbitrary. Compare this to the type of data-induced error that is available in machine learning or other predictive techniques. Where an algorithm might try to find a line of best fit in a scattered plot of known and ascertainable data, the generative tool instead transforms language into more manipulable components, rather than achieve anything approximating understanding. As a result, the errors can be jarring and seem altogether fabricated. For instance, when tasked with summarizing evidence in a criminal case, a generative AI system might confidently fabricate forensic findings or witness statements that were never presented in court. This does not occur due to gaps or flaws in its dataset in the conventional sense but rather because the model lacks genuine comprehension. Instead of drawing from an anchored knowledge base, it generates responses based on statistical relationships between words and phrases, producing what seems linguistically coherent but is factually erroneous. These hallucinations are particularly difficult to correct because they do not follow a consistent or predictable pattern, unlike errors in predictive AI, which can often be traced back to biases in training data or specific algorithmic miscalculations that can be systematically refined.

C. Evaluative AI Error

Where predictive AI uses algorithms to excavate patterns in existing data, and where generative AI uses data to create new outputs that may not necessarily reflect input data, evaluative AI uses probabilistic reasoning to make determinations that are specific to an isolated input. This type of AI is *evaluative* in the sense that it brings its artificially-intelligent capabilities to bear on a given item. Take, as one example, facial recognition technology, which is being increasingly used in criminal investigations. Ferguson explains the concept elegantly:

---

[137] Xi Yang, Kaiwen Yang, Tianxu Cui, Min Chen & Liyan He, *A Study of Text Vectorization Method Combining Topic Model and Transfer Learning*, PROCESSES, Feb. 11, 2022, at 2–4.

ISSN 0041-9915 (print) 1942-8405 (online) ● DOI 10.5195/lawreview.2026.1167
http://lawreview.law.pitt.edu

> The simple idea behind facial recognition is to have a computer program automatically match a digital image of a face with a similar digital image of a face in a stored database. To work, a computer, a program is run on existing digital photographs or video surveillance cameras turning images into a digital network of identifiable objects and faces.[138]

However, as Ferguson further explains, the technology is more complex than mere pattern matching. Individual faces are "broken down into identifiable component parts," enabling the technology to reconcile its observation of an individual feature against not just a database of whole images, but against specific features in those whole images (like noses and eyes).[139] Recent improvements in facial recognition technology have enabled consideration of "skin textures, shadows, three-dimensional models" and even combinations of some or all the above.[140] To perform these analyses, systems need to be populated with "faces, classify them, train the data, and test the training sets, so the systems can identify the overlapping nodal points of any face in the system."[141]

Facial recognition technology is evaluative in the sense that it is fundamentally focused on evaluating a single input—such as a face—and making a determination based on that input. It does not generalize or predict future data, nor does it generate new data. Instead, it takes an individual piece of information (a face, in this case) and assesses it according to predefined criteria, weighing the relevant features to make a decision about what it matches or who it corresponds to. This concept is similar to other evaluative processes, such as DNA matching. In DNA analysis, a sample is compared against a known database of sequences to determine whether the sample matches a specific individual.[142] The system evaluates the input (the DNA sample) and judges whether it matches or does not match another sample in the database.[143]

---

[138] Ferguson, *supra* note 97, at 1109–12.

[139] *Id.* at 1110.

[140] Andrew Guthrie Ferguson, *Video Analytics and Fourth Amendment Vision*, 103 TEX. L. REV. (forthcoming 2026) (available online at https://texaslawreview.org/video-analytics-and-fourth-amendment-vision/) [https://perma.cc/HXV7-244R].

[141] Ferguson, *supra* note 97, at 1112.

[142] *DNA Evidence: Basics of Analyzing*, U.S. DEP'T OF JUST., NAT'L INST. OF JUST. (2012), https://nij.ojp.gov/topics/articles/dna-evidence-basics-analyzing [https://perma.cc/7GCV-SY3D].

[143] *Id.*

ISSN 0041-9915 (print) 1942-8405 (online) ● DOI 10.5195/lawreview.2026.1167
http://lawreview.law.pitt.edu

An example that deeply implicates constitutional rights is "TrueAllele," a software system developed by Pittsburgh-based Cybergenetics Inc.[144] TrueAllele consists of two products that enjoy broad use in criminal investigation and adjudication: TrueAllele Databank, which "interpret[s] single-source reference samples," and TrueAllele Casework "for interpreting more complex DNA mixtures."[145] TrueAllele is the focal point of *People v. Wakefield*, a 2022 case in the Court of Appeals in New York that affirmed a murder conviction that was in part supported by TrueAllele's AI-enabled analysis of DNA.[146] In the underlying investigation, police obtained the defendant's DNA from various sources, such as the victim's clothing and an amplifier cord that was used as part of the crime.[147] There, the Court described TrueAllele's technology as follows:

> TrueAllele Casework System applies a continuous probabilistic genotyping method of calculating a likelihood ratio—using all of the information generated on the electropherogram, including peaks that fall below a laboratory's stochastic threshold. The likelihood ratio in its modern form was developed by Alan Turing during World War II as a code-breaking method. TrueAllele uses a probability model to assess the values of a genotype objectively. It does not consider a reference sample for any particular DNA profile. Following protocol, once the genotypes were inferred based on mathematical computations from all the data in the electropherograms, the system compared defendant's genotype to all of the statistical genotype possibilities and calculated likelihood ratios as to the presence of defendant's genotype. The ratios were exponentially greater than those generated by the methods employed by the Lab. Specifically, TrueAllele concluded that it was 5.88 billion times more probable that defendant was a contributor to the mixture on the amplifier cord than an unrelated black person, that it was 170 quintillion times more probable that defendant was a contributor to the mixture on the outside rear shirt collar than an unrelated black person, that it was 303 billion times more probable that defendant was a contributor to the mixture on the outside front shirt collar than an unrelated black person, and that it

[144] *Casework*, CYBERGENETICS, https://www.cybgen.com/products/casework [https://perma.cc/A5ED-XECM] (last visited Dec. 29, 2025); *Database*, CYBERGENETICS, https://www.cybgen.com/products/database/ [https://perma.cc/GAU4-QQE4] (last visited Dec. 29, 2025). For an academic evaluation of TrueAllele's method, see David W. Bauer, Nasir Butt, Jennifer M. Hornyak & Mark W. Perlin, *Validating TrueAllele® Interpretation of DNA Mixtures Containing up to Ten Unknown Contributors*, 65 J. FORENSIC SCIS. 380 (2020).

[145] People v. Wakefield, 195 N.E.3d 19, 23–24 (N.Y. 2022).

[146] *Id.* at 21.

[147] *Id.* at 21–22.

ISSN 0041-9915 (print) 1942-8405 (online) ● DOI 10.5195/lawreview.2026.1167
http://lawreview.law.pitt.edu

> was 56.1 million times more probable that defendant was a contributor to the mixture on the victim's dorsal forearm than an unrelated black person.[148]

Here, TrueAllele suggests its uses are predominantly evaluative because it considers the specific genetic markers in a given DNA sample, compares those markers to profiles in a known reference database, and calculates how likely it is that the sample came from a particular individual.[149] The system neither synthesizes new genetic information (as a generative AI might) nor attempts to forecast broader trends or patterns (as a predictive AI might).[150] Instead, TrueAllele's primary function is to make a probabilistic determination about whether the DNA in question matches a known profile.[151] By design, it weighs the evidence gleaned from the input sample (in this instance, the DNA mixture) against pre-established reference points (databases) to reach a concrete, single-case conclusion.[152] In other words, its "intelligence" is geared toward drawing an evaluative inference—here, the probability that the defendant's DNA is present in the relevant mixture—rather than extrapolating beyond the confines of the specific investigative question.

IV. ALIGNING HARMLESS ERROR AND THE AI ERROR TYPOLOGY

A. "Arbitrariness" as a Doctrinal Fault Line

In the previous section, I established a typology of AI applications in criminal adjudication: predictive AI, generative AI, and evaluative AI. Each has a different capacity to err. Predictive tools fundamentally err in ways more closely associated with human cognitive failure, though the fact that said errors often occur in a manner ensconced in advanced data science technique makes them sometimes harder to apprehend. However, these errors tend not to be *arbitrary* in the sense that they involve data or processing issues with extant information. In other words, failure is a function of the source material or the treatment of the source material. Generative AI failure is fundamentally related to its successes: The tools are aimed at performing more creative synthesis and therefore will make errors that are not necessarily traceable to any specific collection, construction, or treatment of underlying data sets. As such, the errors can be inconsistent, unauditable, and unpredictable.

---

[148] *Id.* at 22–23.

[149] *Id.* at 24.

[150] *Id.*

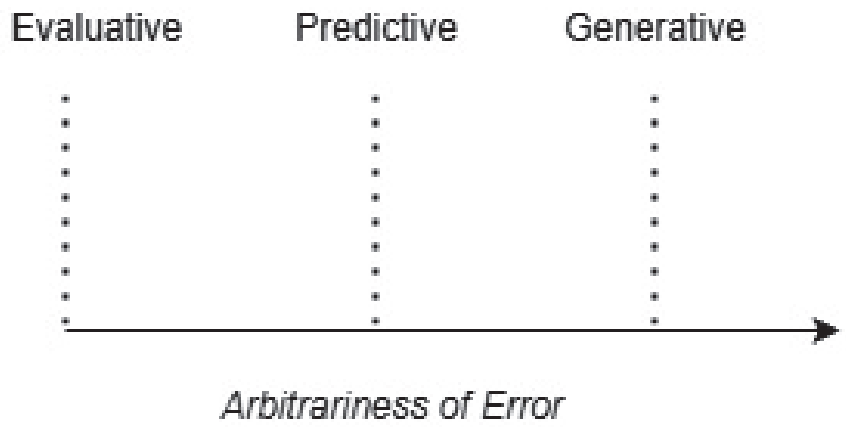
[151] *Id.*

[152] *Id.*

ISSN 0041-9915 (print) 1942-8405 (online) ● DOI 10.5195/lawreview.2026.1167
http://lawreview.law.pitt.edu

At this juncture, it is worth considering the interaction between this typology—wherein different types of AI produce different types of errors ranging from data-sourced to arbitrary—and the justifications for harmless error. Recall how Figure 5 below (previously Figure 4) illustrates how arbitrariness of error is distributed across the categories:

Figure 5



Generative AI errors are the most arbitrary. As an initial matter, generative AI is distinctly not deterministic; rather than operate in a rule-bound atmosphere, it aspires to creative acts like producing language—or, in the case of professional applications—work product. Because the ultimate item that a generative AI tool produces is informed by, but not rooted in, the underlying data that the technology is trained on, its information is not traceable in the traditional sense. Take the previously discussed example of large language model hallucination. If a generative AI application hallucinates information in a police report, or in defense counsel's brief or in its review of transcripts and interviews, those hallucinations are harder to disentangle because they are untethered to real, underlying information.

Here, consider the highly simplified but useful distinction between fabrication and interpretation. Where generative AI tools have the capacity to fabricate, predictive and evaluative AI tools have a problematic capacity to misinterpret. Think of fabrication as the ability to create new information (e.g., grammatically correct novel sentences but *also* utterly invented case law) and interpretation as efforts to contend and make sense of existing information. The information-bound nature of the latter means that even where predictive AI might miscalculate risk assessments based on historical biases, and evaluative AI might misinterpret DNA mixtures or facial recognition results, the mistakes can be unwound (or at least audited) with traditional faculties.

ISSN 0041-9915 (print) 1942-8405 (online) ● DOI 10.5195/lawreview.2026.1167
http://lawreview.law.pitt.edu

Errors made by evaluative AI are the least arbitrary. Where there might be a substantial disjuncture between a generative AI error and the normal distribution of correct outputs, and where predictive AI outcomes might be rooted in underlying data but misfire as a result of processing or analysis failures, evaluative AI errors tend to result from the inherent limitations of single-instance analysis—factors like the quality of the sample, the algorithm's method for weighting specific features, and the completeness of the reference database. In DNA matching, for example, a miscalibration in how a system interprets low-level or partial genetic markers can lead to incorrect conclusions about whether a suspect's DNA is present in a mixture. Because evaluative AI is tasked with making a determination on one discrete input rather than predicting future events or creating new data, any errors that occur typically stem from this narrowly focused process and remain confined to that specific evaluation. Once more, the errors that result from evaluative AI can be caused—and indeed, are likely to be caused—by incorrect computing, bad underlying data, and faulty model selection, for instance, but are highly unlikely to be random. Predictive AI tools, such as risk assessment algorithms and predictive policing models, similarly err in a much less arbitrary manner than generative AI tools and for largely the same reasons: Their reasoning techniques operate with more fidelity to underlying information, and so the universe of possible errors is fairly circumscribed.

Consider the relative arbitrariness of the errors in the typology of generative, predictive, and evaluative AI in light of the doctrinal justifications for harmless error review. At its core, harmless error review is about the extent to which reviewing courts can determine an error's impact. As Meltzer explains, harmless error conceives of two "remedial imperatives": (1) the imperative "to redress individual violations," and (2) the imperative to "reinforce structural values, including those underlying the separation of powers and the rule of law."[153] A prerequisite to discharging the court's remedial obligations is that the problems be assessable. Put differently, the imperatives as articulated by Meltzer will rely on an appellate court's ability to identify the impugned error and determine its relationship to both the substantial rights at issue and the ultimate outcome. If the error is arbitrary, assessment becomes proportionally more difficult. Where mistakes emanate from spontaneous distortions, as is potentially the case with a range of generative AI applications, a court's inquiry into both the nature of the error, its relationship to a constitutional right and its impact on the outcome—or the jury's pre-outcome assessment—is obfuscated. Consider, as one example, a generative AI tool used in a criminal case for transcription. The tool may have been used as part of the police investigation to supplant an officer's analog notetaking. In the event that the AI

[153] Daniel J. Meltzer, *Harmless Error and Constitutional Remedies*, 61 U. CHI. L. REV. 1, 24–25 (1994).

ISSN 0041-9915 (print) 1942-8405 (online) ● DOI 10.5195/lawreview.2026.1167
http://lawreview.law.pitt.edu

application hallucinates (potentially incriminating) statements, courts are placed in the position of having to determine the statement's veracity. Moreover, courts will have to stave off bad faith attempts at using generative AI's arbitrariness as a cajole; for instance, defendants who wish to cast aspersions on legitimately procured evidence on the grounds that it could be an AI hallucination.

Here, the traceability problems emerge in two distinct ways. First, the spontaneous nature of the fabricated material means that courts will have to assess the veracity of the transcript in the utter absence of corroborating material. A simple resolution would be to discard as "hallucinated" any material that has no other audio or textual source, but such a bright line would cast aspersions on legitimate witness statements for which there may not be backup, corroborating material. Second and perhaps more concerningly, the arbitrary fabrications could have colored other elements of the investigation. For instance, investigators may use fabricated evidence to guide discussion with other witnesses, or juries may consider fabricated material in relation to other, legitimate evidence. Here, the *raison d'etre* for harmless error review is upended; to the extent that the doctrine demands that courts ask whether an error is dispositive of the outcome, it requires being able to disentangle the relationship between that error and the outcome. Recall also that the arbitrariness of a generative AI error means that the distortion does not often bear the hallmarks of a falsity; it is offered confidently and therefore harder to detect in the first instance. Where fabricated evidence is commingled with legitimate evidence, for instance, errors become even less traceable.

The key insight here is that arbitrariness challenges courts' ability to assess and thus measure constitutional harms. The harmless error doctrine assumes this capacity; if an error is not structural, its impact on the ultimate outcome is considered and re-considered. Indeed, in *Gonzalez-Lopez*, Justice Scalia plainly stated that the Court rested its "conclusion of structural error upon the difficulty of assessing the effect of the error," making clear the extent to which assessing outcome-determinativeness is vital to harmless error review.[154] Yet even structural errors have to be assessed meaningfully. In *Neder*, Justice Scalia observed that constitutional error is a binary classification exercise; "a constitutional error is either structural or it is not." But the Court could not escape making a latent assessment of the severity of the error, noting that:

> Some structural errors, like the complete absence of counsel or the denial of a public trial, are visible at first glance. Others, like deciding whether the trial judge was biased or whether there was racial discrimination in the grand jury selection,

[154] United States v. Gonzalez-Lopez, 548 U.S. 140, 149 n.4 (2006).

ISSN 0041-9915 (print) 1942-8405 (online) ● DOI 10.5195/lawreview.2026.1167
http://lawreview.law.pitt.edu

> require a more fact-intensive inquiry. Deciding whether the jury made a finding "functionally equivalent" to the omitted or misdescribed element is similar to structural-error analysis of the latter sort.[155]

Compare this to AI errors that are less arbitrary, as in predictive or evaluative AI. There, the source of the errors is traceable—i.e., reside in the underlying data and/or the methodology used to treat the data—and therefore lend themselves more readily to assessment.

B. A Modest Proposal for Procedural Adjustment

Recognizing that AI error presents differently (either through generative, predictive or evaluative tools) and are of varying degrees of assessability (depending on their relative arbitrariness), what approaches should courts consider to accommodate the spectre of AI error? Here, I make a modest proposal: identifying the "AI error threshold." To identify the AI error threshold, courts need to answer two questions. First, does the error at issue emanate from a predictive or evaluative process—i.e., a process rooted in data? If yes, these errors lend themselves to some measure of scientific deconstruction and thus do not pose conceptual difficulties for traditional harmless error review. For instance, an incorrect facial recognition output presents as a binary failure (is it the defendant or not?) with an ascertainable bank of faces and a likely documented methodology tied to its probabilistic model. Moreover, the fact of its discrete output—here, suspect identification—means that judges have ordinary tools for disentangling its effects. Put simply, the technology makes a flawed classification the same way a person makes a flawed classification, and is thus remediable in the same fashion, albeit with more technical demands.

Second, does the error emanate from a spontaneous or creative AI output? In other words, is it a function of a generative application that cannot be accounted for by auditing underlying data? If yes, then the court is faced with an error that poses a similar challenge as the *Neder* or *Gonzalez-Lopez* courts vis-à-vis structural constitutional error, wherein "the difficulty of assessing the effect of the error" counsels against applying harmless error review.[156]

CONCLUSION

The increased use of artificial intelligence in criminal adjudication and these technologies' propensity to err places a great strain on the harmless error doctrine, constitutional criminal procedure's current method for absorbing mistakes. Harmless

[155] Neder v. United States, 527 U.S. 1, 37 (1999).

[156] *Gonzalez-Lopez*, 548 U.S. at 149 n.4.

ISSN 0041-9915 (print) 1942-8405 (online) ● DOI 10.5195/lawreview.2026.1167
http://lawreview.law.pitt.edu

error review's assumption that courts can make determinations about the impact of an error on both substantial rights and the ultimate outcome presupposes an auditability that AI does not always afford observers. Indeed, generative AI in particular can be opaque in this regard; to the extent that it has the ability to *create*, its truths and falsehoods are equally creative. For errors, this means being untethered to underlying data and therefore hard to either detect, correct, or assess. By contrast, errors made by predictive or evaluative AI tend to emanate from knowable data and methodological choices and thus remain amenable to analog harmless error review. Here, I propose thinking about the specter of AI error in terms of "arbitrariness," wherein the more arbitrary an error, the more scrutiny it demands as a prospectively structural error—and therefore definitionally harmful.

ISSN 0041-9915 (print) 1942-8405 (online) ● DOI 10.5195/lawreview.2026.1167
http://lawreview.law.pitt.edu