

INTERNATIONAL JOURNAL FOR LEGAL RESEARCH AND ANALYSIS



Open Access, Refereed Journal Multi Disciplinary
Peer Reviewed

www.ijlra.com

DISCLAIMER

No part of this publication may be reproduced, stored, transmitted, or distributed in any form or by any means, whether electronic, mechanical, photocopying, recording, or otherwise, without prior written permission of the Managing Editor of the *International Journal for Legal Research & Analysis (IJLRA)*.

The views, opinions, interpretations, and conclusions expressed in the articles published in this journal are solely those of the respective authors. They do not necessarily reflect the views of the Editorial Board, Editors, Reviewers, Advisors, or the Publisher of IJLRA.

Although every reasonable effort has been made to ensure the accuracy, authenticity, and proper citation of the content published in this journal, neither the Editorial Board nor IJLRA shall be held liable or responsible, in any manner whatsoever, for any loss, damage, or consequence arising from the use, reliance upon, or interpretation of the information contained in this publication.

The content published herein is intended solely for academic and informational purposes and shall not be construed as legal advice or professional opinion.

**Copyright © International Journal for Legal Research & Analysis.
All rights reserved.**

ABOUT US

The *International Journal for Legal Research & Analysis (IJLRA)* (ISSN: 2582-6433) is a peer-reviewed, academic, online journal published on a monthly basis. The journal aims to provide a comprehensive and interactive platform for the publication of original and high-quality legal research.

IJLRA publishes Short Articles, Long Articles, Research Papers, Case Comments, Book Reviews, Essays, and interdisciplinary studies in the field of law and allied disciplines. The journal seeks to promote critical analysis and informed discourse on contemporary legal, social, and policy issues.

The primary objective of IJLRA is to enhance academic engagement and scholarly dialogue among law students, researchers, academicians, legal professionals, and members of the Bar and Bench. The journal endeavours to establish itself as a credible and widely cited academic publication through the publication of original, well-researched, and analytically sound contributions.

IJLRA welcomes submissions from all branches of law, provided the work is original, unpublished, and submitted in accordance with the prescribed submission guidelines. All manuscripts are subject to a rigorous peer-review process to ensure academic quality, originality, and relevance.

Through its publications, the *International Journal for Legal Research & Analysis* aspires to contribute meaningfully to legal scholarship and the development of law as an instrument of justice and social progress.

PUBLICATION ETHICS, COPYRIGHT & AUTHOR RESPONSIBILITY STATEMENT

The *International Journal for Legal Research and Analysis (IJLRA)* is committed to upholding the highest standards of publication ethics and academic integrity. All manuscripts submitted to the journal must be original, unpublished, and free from plagiarism, data fabrication, falsification, or any form of unethical research or publication practice. Authors are solely responsible for the accuracy, originality, legality, and ethical compliance of their work and must ensure that all sources are properly cited and that necessary permissions for any third-party copyrighted material have been duly obtained prior to submission. Copyright in all published articles vests with IJLRA, unless otherwise expressly stated, and authors grant the journal the irrevocable right to publish, reproduce, distribute, and archive their work in print and electronic formats. The views and opinions expressed in the articles are those of the authors alone and do not reflect the views of the Editors, Editorial Board, Reviewers, or Publisher. IJLRA shall not be liable for any loss, damage, claim, or legal consequence arising from the use, reliance upon, or interpretation of the content published. By submitting a manuscript, the author(s) agree to fully indemnify and hold harmless the journal, its Editor-in-Chief, Editors, Editorial Board, Reviewers, Advisors, Publisher, and Management against any claims, liabilities, or legal proceedings arising out of plagiarism, copyright infringement, defamation, breach of confidentiality, or violation of third-party rights. The journal reserves the absolute right to reject, withdraw, retract, or remove any manuscript or published article in case of ethical or legal violations, without incurring any liability.

AN EXAMINATION OF EXPLAINABILITY AND DUE PROCESS CONCERNS IN THE USE OF ARTIFICIAL INTELLIGENCE IN JUDICIAL AND ADMINISTRATIVE DECISION MAKING

AUTHORED BY - OGHENEHORO EVI ENI

*Lawyer | Immigration & Policy Analyst (U.S) | Administrative & Public Law | Legal
Research, Drafting & Compliance | Fertility Law Expert*

Abstract

The expanding use of artificial intelligence in judicial and administrative decision making has introduced a set of due process questions that existing constitutional doctrine was not designed to answer. Courts and agencies increasingly rely on automated risk assessment instruments, eligibility algorithms, and predictive tools to inform decisions that affect liberty, property, and access to public benefits, yet the internal logic of these systems is frequently inaccessible to the very people whose interests are at stake. This article examines the tension between the opacity of many artificial intelligence systems and the procedural guarantees of notice, an opportunity to be heard, and a reasoned explanation that due process is understood to require. Drawing on the doctrinal framework established in Mathews v. Eldridge and its progeny, the article traces how American courts, most notably the Wisconsin Supreme Court in State v. Loomis, have responded to algorithmic decision making with caution rather than clear constraint, leaving important explainability questions unresolved. It then situates the American experience within a comparative frame by examining the transparency and human oversight obligations imposed by the European Union under the Artificial Intelligence Act and the General Data Protection Regulation. The article argues that a coherent framework for explainable and accountable automated decision making requires three complementary elements: a baseline right to a meaningful explanation calibrated to the significance of the interest at stake, structural obligations on system developers to build interpretability into design rather than treat it as an afterthought, and independent auditing mechanisms capable of verifying that automated tools perform as claimed. The article concludes that without such a framework, the promise of efficiency that artificial intelligence offers to courts and agencies risks being purchased at the expense of the very legitimacy that procedural fairness is meant to secure.

Keywords: *artificial intelligence, due process, explainability, algorithmic accountability, administrative law, judicial decision making.*

I. Introduction

Artificial intelligence has moved from the periphery of legal practice to the center of some of the most consequential decisions that governments make about individuals. Judges consult algorithmic risk scores when deciding whether to grant bail or how long a sentence should run. State agencies use automated systems to determine eligibility for unemployment benefits, Medicaid coverage, and housing assistance. Immigration authorities and law enforcement agencies increasingly rely on predictive tools to allocate scrutiny and resources. In each of these contexts, a system trained on historical data now performs work that was once entrusted exclusively to human judgment, and the outcomes it produces can be as consequential as a term of imprisonment or the loss of a family's only source of income¹.

The appeal of automation to courts and agencies is not difficult to understand. Caseloads are large, budgets are constrained, and algorithmic tools promise consistency and speed that human decision makers, however well trained, cannot always match. Yet the same qualities that make automated systems attractive, namely their capacity to process vast quantities of data through statistical models that no single person fully comprehends, also make them resistant to the kind of scrutiny that due process has traditionally demanded. A defendant sentenced in part on the basis of a proprietary risk score cannot easily interrogate how that score was calculated. A benefits recipient whose eligibility is terminated by an automated eligibility engine may receive a notice that states only that she no longer qualifies, without any indication of which factor changed or why². This gap between the complexity of algorithmic reasoning and the comprehensibility that legal process requires is the central subject of this article.

The term explainability, as used throughout this article, refers to the degree to which the internal logic of an automated system, and the specific factors that produced a particular output, can be rendered intelligible to a human decision maker, a reviewing court, and the individual affected by the decision. Explainability is distinct from transparency in the narrow

¹See generally Danielle Keats Citron, Technological Due Process, 85 Wash. U. L. Rev. 1249, 1249 to 1253 (2008) (describing the migration of adjudicatory functions from human caseworkers to automated systems across a range of government benefit programs).

²See Danielle Keats Citron & Frank Pasquale, The Scored Society: Due Process for Automated Predictions, 89 Wash. L. Rev. 1, 4 to 5 (2014) (arguing that automated scoring systems often deny affected individuals meaningful information about the basis for adverse decisions).

sense of disclosing source code or underlying data, since a system can be technically transparent yet functionally opaque if its statistical reasoning exceeds ordinary human comprehension³. It is this functional gap, rather than mere secrecy, that poses the more difficult constitutional and administrative law problem, because even a fully disclosed algorithm may remain unintelligible to the judge, caseworker, or litigant who must rely on it.

II. Due Process Foundations, From *Goldberg* to *Mathews*

The Due Process Clauses of the Fifth and Fourteenth Amendments provide that no person shall be deprived of life, liberty, or property without due process of law. The Supreme Court has long recognized that the content of that guarantee is not fixed but varies according to the nature of the interest at stake and the risk that existing procedures will produce an erroneous outcome⁴. In *Goldberg v. Kelly*, the Court held that recipients of public assistance were entitled to an evidentiary hearing before the termination of benefits, reasoning that the loss of welfare payments could deprive a family of the very means of subsistence while an appeal was pending⁵. Six years later, in *Mathews v. Eldridge*, the Court declined to extend that pretermination hearing requirement to Social Security disability benefits, concluding that a post termination hearing satisfied due process given the character of the interest involved and the relatively low risk of an erroneous initial determination⁶.

The doctrinal legacy of *Mathews* is the three factor balancing test that continues to govern procedural due process analysis today. Courts must weigh, first, the private interest affected by the government action, second, the risk of an erroneous deprivation through the procedures used and the probable value of additional or substitute safeguards, and third, the government's interest, including the administrative and fiscal burdens that additional procedures would impose⁷. This balancing framework was designed for a world in which the relevant question was how much process a human decision maker owed to an individual before or after acting. It assumes, implicitly, that the accuracy of a decision can be improved through familiar procedural devices such as notice of the evidence relied upon, an opportunity to submit contrary evidence, and a statement of reasons capable of being tested on appeal.

Automated decision making complicates each element of that assumption. The private

³See Joshua A. Kroll et al., *Accountable Algorithms*, 165 U. Pa. L. Rev. 633, 638 to 640 (2017) (distinguishing transparency of code from meaningful accountability and arguing that disclosure alone does not guarantee comprehension).

⁴*Mathews v. Eldridge*, 424 U.S. 319, 334 (1976).

⁵*Goldberg v. Kelly*, 397 U.S. 254, 264 (1970).

⁶*Mathews*, 424 U.S. at 349.

⁷*Mathews*, 424 U.S. at 335.

interest at stake, whether liberty in a sentencing context or a benefit payment in an administrative one, is unchanged by the fact that a machine rather than a person made or informed the decision, and if anything the scale at which automated systems operate can multiply the aggregate interests affected across thousands of similarly situated individuals⁸. The second *Mathews* factor, the risk of erroneous deprivation and the value of additional safeguards, becomes considerably harder to evaluate when the decision maker's reasoning process is not fully intelligible even to the officials who rely on it. A safeguard such as a hearing has diminished value if the hearing officer cannot meaningfully explain, question, or correct the automated output that forms the basis of the challenged decision⁹. The third factor, the government's administrative interest, is precisely the interest that automation is meant to serve, which creates a structural tension: the very efficiency gains that justify adopting an algorithmic tool are the ones most threatened by procedural safeguards robust enough to render that tool's operation explainable.

Professor Danielle Keats Citron was among the first scholars to argue that this tension required a distinct doctrinal response, which she termed technological due process. Citron observed that automated systems collapse the traditional distinction between adjudication, which affords individualized process, and rulemaking, which affords broader public participation, because programmers effectively make policy decisions when translating legal standards into code, yet those decisions escape the scrutiny that either adjudicatory hearings or notice and comment rulemaking would ordinarily provide¹⁰. Her proposed remedy centered on audit trails, structured testing before deployment, and mechanisms that would allow both affected individuals and outside reviewers to reconstruct how a given automated decision was reached¹¹. Citron and Pasquale later extended this analysis to privately developed scoring systems used by government and industry alike, proposing an inspection right that would allow individuals to examine and correct the data underlying an adverse algorithmic determination¹². These proposals remain largely aspirational. No federal statute currently affords a general right to inspect or contest the logic of a government algorithm, and constitutional due process doctrine, rooted in the flexible *Mathews* balancing test, has not yet been authoritatively extended to require explainability as a distinct procedural safeguard.

⁸See Citron, *supra* note 1, at 1281 to 1283 (discussing how automation can propagate a single design error across an entire class of beneficiaries simultaneously).

⁹See Citron & Pasquale, *supra* note 2, at 18 to 20 (arguing that hearing officers often defer to automated scores rather than subject them to independent scrutiny).

¹⁰Citron, *supra* note 1, at 1256 to 1257.

¹¹*Id.* at 1305 to 1308.

¹²Citron & Pasquale, *supra* note 2, at 25 to 28.

III. The Rise of Algorithmic Decision Making in Judicial and Administrative Contexts

Automated tools now inform decisions across nearly every stage of the criminal justice process. Pretrial risk assessment instruments generate scores that predict the likelihood a defendant will fail to appear or reoffend if released, and judges in many jurisdictions consult these scores when setting bail or conditions of release. At sentencing, tools such as the Correctional Offender Management Profiling for Alternative Sanctions instrument, commonly known as COMPAS, combine static criminal history data with dynamic factors drawn from a defendant's responses to a structured questionnaire to produce a numerical risk classification¹³. Investigative journalism examining COMPAS scores used in Broward County, Florida, reported that the tool was substantially more likely to falsely flag Black defendants as high risk for future crime than white defendants, and substantially more likely to falsely flag white defendants as low risk, a finding that generated a sustained academic debate over how fairness should be defined and measured in predictive criminal justice tools¹⁴. Whatever position one takes in that methodological debate, the episode illustrates a recurring problem, namely that the statistical properties of a proprietary tool are often unknown to the courts that rely on it until outside researchers manage to reconstruct them from public records.

Administrative agencies have adopted automated decision making with similar enthusiasm and similarly uneven results. Beginning in 2013, the Michigan Unemployment Insurance Agency deployed an automated system known as the Michigan Integrated Data Automated System, or MiDAS, to detect suspected fraud among unemployment benefit claimants. The system flagged tens of thousands of claimants for fraud, often based on data mismatches or missed multiple choice questionnaires, and issued automatic determinations imposing steep penalties without meaningful human review¹⁵. A subsequent state review found that the overwhelming majority of these automated fraud determinations were incorrect, and affected claimants brought suit alleging that the system deprived them of property without the notice and opportunity to respond that due process requires. The United States Court of Appeals for the Sixth Circuit allowed several of these claims to proceed, reasoning that the plaintiffs had plausibly alleged a due process violation where the automated determinations were issued

¹³See *State v. Loomis*, 881 N.W.2d 749, 753 to 754 (Wis. 2016) (describing the COMPAS instrument and its use in a presentence investigation report).

¹⁴Julia Angwin et al., *Machine Bias*, ProPublica, May 23, 2016 (analyzing COMPAS risk scores assigned to defendants in Broward County, Florida, and finding significant racial disparities in error rates).

¹⁵See *Cahoo v. SAS Analytics Inc.*, 912 F.3d 887, 894 to 897 (6th Cir. 2019) (recounting the operation of the MiDAS system and allegations that it issued fraud determinations without adequate human review or notice).

with minimal individualized notice and no meaningful opportunity to contest the underlying findings before penalties attached¹⁶.

A comparable pattern emerged in Idaho, where the state Medicaid agency adopted a computerized formula to calculate the budget allotments available to individuals with developmental disabilities under a home and community based waiver program. Plaintiffs receiving reduced allotments alleged that the state provided no explanation of how the formula generated their specific budget figures, making it impossible to determine whether an error had occurred or to mount an effective challenge. The United States Court of Appeals for the Ninth Circuit held that due process required the state to disclose the data and reasoning underlying an individual budget calculation, reasoning that a generalized description of the formula was not a substitute for information sufficient to allow a recipient to identify and contest a specific error¹⁷. Together, the Michigan and Idaho litigation demonstrate that when automated systems replace individualized casework entirely, courts have been willing to find due process violations, but only after the harm has already been inflicted on large numbers of people and only through protracted litigation rather than through any systematic requirement of explainability built into the design of the system itself.

IV. Explainability as a Legal and Technical Problem

Explainability is sometimes treated in legal commentary as a single, binary attribute that a system either has or lacks, but the underlying technical literature describes a spectrum of interpretability that depends heavily on the type of model involved. Simple linear models and decision trees can often be explained in terms a lay person can follow, because the relationship between an input factor and the resulting output is direct and can be stated in ordinary language. More complex models, including ensemble methods and neural networks, achieve greater predictive accuracy in many settings precisely because they capture nonlinear relationships among dozens or hundreds of variables that resist any simple verbal description¹⁸. This tradeoff between accuracy and interpretability is not a temporary limitation that better engineering will necessarily resolve. For some classes of problems, the most accurate available model is also among the least interpretable, which means that a legal requirement of explainability may, at

¹⁶Cahoo, 912 F.3d at 901 to 905.

¹⁷K.W. ex rel. D.W. v. Armstrong, 789 F.3d 962, 972 to 975 (9th Cir. 2015).

¹⁸See Cary Coglianese & David Lehr, *Regulating by Robot*, *Administrative Decision Making in the Machine Learning Era*, 105 Geo. L.J. 1147, 1156 to 1160 (2017) (describing the tradeoff between predictive accuracy and interpretability across different classes of machine learning models).

least at present, come at some cost to predictive performance¹⁹.

Computer scientists have developed a range of techniques intended to approximate an explanation for the output of an otherwise opaque model, including methods that identify which input features most influenced a particular prediction and methods that construct a simplified surrogate model to approximate the behavior of a more complex one in the vicinity of a specific case. These techniques can provide useful information, but they are approximations rather than a full account of the model's internal reasoning, and researchers have shown that different explanation techniques applied to the same model and the same output can produce inconsistent or even contradictory accounts of which factors mattered most²⁰. This instability matters a great deal in a legal setting, because a due process regime that offers an individual an explanation which does not reliably reflect the actual reasoning behind the challenged decision provides only the appearance of process rather than its substance.

The legal system has additional reasons to distinguish between global explainability, meaning an account of how a model behaves across the full range of cases it might encounter, and local explainability, meaning an account of why the model reached a particular result in an individual case. Due process, oriented as it is toward the individual litigant or claimant, is primarily concerned with local explainability, since a defendant contesting a risk score or a claimant contesting a benefits termination needs to understand the specific reasons for the decision affecting them rather than a general description of the system's overall design. Much of the transparency that vendors and agencies have offered to date, by contrast, has taken the form of global documentation, such as technical manuals describing a tool's development and validation, which does little to help an individual identify what went wrong in a specific case²¹.

V. Judicial Reliance on Opaque Risk Assessment Tools, *State v. Loomis*

The most significant American judicial engagement with the due process implications of algorithmic risk assessment arose from the sentencing of Eric Loomis, who pleaded guilty to charges arising from a drive by shooting in Wisconsin. The presentence investigation report prepared for Loomis's sentencing hearing included a COMPAS risk assessment indicating that he presented a high risk of recidivism, and the sentencing court referenced this assessment

¹⁹Coglianese & Lehr, *supra* note 15, at 1160 to 1163.

²⁰Kroll et al., *supra* note 3, at 657 to 660 (surveying interpretability techniques and cautioning that post hoc explanations do not necessarily reveal the true basis for a model's decision).

²¹See Citron & Pasquale, *supra* note 2, at 13 to 15 (distinguishing between disclosure of a scoring system's general methodology and disclosure sufficient to allow an individual to contest a specific score).

among the factors supporting a sentence within the maximum available range²². Loomis moved for a new sentencing hearing, arguing that reliance on the COMPAS assessment violated his due process rights because the proprietary nature of the underlying algorithm prevented him from challenging its scientific validity, because the tool relied on group based data rather than an individualized assessment, and because it improperly considered gender as a factor²³.

The Wisconsin Supreme Court rejected each of these arguments but did so in terms that reveal the limits of existing due process doctrine when applied to automated tools. The court reasoned that Loomis had access to the same COMPAS report that the sentencing court reviewed and therefore had an opportunity to challenge the accuracy of the information it contained, even though he could not access the proprietary methodology by which that information was converted into a numerical risk score²⁴. The court further reasoned that because the circuit court stated it would have imposed the same sentence even without considering the COMPAS assessment, any error in relying on the tool was not determinative and therefore did not require resentencing²⁵.

Perhaps most significantly for the purposes of this article, the Wisconsin Supreme Court did not hold that due process requires disclosure of an algorithm's internal methodology. Instead, the court crafted a set of procedural cautions, directing that any presentence investigation report incorporating a COMPAS assessment include written advisories informing the sentencing court that the tool's proprietary nature prevents disclosure of how risk scores are calculated, that COMPAS scores have not been cross validated for a Wisconsin population, that some studies have raised questions about whether such tools disproportionately classify minority offenders as higher risk, and that the tool was designed to assess group recidivism rates rather than to predict the risk that a specific individual will reoffend²⁶. The court also cautioned that COMPAS scores may not be used as the determinative factor in deciding whether to incarcerate an offender or in determining the severity of a sentence²⁷.

These cautions represent a meaningful acknowledgment of the risks that opaque algorithmic tools pose, but they fall well short of an explainability requirement. A warning that a tool's methodology cannot be disclosed is not the same as a right to understand why a particular score was assigned, and a rule that a score may not be the sole determinative factor does not address the more fundamental difficulty that neither the sentencing judge nor the

²²State v. Loomis, 881 N.W.2d 749, 754 to 755 (Wis. 2016).

²³Id. at 757.

²⁴Id. at 761 to 762.

²⁵Id. at 767 to 769.

²⁶Id. at 769 to 770.

²⁷Id. at 772.

defendant can verify how heavily the score actually influenced the sentence that was imposed. Commentators have observed that the practical effect of *Loomis* is to shift the burden onto defendants to demonstrate misuse of a tool whose inner workings remain inaccessible to them, a burden that is difficult if not impossible to satisfy without access to the underlying algorithm²⁸. The United States Supreme Court denied certiorari in the related case seeking further review, leaving *Loomis* as the leading, and largely unchallenged, statement of American constitutional doctrine on this question²⁹.

The caution that Wisconsin courts have applied to COMPAS has not been uniformly adopted elsewhere, and the tool continues to be used in jurisdictions including New York, Pennsylvania, California, and Florida, often without the specific advisories that the Wisconsin Supreme Court required³⁰. The result is a patchwork of practice in which the degree of procedural protection an individual receives depends substantially on the jurisdiction in which sentencing occurs, a state of affairs that sits uneasily with the constitutional promise that due process protections should not vary arbitrarily by geography for a right as fundamental as the liberty interest at stake in a criminal sentence.

VI. Automated Administrative Decision Making and the Erosion of Individualized Process

If judicial reliance on algorithmic tools raises due process concerns in individual cases, automated administrative decision making raises them at scale. Unlike a sentencing court, which retains ultimate discretion and reviews a risk score alongside other case specific information, many administrative systems are designed to issue determinations with minimal or no human review, precisely because the volume of claims processed makes individualized review impractical³¹. The Michigan and Idaho litigation discussed above illustrates the consequences of this design choice when it goes wrong. In Michigan, the MiDAS system issued fraud determinations based on automated cross referencing of data that frequently proved unreliable, and because the system was designed to issue determinations automatically, tens of thousands of claimants experienced wage garnishment, seizure of tax refunds, and referral for

²⁸See, for example, Sarah Kolo, *State v. Loomis*, Wisconsin Sentencing Court Considers Algorithmic Risk Assessment in Sentencing, 33 Berkeley Tech. L.J. 539, 550 to 554 (2018) (arguing that the procedural safeguards adopted in *Loomis* do not adequately address the opacity of proprietary risk assessment tools).

²⁹See *Loomis v. Wisconsin*, 137 S. Ct. 2290 (2017) (denying certiorari).

³⁰See Kolo, *supra* note 21, at 541 to 543 (surveying jurisdictions using COMPAS or comparable risk assessment tools at sentencing).

³¹See Coglianese & Lehr, *supra* note 15, at 1180 to 1183 (describing the efficiency rationale that leads agencies to adopt largely automated adjudication for high volume benefit programs).

criminal prosecution before human caseworkers meaningfully reviewed their cases³².

The structural feature common to both the Michigan and Idaho cases is what might be termed the collapse of individualized notice into generic notice. Due process doctrine has long required that notice be reasonably calculated to apprise an individual of the pendency of an action and afford an opportunity to present objections³³. When a human caseworker denies a benefit, the caseworker ordinarily identifies the specific factual basis for the denial because that basis is what the caseworker herself relied upon. When an algorithm denies a benefit, the notice generated by the system often states only the output of the calculation, namely that eligibility has changed or that fraud has been detected, without necessarily identifying which input or combination of inputs produced that output, because the system itself may not have been designed to make that information available in a form suitable for a plain language notice³⁴. This is not merely a drafting failure that better notice letters could cure. It reflects a deeper design choice about what information the system captures and retains in a form that can later be communicated to an affected individual.

The Ninth Circuit's decision in *K.W. v. Armstrong* speaks directly to this problem. The court rejected the state's argument that publishing the general formula used to calculate Medicaid waiver budgets satisfied due process, holding instead that recipients were entitled to the specific data points that the formula applied to their individual circumstances, because only that individualized information would allow a recipient to identify a data entry error, a miscalculation, or an eligibility determination based on outdated information³⁵. This reasoning suggests a workable, if demanding, standard, namely that explainability in the administrative context requires not an account of how a system works in general but a case specific account of how it produced the particular result being challenged, sufficient to allow the individual to identify and contest a specific error.

Applying this standard consistently across the range of automated systems now in use would require significant changes to how agencies design and document their decision processes. Many systems currently in use were procured from private vendors under contracts that treat the underlying methodology as a trade secret, mirroring the posture the state of Wisconsin took in defending the use of COMPAS at sentencing³⁶. Absent a legal requirement

³²Cahoo, 912 F.3d at 895 to 897.

³³See *Mullane v. Central Hanover Bank & Trust Co.*, 339 U.S. 306, 314 (1950).

³⁴See Citron, *supra* note 1, at 1289 to 1291 (describing how automated notices in public benefits systems frequently fail to specify the particular factual basis for an adverse determination).

³⁵*K.W.*, 789 F.3d at 973 to 974.

³⁶See Robert Brauneis & Ellen P. Goodman, *Algorithmic Transparency for the Smart City*, 20 *Yale J.L. & Tech.* 103, 122 to 126 (2018) (documenting widespread reliance on trade secret protections to resist disclosure requests concerning government algorithmic systems).

that procurement contracts preserve the government's ability to disclose case specific information to affected individuals, agencies may find themselves unable to satisfy due process even where they wish to do so, because the technical capacity to generate an individualized explanation was never built into the system and the contractual terms under which the system was purchased may forbid disclosure of the information necessary to reconstruct one after the fact.

VII. Comparative Perspectives: The European Union's Response

The European Union has approached the explainability problem through a more comprehensive regulatory architecture than currently exists in the United States, combining data protection law with a dedicated artificial intelligence statute. Article 22 of the General Data Protection Regulation provides that a data subject has the right not to be subject to a decision based solely on automated processing, including profiling, which produces legal effects concerning the data subject or similarly significantly affects the data subject, subject to certain exceptions³⁷. Where an exception applies, the data controller must implement suitable safeguards, including at minimum the right to obtain human intervention, to express one's point of view, and to contest the decision³⁸. Commentators have debated at length whether the regulation's recitals establish a free standing right to an explanation of automated decisions, since the operative text of Article 22 does not use that phrase explicitly, though the recitals reference a right to obtain an explanation of the decision reached after such assessment³⁹.

The Artificial Intelligence Act, formally Regulation 2024/1689, adopts a risk based approach that imposes graduated obligations depending on the potential harm a given application of artificial intelligence poses. Systems used in the administration of justice and in the management of public benefits and services fall within the Act's high risk category, triggering a set of obligations concerning data governance, technical documentation, transparency, and human oversight⁴⁰. Article 13 requires that high risk systems be designed and developed so that their operation is sufficiently transparent to enable deployers to interpret a system's output and use it appropriately, and requires that such systems be accompanied by

³⁷Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 (General Data Protection Regulation), art. 22.

³⁸Id. art. 22(3).

³⁹Regulation (EU) 2016/679, recital 71; see Sandra Wachter, Brent Mittelstadt & Luciano Floridi, Why a Right to Explanation of Automated Decision Making Does Not Exist in the General Data Protection Regulation, 7 Int'l Data Priv. L. 76, 78 to 82 (2017) (arguing that the operative provisions of the regulation do not clearly establish an enforceable right to explanation).

⁴⁰Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 (Artificial Intelligence Act), art. 6 and annex III.

instructions for use containing information about the system's characteristics, capabilities, and limitations⁴¹. Article 14 requires that high risk systems be designed to allow for effective human oversight, including measures that enable the individuals responsible for that oversight to correctly interpret a system's output and, where appropriate, to decide not to use the system or to disregard, override, or reverse its output⁴².

The European framework is not without its own limitations. Article 13 requires transparency sufficient to enable a deployer, meaning the entity operating the system, to interpret its output, but does not on its face require that individuals affected by a given decision receive an explanation calibrated to their own circumstances in the way that the Ninth Circuit required in *K.W. v. Armstrong*. Scholars examining the drafting history of the Act have observed that earlier versions of the transparency provisions were narrowed during negotiation, shifting the emphasis from empowering the end user to understand a specific decision toward ensuring that the professional deploying the system understands its general capabilities and limitations⁴³. Nonetheless, the combination of the Act's human oversight requirements and the General Data Protection Regulation's rights to obtain human intervention and to contest an automated decision provides a more explicit statutory foundation than currently exists in American law, where explainability obligations must be inferred from a constitutional balancing test never designed with automated systems in mind.

The comparison also reveals a structural difference in regulatory philosophy that is worth making explicit. American due process doctrine operates retrospectively, through litigation initiated by an individual who has already suffered an adverse decision and who must establish, often years after the fact and at considerable cost, that the process afforded was constitutionally inadequate, as occurred in both the Michigan and Idaho litigation described above. The European framework, by contrast, operates prospectively, imposing design and documentation obligations before a high risk system may be placed into service at all, backed by administrative enforcement rather than relying solely on private litigation⁴⁴. A prospective regulatory model of this kind offers the possibility of catching design flaws before they are deployed at scale, rather than after tens of thousands of individuals have already been affected, though it also depends on the existence of a regulatory body with the technical capacity and

⁴¹Id. art. 13.

⁴²Id. art. 14.

⁴³See Clara Rachel, Bridging the Transparency Gap, What Can Explainable Artificial Intelligence Learn from the Artificial Intelligence Act, arXiv preprint (2023), at 6 to 9 (tracing the evolution of the transparency provisions through successive drafts of the Artificial Intelligence Act).

⁴⁴See Regulation (EU) 2024/1689, arts. 16, 43, and 71 (establishing conformity assessment procedures and administrative penalties for noncompliance).

political will to conduct meaningful conformity assessments, a capacity that remains largely untested in practice.

VIII. Toward a Framework for Explainable and Accountable Decision Making

The doctrinal and comparative analysis above suggests that a coherent response to the explainability problem should rest on three complementary elements, none of which is fully realized in either the American or the European approach as currently constituted.

A. A Calibrated Right to Explanation

The first element is a right to explanation whose content varies according to the significance of the interest at stake, consistent with the flexible structure of the *Mathews* balancing test rather than as a rigid, one size fits all rule. Where an automated tool contributes to a decision affecting physical liberty, such as a pretrial detention or sentencing determination, the individual should be entitled to a case specific account of the factors that most influenced the tool's output, sufficient to allow meaningful cross examination of the tool's reliability in that particular case, along the lines the Ninth Circuit required in the Medicaid budgeting context⁴⁵. Where an automated tool merely assists in routine administrative processing of low stakes matters, a less demanding standard focused on the accuracy of the underlying data and the availability of human review upon request may suffice. This calibrated approach avoids imposing uniform and potentially prohibitive explainability burdens on every application of artificial intelligence in government, while ensuring that the most consequential decisions receive the most rigorous scrutiny.

B. Design Obligations for Interpretability

The second element addresses the point, discussed in Part VI, that explainability cannot be added to a system after the fact if the system was never designed to capture the information necessary to generate an individualized explanation. Procurement rules governing government acquisition of automated decision systems should require that vendors build in logging and interpretability features from the outset, and should prohibit contractual terms that would allow a vendor's trade secret interest to override the government's ability to satisfy its constitutional

⁴⁵See K.W., 789 F.3d at 973 to 974; Kroll et al., *supra* note 3, at 692 to 694 (proposing that the rigor of an explanation requirement should scale with the severity of the decision at issue).

obligations to the individuals a given system affects⁴⁶. This obligation parallels the European Union's Article 13 and Article 14 requirements, but should be strengthened, consistent with the recommendation in Part IX below, to require that the interpretability features built into a system be sufficient to generate the individualized, case specific explanations that Part A above describes, not merely the general operating instructions that satisfy the current text of the Artificial Intelligence Act.

C. Independent Auditing

The third element responds to the pattern, visible in both the ProPublica investigation of COMPAS and the eventual state review of the MiDAS system, in which serious flaws in an automated tool came to light only through the efforts of outside researchers or investigative journalists working with limited access to the underlying system, well after the tool had already affected large numbers of people⁴⁷. A framework for accountable automated decision making should include a mechanism for independent, pre deployment and periodic post deployment auditing, conducted by reviewers with genuine access to system data and methodology and with the technical expertise to evaluate claims about accuracy and fairness, reporting to a body with authority to require corrective action or to suspend a system's use pending correction. The European Union's conformity assessment procedures under the Artificial Intelligence Act represent a step toward this model, though their effectiveness will depend heavily on the resources and independence of the bodies charged with conducting these assessments⁴⁸.

IX. Recommendations

- The legislature should codify a calibrated right to explanation for government use of automated decision systems, specifying the minimum content of an adequate explanation according to the category of interest affected, rather than leaving this question to be worked out case by case through constitutional litigation under the general *Mathews* framework. Such legislation could draw on the individualized disclosure standard articulated in *K.W. v. Armstrong* as a floor for decisions affecting benefits eligibility, and could require an even more demanding standard for decisions affecting physical liberty.

⁴⁶See Brauneis & Goodman, *supra* note 27, at 148 to 152 (proposing procurement reforms that would preserve government access to information necessary for algorithmic transparency notwithstanding vendor trade secret claims).

⁴⁷See Angwin et al., *supra* note 12; Cahoo, 912 F.3d at 895 to 896 (describing the state review that ultimately found the overwhelming majority of MiDAS fraud determinations to be incorrect).

⁴⁸Regulation (EU) 2024/1689, arts. 43 and 74.

- Procurement statutes and regulations should require that any automated decision system purchased or licensed by a government agency include, as a condition of the contract, the technical capacity to generate case specific explanations and should void any contractual term that would prevent the agency from disclosing the information necessary to satisfy this obligation when a due process challenge arises.
- The courts applying the *Mathews* balancing test to automated decision making should treat the absence of a case specific, reasonably intelligible explanation as itself a factor weighing heavily in the second *Mathews* factor, the risk of erroneous deprivation, rather than treating the mere existence of an opportunity to be heard as sufficient regardless of whether the individual can meaningfully use that opportunity to challenge the actual basis for the decision.
- The jurisdictions that continue to permit judicial reliance on proprietary risk assessment tools at sentencing should, at minimum, adopt cautions comparable to those required in *State v. Loomis*, and should consider whether the Wisconsin approach goes far enough, given the persuasive academic criticism that a warning about a tool's limitations does not substitute for a genuine opportunity to interrogate the basis for a specific score.

X. Conclusion

The introduction of artificial intelligence into judicial and administrative decision making has outpaced the development of legal doctrine equipped to address it. *Mathews v. Eldridge* offers a flexible framework capable in principle of accommodating new forms of decision making, but its application to automated systems has so far produced cautious, incremental results rather than a clear standard. *State v. Loomis* illustrates both the promise and the limits of judicial engagement with this problem, acknowledging the risks that opaque risk assessment tools pose while stopping well short of requiring the kind of individualized explanation that would allow a defendant to meaningfully contest a score's accuracy. The Michigan and Idaho litigation demonstrates that when automated systems are permitted to operate with minimal human oversight, the consequences of design flaws can affect tens of thousands of people before any correction occurs, and that courts, when finally confronted with the resulting harm, have been willing to require the kind of individualized, case specific disclosure that this article argues should be built into these systems from the outset rather than extracted from them years later through litigation.