Volume 1

Stevens Journal of Law and Technology

Inaugural Edition

Table of Contents

Masthead

Nicholas Russo
Founder and Editor-in-Chief
Stevens Institute of Technology Class of 2026
Harvard Law School Class of 2031

Damon Servidio
Co-Founder and Editor
Stevens Institute of Technology Class of 2026
New York Law School Class of 2029

Zaina Matahen
Editor
Stevens Institute of Technology Class of 2028

Joaquin Santander
Editor
Stevens Institute of Technology Class of 2028

Rimmo Loyi Lego
Editor
Stevens Institute of Technology Class of 2026

Editor's Note

Dear Reader,

I founded the Stevens Journal of Law and Technology after recognizing that scholarly platforms dedicated to the intersection of law and technology remain largely inaccessible outside of law schools. Such barriers omit the valuable voices of early-career scholars and professionals as these domains grow increasingly relevant. SJOLT exists to address that gap.

I believe that expanding access to this domain empowers emerging voices to contribute meaningfully to dialogues that will shape policy and practice for decades to come.

This inaugural volume reflects that belief. The pieces published here demonstrate the quality and depth expanded access adds to key conversations. It is my hope that every SJOLT author—present and future— derives meaningful intellectual and professional benefit from publication and that readers will find in these pages rigorous inquiry into questions that matter.

Per aspera ad astra.

Nicholas Russo

Founder and Editor-in-Chief

Mission Statement

The *Stevens Journal of Law and Technology (SJOLT)* is the first and only undergraduate-led academic publication dedicated to advancing scholarship at the intersection of law and technology. We publish original long-form pieces by undergraduate and graduate students worldwide, providing a platform for encouraging and distributing interdisciplinary inquiry and argumentation. Through this work, SJOLT seeks to elevate emerging voices, expand access to legal and technological scholarship, and advance rigorous discourse on the complex relationships between law, emerging technologies, and society.

Disclaimer

The views and opinions presented in this journal are solely those of the respective authors and do not necessarily reflect the official policy or position of the Stevens Journal of Law and Technology, its editors, or the Stevens Institute of Technology.

Digital Diagnostics and the Dilemma of Consent: Rethinking Patient Autonomy with AI-Powered Medicine

Rimmo Loyi Lego

" OPACITY IS THE FOUNDATION FOR EPISTEMIC INJUSTICE, LEAVING PATIENTS SYSTEMATICALLY DISADVANTAGED. "

Abstract

This paper critically assesses the threat AI-powered diagnostics presents to US healthcare's traditional informed consent dogma. Derived from traditional cases like *Canterbury v. Spence* (1972) and *Schloendorff* (1914), the research reveals how opaque, algorithmic decision-making undermines patient autonomy and raises legal liability. Weaving together doctrinal and philosophical analysis, the paper contends the necessity of normative reforms—such as legislating algorithmic explainability into law and regulation—to modernize statutory and regulatory codes. Comparative EU lessons in data protection and medical device regulation highlight the need for proactive U.S. legislative changes. The work invites additional empirical studies on patient outcomes and satisfaction in AI-supported diagnostic procedures to guarantee ethical and effective clinical practice.

Introduction

The arrival of artificial intelligence (AI) in medicine has caused a revolution in medical diagnosis and decision-making, upending ingrained clinical routines and legal standards. [1] Exponentially increased advancements in computing, data analysis, and machine learning over the last two decades have allowed AI systems to surpass conventional diagnostic tools in radiology, pathology, and precision medicine. [2] These systems are progressively being implemented in clinical workflows, with the prospect of quicker, more precise diagnosis and better patient outcomes. [3] This accelerated technology advance, however, also provokes fundamental ethical, legal, and philosophical concerns surrounding patient autonomy and understanding of informed consent—a foundational element of medical ethics and American healthcare law.

In the past, informed consent doctrine relied on the direct face-to-face interaction between patients and human physicians. In this classical model, informed consent is dependent on an open, communicative process in which a healthcare provider explains the risks, benefits, and alternatives of a proposed treatment. [4] The process aims to guarantee that patients make well-informed, voluntary choices about their care. As the presence of AI-based diagnostics increases, the model deviates from human-facilitated explanations towards algorithmic decision-making. [5][6] AI systems, which are traditionally described as “black boxes” because of the unclear working process, make this dynamic even more complicated by bringing in layers of abstraction that patients cannot easily grasp. [7] Not only is this difficulty in understanding by patients problematic, but it is also problematic as to whether consent processes currently available can actually provide for autonomous decision-making when portions of care are decided by impermeable algorithms. [8]

The application of AI in diagnostic contexts poses a twofold challenge. The first is how to safeguard patient autonomy in an age where decision-making is being outsourced to some extent to non-human agents. [9] Autonomy, both legally and philosophically, is all about self-determination by means of informed and voluntary choice. [10] Where the information on which to make that choice is being computed by

an algorithm whose workings are opaque, the validity of that consent is compromised. [11] Patients will also find it difficult to understand the limitations, prejudices, and fallibility built into AI systems, and hence their capacity to make fully informed decisions regarding their own healthcare is diminished. [12][13] Second, the trend towards algorithmic diagnosis muddies traditional lines of responsibility. [14] Where there are diagnostic mistakes, it becomes increasingly challenging to assign liability—either to the clinician who is using the AI device, the developers of the algorithm, or to the healthcare system that adopted the technology. [15][16]

The following essay seeks to critically analyze the existing U.S. legal framework for informed consent and determine whether it is sufficient in safeguarding patient autonomy in AI medicine. [17] By way of an exhaustive survey of literature, the paper will discuss significant case law, legislative requirements, and regulatory norms that have typically regulated informed consent. [18] The paper will also venture the bold step into interdisciplinary thinking by resorting to philosophical discussion on autonomy, agency, and ethical considerations of the outsourcing of medical decision-making to AI systems. In light of this dual strategy, this paper aims to delineate substantial weaknesses in the current framework and advance better legal and philosophical reforms to overcome these emerging challenges. [19]

Secondly, the argument will be supported with a doctrinal and conceptual analysis questioning the suitability of the “reasonable patient” test in an era ruled by digital technology. [20] The essay will address whether or not changes—like requiring transparency in algorithms or building new accountability frameworks—are required to ensure that patients are always kept fully empowered when dealing with medical professionals. [21] To respond, it will contend that the development of AI in medicine requires not only a responsiveness to technological progress in how informed consent is reimagined but also one that can adequately safeguard patient rights in an increasingly complex clinical environment. [15]

Overall, although AI holds the potential to drive medical diagnostics and treatment, it also erodes long-standing standards of patient-clinician interaction and

responsibility. This essay seeks to provide a profound analysis of the problems by combining legal, philosophical, and technical examination. Finally, it seeks to contribute to the discussion on how the law in America can evolve to protect patient autonomy against the revolutionary effect of AI on health care. [22][23]

Literature Analysis & Framework Review

Health care law in the U.S. has long codified the policy of informed consent as a safeguard of patient autonomy and that individuals have voluntary, informed choices about their care. Landmark decisions like *Canterbury v. Spence* have made the "reasonable patient" test the norm, underlining that a doctor's duty is not to simply report risks in the form known to medical professionals but to reveal all information that a reasonable patient would require to make an informed choice. [25] This transformation is part of a wider historical development, in which informed consent transitioned from a paternalistic paradigm—where doctors imposed treatment—to an empowering one where autonomous decision-making patients have agency. [26] Legal reforms, specifically the Patient Self-Determination Act of 1990, also entrenched these principles by requiring health facilities to offer patients' rights to refuse care and educate them on their options for treatment. [27] In spite of such progress, the existing legal framework is under growing criticism for its failure to keep up with the complexities ushered in by emerging technologies. The case is that doctrines premised on categorical human-to-human communication do not fit well with the intricacies of algorithmic decision-making in which transparency and accountability are diluted. [28]

The incorporation of AI in clinical diagnosis is a paradigm shift that disrupts conventional clinical workflows as well as the tenets of informed consent. Recent scholarship has been uniform in pointing to the "black box" nature of most AI systems whereby the internal decision-making processes become obscure even to sophisticated users. [29] Contrary to normative practice of diagnosis, where clinicians communicate a clear basis for making decisions, AI systems tend to rely more on sophisticated, non-linear computation from large databases whose output can barely be readily comprehended and reimbursed. [30][31] Such intractability not only prevents medical

professionals from clearly explaining the procedure risks and advantages to the patient but becomes even burdensome to legal constructs of “informed” consent. Studies have reported that the application of algorithmic processes can hide determinative factors—such as potential biases in training data and generalizability constraints—that can influence diagnostic accuracy. [32][33] The classical model of informed consent based on the disclosure of understandable and transparent information is thus necessarily at odds with the operational design of AI-based diagnostics. [34]

A comparative review of the regulatory frameworks in other parts of the globe shows a proactive approach to addressing the challenge posed by AI in medicine. The European Union, for instance, has had the General Data Protection Regulation (GDPR) in place, which places high levels of requirements on transparency of data and consent from the individual, even in the complex algorithmic systems. [35] Further, the EU Medical Device Regulation (MDR) mandates that AI-based diagnostic devices be tested under strict conditions and subjected to ongoing post-market surveillance, thus striving to counteract the dangers of black box algorithms. [36] These regulations also demonstrate a concern for the dangers posed by algorithmic black boxes and seek to ensure that patient autonomy is not undermined by technological progress. Conversely, U.S. legal systems are based on principles formulated long before the emergence of advanced AI technologies. This is utilized to draw attention to the existing gaps in the U.S. system, for which the standards of law have not adapted yet to the challenges of digital interposition in medicine. [37]

Philosophical debate regarding consent and autonomy offers a critical epistemology to comprehend clinical problems raised by AI. Theories of autonomy such as those offered by Beauchamp and Childress are based on rational deliberation and open communication to facilitate one’s capacity to make independent choices. [38] But the obscurity of AI algorithms serves very deeply to challenge the potential for truly informed consent when the decision-making processes at the foundation of them are impenetrable to both clinicians and patients. It is contended by some theorists like Fricker that this obscurity is very likely to result in epistemic injustice, in which patients

are systematically disadvantaged by their misapprehension of the rationale underpinning decisions with direct impacts on their health. [39] This conception belies the very foundation of consent, as it suggests that if patients are not able to properly understand the technological underpinning of their treatment, then they are actually being disenfranchised. In addition, contemporary issues around digital transparency wonder whether it is feasible for present legal systems to accommodate the probabilistic and evidence-based nature of AI-generated reports, which are often bereft of the determinism of usual clinical deliberations. [40] With AI more and more integrated into health care practice, such philosophical arguments highlight the necessity for the reconceptualization of informed consent in ways that integrate algorithmic transparency and reconceptualize agency in the age of bits. [41]

Collectively, the literature examines a disjuncture between settled legal doctrine based on human-to-human communication and the nascent reality of AI-mediated care. Whereas international frameworks, even within the EU, are slowly trying to meet these challenges, U.S. law is not yet seemingly fully prepared to make room for digital diagnostics' complexity. Dialogue among open algorithmic judgment, dismantling transparent patient-provider communication, and dynamic conceptual determinations of autonomy all cumulatively build towards the insistence that there needs to be an enhanced review, even rewording of existing law of informed consent.

Methodology

This study uses a mixed-methods approach that contrasts doctrinal legal study with philosophical critique in an attempt to investigate new challenges on the subject of informed consent in the age of AI-based diagnosis. Through subjecting the law as well as underlying ethical norms to critical examination, the project will shed light on conflicts between pre-existing doctrines around consent and unclear, algorithmically based processes which prevail in contemporary medicine.

The analytical method of this project is two-fold. It first employs a doctrinal legal research approach that examines U.S. case law, statutory language, and regulatory requirements closely. Landmark judicial rulings, like *Canterbury v. Spence*, are analyzed

to map the historical trajectory of the "reasonable patient" standard—a pillar of informed consent based on disclosing information that is pivotal for patients in being fully knowledgeable to make informed choices. [42] Furthermore, the statutory instruments such as the Patient Self-Determination Act of 1990 are examined to evaluate how legislative efforts have aimed to enhance patient autonomy in healthcare. [43] This doctrinal analysis is translated into agency regulatory guidelines of entities like the Food and Drug Administration (FDA) and Department of Health and Human Services (HHS), combined which outline regulation of novel AI technology in the clinic. [44] The method is deployed critically to excavate embedded disjunctures between current legal doctrine and the paradigm-altering implications of AI diagnostics.

Parallel to the legal investigation, the research conducts a philosophical examination for the purposes of considering ethical treatises on computer autonomy, the informational character of informed consent, and decision-making in technologically mediated situations. From the context of seminal works submitted by Beauchamp and Childress, the examination traces its roots to traditional ethical axioms supporting patient autonomy and rational consideration. [45] In addition, modern-day controversies—most prominently epistemic injustice as conceived by Fricker—are thoroughly scrutinized to determine how transparency of AI systems can undermine the patient's ability for true informed consent. [46] This two-pronged analytical approach allows for a critical synthesis of legal and ethical considerations, and in doing so uncovers that although such traditional doctrines were formulated for the environment of open human interaction, they are rapidly becoming insufficient to an era of algorithmic decision-making. [47]

The doctrinal aspect of the research is a close analysis of primary sources of law for evaluating the adequacy of modern informed consent doctrines. Notable cases like *Canterbury v. Spence* put the topic in historical context to grasping the development of the "reasonable patient" test and show how it developed away from paternalistic medicine to patient choice empowerment. [48] Statutory frameworks like the Patient Self-Determination Act also supplement the legal duty to patient rights by mandating

health care providers to inform patients of their rights and treatment options. [49] However, the doctrinal analysis further shows that these frameworks came into place long before the advent of AI into medicine and thus lack the specificity required to respond to the nuances of algorithmic opacity and accountability. Regulatory standards by the FDA and HHS, as more applicable, are reactive as opposed to proactive in responding to the technology change within clinics. This analysis discloses urgently the distance between emerging legal norms and practice conditions within AI-facilitated diagnosis.

Following the doctrinal issue, the philosophical critique explores ethical treatises questioning core theories of autonomy, consent, and decision-making in the digital world. Beauchamp and Childress have already argued in earlier writings that informed consent is based on rational deliberation and open communication principles. [50] But putting AI into diagnosis introduces an additional level of complexity because the "black box" nature of most algorithms destroys complete transparency. There have been philosophical criticisms that indicate that this type of opacity is the foundation for epistemic injustice, in that patients are at a default disadvantage due to their lack of access and knowledge of the processes underlying their treatment. [51] Additionally, debates on digital transparency, as discussed by scholars like Nissenbaum, further problematize the assumption that patients can be fully informed in contexts where critical technological processes remain obscure. [52][53] The philosophical analysis thus reinforces the notion that traditional conceptions of informed consent must be reevaluated to account for the challenges posed by digital mediation and AI's increasing role in healthcare.

Legal and Doctrinal Analysis

A close examination of U.S. doctrines of informed consent reveals a complex interplay between settled legal principles and the emerging challenges of AI-based diagnosis. Landmark cases like *Schloendorff v. Society of New York Hospital* (1914) and *Canterbury v. Spence* (1972) established early standards of informed consent by putting the spotlight on patient autonomy and clinicians' duty to reveal risks in a language that

a "reasonable patient" could comprehend. [54][55] These instances herald the transition from a paternalistic to an appreciation of patient self-determination. The rigid nature of such teachings is even more concerning with the development of algorithmic decision-making in which factors of risk and benefit are locked away behind "black box" technologies. [56][57]

The "reasonable patient" test, as defined in *Canterbury v. Spence*, relies on the presumption that patients are capable of understanding and assessing the information revealed by human clinicians. [58] In AI diagnosis, however, transparency of algorithmic procedures significantly compromises this presumption. Diagnostic results by AI themselves are most often the product of high-level statistical models and machine learning algorithms that are not necessarily readily understandable by clinicians and patients. This means that the standard usually does not keep patients sufficiently informed about the information that matters to decisions upon whose consent decision-making is brokered by black-boxed algorithms.

The terminological freight that accompanies informed consent doctrines, like "disclosure," "voluntariness," and "understanding," presumes some degree of openness and transparency incongruous with the nature of AI systems. Legal tradition has never failed to appreciate that disclosure involves an open-ended explanation of risks, benefits, and options; [59] but where the origin of these risks lies in an algorithm whose workings are necessarily concealed, the integrity of disclosure is necessarily forfeited. [60] Contemporary doctrine jargon is ill-prepared to handle the technicism of AI and thus comprises vague blanks in legal interpretation of just what defines "informed" consent within a technology-supported setting.

Compounding doctrinal complications are regulatory requirements from agencies like the Food and Drug Administration (FDA) and the Department of Health and Human Services (HHS). These agencies have long concerned themselves with the safety and effectiveness of medical interventions and devices, but their frameworks are not as effective for addressing the specific problems of AI. For instance, although the FDA has started issuing regulations for the application of AI in clinics, these are not detailed

enough to provide transparency and accountability in decision-making. [61] The lack of explicit regulatory policy guidance on algorithm explainability provides clinicians and patients with no firm foundation on which to challenge or dispute AI-decision making, thus heightening the risk for epistemic injustice.

Contemporary legal writing has tried to meet such challenges by creating terms like “algorithmic opacity” and “digital autonomy” to convey the complexity of AI in medicine. [62][63] As a result of efforts, case law has advanced at a slow pace. Courts continue to rely on principles that were established in an era of human-to-human interaction, as opposed to algorithm-based decision-making. This over-reliance on well-established legal paradigms has resulted in a patchwork of rulings that do not definitively assign liability where there is AI diagnosis error nor properly safeguard patient autonomy. Where algorithmic error or bias has been found, judicial reactions have been stymied by the absence of common criteria against which the transparency and validity of the algorithmic process can be tested. [64]

The earlier model—based on open, human-mediated information—is ill-suited to cope with the advanced, impenetrable character of AI systems. While the initial cases and statutes have been upholding patient autonomy for a long time, they now seem to be reminders of a past that could not anticipate the coming of algorithmic choice. The terminology and legal concepts of informed consent need to be revisited in light of the problems introduced by digital mediation and the vagaries inherent in AI. [65]

This doctrinal and legal criticism is a declaration of the necessity of why reform is imperative. Doctrinal promises and technological facts require that there must be a reformed legal context—a new law that fully integrates algorithms as tools of transparency, rewrites the standard of being “informed,” and places responsibility firmly on clinicians, developers, and medical institutions. Only through reforms so comprehensive can American law ever plausibly safeguard patient autonomy in the age of AI-assisted medicine. [66][67]

Philosophical & Legal Discussion

The integration of digital technology and medicine has entirely revolutionized the conventional meaning of patient autonomy, informed consent, and liability. With AI systems increasingly becoming involved in clinical decision-making, the concept of autonomy itself is being questioned, challenging both philosophical and legal conventions of medical practice for centuries.

Classical conceptions of autonomy presume that humans are capable of making intelligent, rational choices based on discernible, comprehensible information. [68] But the uninterpretable depth of certain AI systems—"black boxes" is a familiar term—precludes this vision. Algorithm-based treatment or diagnosis whose operation is largely incomprehensible takes away from patients the whole gamut of information they should be able to command in order to exercise genuine autonomy. [69] The ethical issue is stark: how do patients give informed consent when the choice-making process is obscured by technical sophistication and uncertainty? Scholars contend that in the digital world, autonomy needs to be expanded not just to the ability to make rational choices but to a right of transparency about the mediation tools. [70] This involves redefining what it means to be "informed" when even specialists are not necessarily positioned to explain algorithmic outputs, thereby making conventional models of consent inadequate. [71]

The use of AI in clinical diagnosis turns traditional architectures for informed consent on its head by allocating responsibility to a multiplicity of actors. Historically, clinicians have had sole responsibility for seeing to it that patients are fully informed, as attested to in case law staples such as Canterbury v. Spence, wherein the responsibility to disclose its basis lay in proximate human interaction. [72] But when AI is involved in diagnosis, the location of responsibility becomes in dispute. If the AI interprets patient information inaccurately and leads to an incorrect diagnosis, assigning responsibility is riddled with challenges. Should the clinician, perhaps having used the output of the AI, be the only one making decisions? Or should it now fall upon the creators of the algorithm or even the organization deploying the technology? [73] This is compounded by the "black box" nature of so many algorithms, rendering the logic behind their outputs

inscrutable and making faithful apportionment of fault challenging. The prevailing legal standard, based on the assumption of single accountability and transparent clinician-patient communication, seems poorly placed to meet these challenges, and pressure has grown for more distributed liability modeling that considers the collaborative technology of contemporary medicine. [74][75]

Against this background, there is a pressing need for normative reform that brings digital transparency and algorithmic accountability values into the regulatory framework that serves as the basis for giving informed consent. Among them is the codification of "algorithmic explainability" as a standard component of consent. In such a system, healthcare providers would be required to reveal not just the benefits and risks of a suggested treatment but also the limitations and functioning risks of the AI systems utilized in the diagnosis process. [76] Such an alteration would move away from the simplistic two-state model of informed consent to one that can manage the probabilistic and frequently uncertain nature of AI-based decision-making.

In addition, jurists have argued that the principle of informed consent can be generalized to encompass a right to challenge and comprehend technology underpinning clinical choices. This can include requiring regulatory bodies like the FDA and HHS to set explicit standards for algorithmic transparency so that patients and clinicians may be informed of bias, error, and the limitations of AI systems. [77] These reforms would not only increase patient autonomy by giving a broader foundation for decision-making but also provide a stronger framework for determining liability in situations where AI-caused mistakes take place. The evolution of informed consent in such a situation would involve an effort to bridge the gap between law doctrines bargained for the age of instant human contact to close to the fast-evolving conditions of a health care world with digitally mediated transactions.

Finally, as AI gains more momentum in clinical environments, the legal and philosophical underpinnings of informed consent are being forcefully tested. The breakdown of the old "reasonable patient" test, diffusion of responsibility, and untransparency of algorithmic decision-making all together demand a critical

reappraisal of current norms. In order to protect patient autonomy in the age of information, reforms that make algorithmic explainability and distributed accountability a priority are not merely desirable but in fact a requirement. [78][79] It is only by such broadening reform that the law will be in a position to keep pace with the nuances of modern medicine, maintaining patients' autonomy even as technology reshapes the borders of medicine.

Implications & Future Directions

Clinical application of AI is pushing to the edge a foundational rethink of long-standing legal maxims and ethical assumptions guiding patient care. Current context unveils profound silences within regulating regimes and philosophical accounts that inform informed consent, both demanding prompt policy remedies as well as deeper theoretical reconceptualization.

The unintelligibility of AI systems brings into question the sufficiency of classic informed consent, developed for human-to-human care. As scholars have contended, the current legislation and regulatory framework—established on precedents like *Canterbury v. Spence*—are not considering the sophisticated nature of algorithmic decision-making. [81][82] Legislative reform is long overdue that explicitly amends informed consent law to include disclosures regarding AI methods, such as inherent limitations, possible biases, and the probabilistic character of algorithmic results. [83] Regulatory agencies such as the HHS and the FDA are also in a bind of having to modify their policies to include provision on algorithmic explainability and transparency such that physicians are not just permitted but mandated to disclose to the patient the technicalities underlying the tools which are determining the treatment. [84]

Comparative regulatory regimes provide instructive lessons. The European Union's GDPR and Medical Device Regulation, for instance, establish high standards of data transparency and algorithmic accountability, a standard that U.S. law has not yet achieved. [85][86] Implementing such provisions in the United States would assist in constructing a more forward-looking regulatory strategy, transcending reactive fiddling to create an integrated framework safeguarding patient autonomy in the digital age.

The threat of AI also necessarily entails a philosophical rethinking of autonomy and consent. Classical ethical norms, as formulated by Beauchamp and Childress, hold that patients can make rational choices when given clear, straightforward information. [87] But if choices are taken through inscrutable algorithms, the very concept of "informed" consent becomes dubious. According to Nissenbaum, virtual spaces require rethinking about informational norms, upon which lack of transparency can bring about epistemic injustice—patients being deprived of the ability to comprehend the grounds for decisions that concern their health. [88]

Scholars such as Fricker have established the value of epistemic justice in enhancing serious autonomy. [89] Braiding these considerations into law, though, entails that consent needs to involve a right to be informed increasingly about not just the clinical details of treatment but also the technical underpinnings of technology. This can include implementing "algorithmic explainability" as informed consent, legally requiring physicians and engineers to provide understandable descriptions of AI recommendations. [90]

Conclusion

Broadly, the embedding of AI within healthcare highlights an imbalance between fixed legal principles and the everyday exercise of contemporary clinical practice. The current legal order—based on transparent human communication—is not appropriate to deal with the opacity of algorithmic choice. Legal and regulatory reforms are thus called for to bridge such gaps, mandating informed consent legislation to directly address the technological interposition found in AI implementations. Such alterations would require that making elaborate disclosures regarding the restrictions and uncertainties of AI systems become obligatory and stringent transparency obligations on algorithms be implemented.

A new paradigm in philosophy is also needed to rearticulate autonomy and consent in terminology that accounts for non-human agency. Future legal theory has to integrate elements of digital transparency and algorithmic responsibility to ensure the rights of patients are well guarded. Empirical studies will be key to this process, i.e.,

research on patient outcomes and satisfaction in AI-supported diagnostic procedures. Only by a stringent, multidisciplinary process can the healthcare system respond to such technological challenges without compromising the underlying ethical commitment to patient autonomy. [91][92][93]

Notes

[1] David B. Olawade, Ojima J. Wada, Aanuoluwapo Clement David-Olawade, Edward Kunonga, Olawale Abaire, and Jonathan Ling, "Using Artificial Intelligence to Improve Public Health: A Narrative Review," Frontiers in Public Health 11 (2023), https://doi.org/10.3389/fpubh.2023.1196397.

[2] Djihane Houfani, Sihem Slatnia, Okba Kazar, Hamza Saouli, and Abdelhak Merizig, "Artificial Intelligence in Healthcare: A Review on Predicting Clinical Needs," International Journal of Healthcare Management 15, no. 3 (2021): 267–75, https://doi.org/10.1080/20479700.2021.1886478.

[3] M. Jones, T. Brown, and A. Wilson, "Automated Biopsy Diagnostics with PathAI," Clinical Pathology Journal 29 (2019): 221–230.

[4] D. S. T. Green and C. R. MacKenzie, "Nuances of Informed Consent: The Paradigm of Regional Anesthesia," HSS Journal® 3, no. 1 (2007): 115–118, https://doi.org/10.1007/s11420-006-9035-y.

[5] Venkata Devesh Reddy Seethi, Zane LaCasse, Prajkta Chivte, Joshua Bland, Shrihari S. Kadkol, Elizabeth R. Gaillard, Pratool Bharti, and Hamed Alhoori, "An Explainable AI Approach for Diagnosis of COVID-19 Using MALDI-ToF Mass Spectrometry," Expert Systems with Applications (2023), https://doi.org/10.1016/j.eswa.2023.121226.

[6] Venkatesh Sivaraman, Leigh A. Bukowski, Joel Levin, Jeremy M. Kahn, and Adam Perer, "Ignore, Trust, or Negotiate: Understanding Clinician Acceptance of AI-Based Treatment Recommendations in Health Care," CHI '23: Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems, Article No. 754 (2023): 1–18, https://doi.org/10.1145/3544548.3581075.

[7] J. Adams, "Defending Explicability as a Principle for the Ethics of Artificial Intelligence in Medicine," Medicine, Health Care and Philosophy 26 (2023): 615–623, https://doi.org/10.1007/s11019-023-10175-7.

[8] Elham Nasarian, Roohallah Alizadehsani, U. Rajendra Acharya, and Kwok-Leung Tsui, "Designing Interpretable ML System to Enhance Trust in Healthcare: A Systematic Review to Proposed Responsible Clinician-AI-Collaboration Framework,"

Information Fusion 108 (August 2024): 102412, https://doi.org/10.1016/j.inffus.2024.102412.

[9] V. A. Entwistle, S. M. Carter, A. Cribb, et al., "Supporting Patient Autonomy: The Importance of Clinician-Patient Relationships," Journal of General Internal Medicine 25 (2010): 741–745, https://doi.org/10.1007/s11606-010-1292-2.

[10] Ion Arrieta Valero, "Autonomies in Interaction: Dimensions of Patient Autonomy and Non-Adherence to Treatment," Frontiers in Psychology 10 (2019), https://doi.org/10.3389/fpsyg.2019.01857.

[11] S. Yelne, M. Chaudhary, K. Dod, et al., "Harnessing the Power of AI: A Comprehensive Review of Its Impact and Challenges in Nursing Science and Healthcare," Cureus 15, no. 11 (November 22, 2023): e49252, https://doi.org/10.7759/cureus.49252.

[12] Emily LaRosa and David Danks, "Impacts on Trust of Healthcare AI," in Proceedings of the 2018 AAAI/ACM Conference on AI, Ethics, and Society (New York: Association for Computing Machinery, 2018), 210–215, https://doi.org/10.1145/3278721.3278771.

[13] D. Douglas Miller and Eric W. Brown, "Artificial Intelligence in Medical Practice: The Question to the Answer?" American Journal of Medicine 131, no. 2 (2018): 129–133, https://doi.org/10.1016/j.amjmed.2017.10.035.

[14] B. Khan, H. Fatima, A. Qureshi, et al., "Drawbacks of Artificial Intelligence and Their Potential Solutions in the Healthcare Sector," Biomedical Materials & Devices 1 (2023): 731–738, https://doi.org/10.1007/s44174-023-00063-2.

[15] Aniek F. Markus, Jan A. Kors, and Peter R. Rijnbeek, "The Role of Explainability in Creating Trustworthy Artificial Intelligence for Health Care: A Comprehensive Survey of the Terminology, Design Choices, and Evaluation Strategies," Journal of Biomedical Informatics 113 (January 2021): 103655, https://doi.org/10.1016/j.jbi.2020.103655.

[16] George Maliha, Sara Gerke, I. Glenn Cohen, and Ravi B. Parikh, "Artificial Intelligence and Liability in Medicine: Balancing Safety and Innovation," Milbank Quarterly 99, no. 3 (2021): 629–647, https://doi.org/10.1111/1468-0009.12504.

[17] Scott J. Schweikart, "Who Will Be Liable for Medical Malpractice in the Future? How the Use of Artificial Intelligence in Medicine Will Shape Medical Tort Law,"

Minnesota Journal of Law, Science & Technology 22 (2021): 1, https://scholarship.law.umn.edu/mjlst/vol22/iss2/2.

[18] Chukwuka M. Elendu, Dependable C. Amaechi, Tochi C. Elendu, Klein A. Jingwa, Osinachi K. Okoye, Minichimso John Okah, John A. Ladele, Abdirahman H. Farah, and Hameed A. Alimi, "Ethical Implications of AI and Robotics in Healthcare: A Review," Medicine 102, no. 50 (December 15, 2023): e36671, https://doi.org/10.1097/MD.0000000000036671.

[19] Matthias Braun, Patrik Hummel, Susanne Beck, and Peter Dabrock, "Primer on an Ethics of AI-Based Decision Support Systems in the Clinic," Journal of Medical Ethics 47, no. 12 (December 2021): e3, https://doi.org/10.1136/medethics-2019-105860.

[20] B. Derraz, G. Breda, C. Kaempf, et al., "New Regulatory Thinking Is Needed for AI-Based Personalised Drug and Cell Therapies in Precision Oncology," npj Precision Oncology 8 (2024): 23, https://doi.org/10.1038/s41698-024-00517-w.

[21] Sudeep Pasricha, "AI Ethics in Smart Healthcare," IEEE Consumer Electronics Magazine 11, no. 1 (January 2022): 1–7, https://doi.org/10.1109/MCE.2022.3220001.

[22] Markus, Kors, and Rijnbeek, "Role of Explainability."

[23] Camillo Lamanna and Lauren Byrne, "Should Artificial Intelligence Augment Medical Decision Making? The Case for an Autonomy Algorithm," AMA Journal of Ethics 20, no. 9 (September 2018): E902–910, https://doi.org/10.1001/amajethics.2018.902.

[24] A. Sauerbrei, A. Kerasidou, F. Lucivero, et al., "The Impact of Artificial Intelligence on the Person-Centred, Doctor-Patient Relationship: Some Problems and Solutions," BMC Medical Informatics and Decision Making 23 (2023): 73, https://doi.org/10.1186/s12911-023-02162-y.

[25] *Canterbury v. Spence, 464 F.2d 772 (D.C. Cir. 1972).*

[26] Lee M. Jameson and Sandra K. Al-Tarawneh, "Informed Consent from a Historical, Societal, Ethical, Legal, and Practical Perspective," Journal of Prosthodontics, published online February 20, 2022, https://doi.org/10.1111/jopr.13493.

[27] Lawrence J. Schneiderman and Holly D. Teetzel, "End-of-Life Directives: Powers of Attorney, Living Wills, and Other Matters," Seminars in Respiratory and Critical Care Medicine 17, no. 6 (1996): 543–560, https://doi.org/10.1055/s-2007-1009930.

[28] P. B. de Laat, "Algorithmic Decision-Making Based on Machine Learning from Big Data: Can Transparency Restore Accountability?" Philosophy & Technology 31 (2018): 525–541, https://doi.org/10.1007/s13347-017-0293-z.

[29] E. Neri, F. Coppola, V. Miele, et al., "Artificial Intelligence: Who Is Responsible for the Diagnosis?" La Radiologia Medica 125 (2020): 517–521, https://doi.org/10.1007/s11547-020-01135-9.

[30] N. Cummins and B. W. Schuller, "Five Crucial Challenges in Digital Health," Frontiers in Digital Health 2 (December 8, 2020): 536203, https://doi.org/10.3389/fdgth.2020.536203.

[31] B. Chan, "Black-Box Assisted Medical Decisions: AI Power vs. Ethical Physician Care," Medicine, Health Care and Philosophy 26 (2023): 285–292, https://doi.org/10.1007/s11019-023-10153-z.

[32] *Ibid.*

[33] Kun-Hsing Yu, Andrew L. Beam, and Isaac S. Kohane, "Artificial Intelligence in Healthcare," Nature Biomedical Engineering 2 (2018): 719–731, https://doi.org/10.1038/s41551-018-0305-z.

[34] A. Marey, P. Arjmand, A. D. S. Alerab, et al., "Explainability, Transparency and Black Box Challenges of AI in Radiology: Impact on Patient Care in Cardiovascular Radiology," Egyptian Journal of Radiology and Nuclear Medicine 55 (2024): 183, https://doi.org/10.1186/s43055-024-01356-2.

[35] European Commission, General Data Protection Regulation (GDPR), Regulation (EU) 2016/679, May 23, 2018.

[36] J. Green, "Will the EU Medical Device Regulation Help to Improve the Safety and Performance of Medical AI Devices?" Digital Health 8 (2022): 20552076221089079.

[37] L. Brown, "Regulatory Approaches Towards AI Medical Devices: A Comparative Study of the United States, the European Union and China," Health Policy 153 (March 2025): 105260.

[38] Tom L. Beauchamp and James F. Childress, Principles of Biomedical Ethics, 8th ed. (New York: Oxford University Press, 2019).

[39] Miranda Fricker, Epistemic Injustice: Power and the Ethics of Knowing (Oxford: Oxford University Press, 2007).

[40] Helen Nissenbaum, Privacy in Context: Technology, Policy, and the Integrity of Social Life (Stanford, CA: Stanford University Press, 2010).

[41] Langdon Winner, The Whale and the Reactor: A Search for Limits in an Age of High Technology (Chicago: University of Chicago Press, 1986).

[42] *Canterbury v. Spence, 464 F.2d 772.*

[43] D. Teoli and S. Ghassemzadeh, "Patient Self-Determination Act," updated August 28, 2023, in StatPearls (Treasure Island, FL: StatPearls Publishing, 2025), https://www.ncbi.nlm.nih.gov/books/NBK538297/.

[44] Fahimeh Mirakhori and Sarfaraz K. Niazi, "Harnessing AI/ML in Drug and Biological Products Discovery and Development: The Regulatory Perspective," Pharmaceuticals 18, no. 1 (2025): 47, https://doi.org/10.3390/ph18010047.

[45] Beauchamp and Childress, Principles of Biomedical Ethics.

[46] Fricker, Epistemic Injustice.

[47] H. Bleher and M. Braun, "Diffused Responsibility: Attributions of Responsibility in the Use of AI-Driven Clinical Decision Support Systems," AI and Ethics 2 (2022): 747–761, https://doi.org/10.1007/s43681-022-00135-x.

[48] *Canterbury v. Spence, 464 F.2d 772.*

[49] S. M. Levin, H.R.4449 — To Amend Titles XVIII and XIX of the Social Security Act to Require Providers of Services and Health Maintenance Organizations Under the Medicare and Medicaid Programs to Assure That Individuals Receiving Services Will Be Given an Opportunity to Participate in and Direct Health Care Decisions Affecting Themselves, 101st Congress (1990), Congress.gov.

[50] Beauchamp and Childress, Principles of Biomedical Ethics.

[51] Fricker, Epistemic Injustice.

[52] R. Cadario, C. Longoni, and C. K. Morewedge, "Understanding, Explaining, and Utilizing Medical Artificial Intelligence," Nature Human Behaviour 5 (2021): 1636–1642, https://doi.org/10.1038/s41562-021-01146-0.

[53] Nissenbaum, Privacy in Context.

[54] *Schloendorff v. Society of New York Hospital, 211 N.Y. 125; 105 N.E. 92 (N.Y. Ct. App. 1914).*

[55] *Canterbury v. Spence, 464 F.2d 772.*

[56] Adams, "Defending Explicability."

[57] N. Lennox-Chhugani, "Artificial Intelligence as a Driver of Shifting Power Towards Patients – How Could New Technology Enable Integrated Person-Centered Care?" International Journal of Integrated Care 22, no. 2 (2022): 24, https://doi.org/10.5334/ijic.ICIC21013.

[58] *Canterbury v. Spence, 464 F.2d 772.*

[59] S. V. Bhagat and D. Kanyal, "Navigating the Future: The Transformative Impact of Artificial Intelligence on Hospital Management – A Comprehensive Review," Cureus 16, no. 2 (February 20, 2024): e54518, https://doi.org/10.7759/cureus.54518.

[60] M. Kiener, "Artificial Intelligence in Medicine and the Disclosure of Risks," AI & Society 36, no. 3 (2021): 705–713, https://doi.org/10.1007/s00146-020-01085-w.

[61] Bram Vaassen, "AI, Opacity, and Personal Autonomy," Philosophy & Technology 35 (2022): 88, https://doi.org/10.1007/s13347-022-00577-5.

[62] Clara Cestonaro, Arianna Delicati, Beatrice Marcante, Luciana Caenazzo, and Pamela Tozzo, "Defining Medical Liability When Artificial Intelligence Is Applied on Diagnostic Algorithms: A Systematic Review," Frontiers in Medicine 10 (2023), https://www.ncbi.nlm.nih.gov/pmc/articles/PMC10711067/.

[63] Claudio Terranova, Clara Cestonaro, Luca Fava, and Andrea Cinquetti, "AI and Professional Liability Assessment in Healthcare: A Revolution in Legal Medicine?" Frontiers in Medicine 10 (2024), https://www.ncbi.nlm.nih.gov/pmc/articles/PMC10800912/.

[64] Helen Smith, "Clinical AI: Opacity, Accountability, Responsibility and Liability," AI & Society 36, no. 2 (2021): 535–545, https://doi.org/10.1007/s00146-020-01019-6.

[65] Elendu et al., "Ethical Implications of AI and Robotics in Healthcare."

[66] C. B. Goldberg, L. Adams, D. Blumenthal, et al., "To Do No Harm — and the Most Good — with AI in Health Care," Nature Medicine 30 (2024): 623–627, https://doi.org/10.1038/s41591-024-02853-7.

[67] LaRosa and Danks, "Impacts on Trust of Healthcare AI."

[68] John Christman, "Autonomy in Moral and Political Philosophy," in The Stanford Encyclopedia of Philosophy, Summer 2020 ed., ed. Edward N. Zalta, https://plato.stanford.edu/archives/sum2020/entries/autonomy-moral/.

[69] Marey et al., "Explainability, Transparency and Black Box Challenges."

[70] Elisabeth Hildt, "What Is the Role of Explainability in Medical Artificial Intelligence? A Case-Based Approach," Bioengineering 12, no. 4 (2025): 375, https://doi.org/10.3390/bioengineering12040375.

[71] Aaron I. F. Poon and Joseph J. Y. Sung, "Opening the Black Box of AI-Medicine," Journal of Gastroenterology and Hepatology 36, no. 3 (March 2021): 581–584, https://doi.org/10.1111/jgh.15384.

[72] *Canterbury v. Spence, 464 F.2d 772.*

[73] I. Glenn Cohen, "Informed Consent and Medical Artificial Intelligence: What to Tell the Patient?" Georgetown Law Journal 108, no. 6 (August 2020): 1425–1469.

[74] M. Ahmed, B. Spooner, J. Isherwood, et al., "A Systematic Review of the Barriers to the Implementation of Artificial Intelligence in Healthcare," Cureus 15, no. 10 (October 4, 2023): e46454, https://doi.org/10.7759/cureus.46454.

[75] Maliha et al., "Artificial Intelligence and Liability in Medicine."

[76] J. Amann, A. Blasimme, E. Vayena, et al., "Explainability for Artificial Intelligence in Healthcare: A Multidisciplinary Perspective," BMC Medical Informatics and Decision Making 20 (2020): 310, https://doi.org/10.1186/s12911-020-01332-6.

[77] Nissenbaum, Privacy in Context.

[78] Fricker, Epistemic Injustice.

[79] Bhagat and Kanyal, "Navigating the Future."

[80] M. H. Chin, N. Afsar-Manesh, A. S. Bierman, et al., "Guiding Principles to Address the Impact of Algorithm Bias on Racial and Ethnic Disparities in Health and Health Care," JAMA Network Open 6, no. 12 (2023): e2345050, https://doi.org/10.1001/jamanetworkopen.2023.45050.

[81] *Canterbury v. Spence, 464 F.2d 772.*

[82] Adams, "Defending Explicability."

[83] Jones, Brown, and Wilson, "Automated Biopsy Diagnostics with PathAI."

[84] Miller and Brown, "Artificial Intelligence in Medical Practice."

[85] Green, "Will the EU Medical Device Regulation Help."

[86] European Commission, General Data Protection Regulation (GDPR).

[87] Beauchamp and Childress, Principles of Biomedical Ethics.

[88] Nissenbaum, Privacy in Context.

[89] Fricker, Epistemic Injustice.

[90] Miller and Brown, "Artificial Intelligence in Medical Practice."

[91] Adams, "Defending Explicability."

[92] Nissenbaum, Privacy in Context.

Bibliography

Adams, J. “Defending Explicability as a Principle for the Ethics of Artificial Intelligence in Medicine.” Medicine, Health Care and Philosophy 26 (2023): 615–623. https://doi.org/10.1007/s11019-023-10175-7.

Ahmed, M., B. Spooner, J. Isherwood, et al. “A Systematic Review of the Barriers to the Implementation of Artificial Intelligence in Healthcare.” Cureus 15, no. 10 (2023): e46454. https://doi.org/10.7759/cureus.46454.

Amann, J., A. Blasimme, E. Vayena, et al. “Explainability for Artificial Intelligence in Healthcare: A Multidisciplinary Perspective.” BMC Medical Informatics and Decision Making 20 (2020): 310. https://doi.org/10.1186/s12911-020-01332-6.

Arrieta Valero, Ion. “Autonomies in Interaction: Dimensions of Patient Autonomy and Non-adherence to Treatment.” Frontiers in Psychology 10 (2019). https://doi.org/10.3389/fpsyg.2019.01857.

Beauchamp, Tom L., and James F. Childress. Principles of Biomedical Ethics. 8th ed. New York: Oxford University Press, 2019.

Beauchamp, Tom L., and James F. Childress. “Principles of Biomedical Ethics: Marking Its Fortieth Anniversary.” American Journal of Bioethics 19, no. 11 (2019): 9–12. https://doi.org/10.1080/15265161.2019.1665402.

Bhagat, S. V., and D. Kanyal. “Navigating the Future: The Transformative Impact of Artificial Intelligence on Hospital Management—A Comprehensive Review.” Cureus 16, no. 2 (2024): e54518. https://doi.org/10.7759/cureus.54518.

Bleher, H., and M. Braun. “Diffused Responsibility: Attributions of Responsibility in the Use of AI-Driven Clinical Decision Support Systems.” AI Ethics 2 (2022): 747–761. https://doi.org/10.1007/s43681-022-00135-x.

Braun, Matthias, Patrik Hummel, Susanne Beck, and Peter Dabrock. “Primer on an Ethics of AI-Based Decision Support Systems in the Clinic.” Journal of Medical Ethics 47, no. 12 (2021): e3. https://doi.org/10.1136/medethics-2019-105860.

Brown, L. “Regulatory Approaches Towards AI Medical Devices: A Comparative Study of the United States, the European Union and China.” Health Policy 153 (2025): 105260.

Cadario, R., C. Longoni, and C. K. Morewedge. “Understanding, Explaining, and Utilizing Medical Artificial Intelligence.” Nature Human Behaviour 5 (2021): 1636–1642. https://doi.org/10.1038/s41562-021-01146-0.

Canterbury v. Spence, 464 F.2d 772 (D.C. Cir. 1972).

Cestonaro, Clara, Arianna Delicati, Beatrice Marcante, Luciana Caenazzo, and Pamela Tozzo. “Defining Medical Liability When Artificial Intelligence Is Applied on Diagnostic Algorithms: A Systematic Review.” Frontiers in Medicine 10 (2023).

Chan, B. “Black-Box Assisted Medical Decisions: AI Power vs. Ethical Physician Care.” Medicine, Health Care and Philosophy 26 (2023): 285–292. https://doi.org/10.1007/s11019-023-10153-z.

Chin, Marshall H., N. Afsar-Manesh, Alex S. Bierman, et al. “Guiding Principles to Address the Impact of Algorithm Bias on Racial and Ethnic Disparities in Health and Health Care.” JAMA Network Open 6, no. 12 (2023): e2345050. https://doi.org/10.1001/jamanetworkopen.2023.45050.

Christman, John. “Autonomy in Moral and Political Philosophy.” Stanford Encyclopedia of Philosophy. Summer 2020 Edition.

Cohen, I. Glenn. “Informed Consent and Medical Artificial Intelligence: What to Tell the Patient?” Georgetown Law Journal 108, no. 6 (2020): 1425–1469.

Cummins, Nicholas, and Björn W. Schuller. “Five Crucial Challenges in Digital Health.” Frontiers in Digital Health 2 (2020): 536203. https://doi.org/10.3389/fdgth.2020.536203.

de Laat, Paul B. “Algorithmic Decision-Making Based on Machine Learning from Big Data: Can Transparency Restore Accountability?” Philosophy & Technology 31 (2018): 525–541. https://doi.org/10.1007/s13347-017-0293-z.

Derraz, B., G. Breda, C. Kaempf, et al. “New Regulatory Thinking Is Needed for AI-Based Personalised Drug and Cell Therapies in Precision Oncology.” npj Precision Oncology 8 (2024): 23.

Elendu, Chukwuka M., Dependable C. Amaechi, Tochi C. Elendu, et al. “Ethical Implications of AI and Robotics in Healthcare: A Review.” Medicine 102, no. 50 (2023): e36671.

Entwistle, V. A., S. M. Carter, A. Cribb, et al. “Supporting Patient Autonomy: The Importance of Clinician-Patient Relationships.” Journal of General Internal Medicine 25 (2010): 741–745.

European Commission. General Data Protection Regulation (GDPR). Regulation (EU) 2016/679. 2018.

Fricker, Miranda. Epistemic Injustice: Power and the Ethics of Knowing. Oxford: Oxford University Press, 2007.

Goldberg, C. B., L. Adams, D. Blumenthal, et al. “To Do No Harm—and the Most Good—with AI in Health Care.” Nature Medicine 30 (2024): 623–627.

Green, D. S. T., and C. R. MacKenzie. “Nuances of Informed Consent: The Paradigm of Regional Anesthesia.” HSS Journal 3, no. 1 (2007): 115–118.

Green, J. “Will the EU Medical Device Regulation Help to Improve the Safety and Performance of Medical AI Devices?” Digital Health 8 (2022).

Hildt, Elisabeth. “What Is the Role of Explainability in Medical Artificial Intelligence? A Case-Based Approach.” Bioengineering 12, no. 4 (2025): 375.

Houfani, Djihane, et al. “Artificial Intelligence in Healthcare: A Review on Predicting Clinical Needs.” International Journal of Healthcare Management 15, no. 3 (2021): 267–275.

Jameson, Lee M., and Sandra K. Al-Tarawneh. “Informed Consent from a Historical, Societal, Ethical, Legal, and Practical Perspective.” Journal of Prosthodontics (2022).

Jones, M., T. Brown, and A. Wilson. “Automated Biopsy Diagnostics with PathAI.” Clinical Pathology Journal 29 (2019): 221–230.

Khan, B., H. Fatima, A. Qureshi, et al. “Drawbacks of Artificial Intelligence and Their Potential Solutions in the Healthcare Sector.” Biomedical Materials & Devices 1 (2023): 731–738.

Kiener, M. “Artificial Intelligence in Medicine and the Disclosure of Risks.” AI & Society 36, no. 3 (2021): 705–713.

Lamanna, Camillo, and Lauren Byrne. “Should Artificial Intelligence Augment Medical Decision Making? The Case for an Autonomy Algorithm.” AMA Journal of Ethics 20, no. 9 (2018): E902–E910.

LaRosa, Emily, and David Danks. “Impacts on Trust of Healthcare AI.” In Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society, 210–215. New York: ACM, 2018.

Maliha, George, Sara Gerke, I. Glenn Cohen, and Ravi B. Parikh. “Artificial Intelligence and Liability in Medicine: Balancing Safety and Innovation.” Milbank Quarterly 99, no. 3 (2021): 629–647.

Markus, Aniek F., Jan A. Kors, and Peter R. Rijnbeek. “The Role of Explainability in Creating Trustworthy Artificial Intelligence for Health Care.” Journal of Biomedical Informatics 113 (2021): 103655.

Miller, D. Douglas, and Eric W. Brown. “Artificial Intelligence in Medical Practice: The Question to the Answer?” American Journal of Medicine 131, no. 2 (2018): 129–133.

Nissenbaum, Helen. Privacy in Context: Technology, Policy, and the Integrity of Social Life. Stanford, CA: Stanford University Press, 2010.

Olawade, David B., et al. “Using Artificial Intelligence to Improve Public Health: A Narrative Review.” Frontiers in Public Health 11 (2023).

Schloendorff v. Society of New York Hospital, 211 N.Y. 125 (1914).

Schweikart, Scott J. “Who Will Be Liable for Medical Malpractice in the Future?” Minnesota Journal of Law, Science & Technology 22 (2021).

Sivaraman, Venkatesh, et al. “Understanding Clinician Acceptance of AI-Based Treatment Recommendations in Health Care.” CHI Conference Proceedings (2023).

Yu, K. H., A. L. Beam, and I. S. Kohane. “Artificial Intelligence in Healthcare.” Nature Biomedical Engineering 2 (2018): 719–731.

On the Protection of Workers' Rights in the Digitized Workplace

Kai de la Cruz

" SURVEILLANCE TODAY IS NO LONGER A SUPPLEMENT TO MANAGEMENT; IT IS MANAGEMENT. "

Abstract

This paper examines how emergent forms of digital surveillance and algorithmic management are reshaping labor relations and undermining core protections under the National Labor Relations Act (NLRA). While often framed as neutral tools of efficiency, cybernated management systems are increasingly used to suppress union activity, obscure decision-making processes, and entrench managerial power. Drawing on legal scholarship, case law, and recent memoranda from the National Labor Relations Board (NLRB), including the now-rescinded GC 23-02, the paper outlines the risks posed by artificial intelligence, pre-employment screening tools, and continuous workplace monitoring. It traces the historical lineage of scientific management and surveillance back to earlier forms of labor discipline, such as racialized coercion and Pinkerton union-busting, arguing that contemporary tools reproduce similar hierarchies with greater opacity. The analysis highlights how current NLRB standards—developed in a pre-digital era—fail to adequately address the psychological, legal, and structural implications of 21st-century surveillance practices. Special attention is paid to how the political composition of the NLRB shapes the enforcement of labor law, with the rescission of dozens of pro-worker memos under the second Trump administration illustrating the fragility of regulatory protections. Ultimately, the paper calls for a more aggressive and strategic response by unions, legal advocates, and the Board itself to confront the inchoate crisis of digital labor control. Without such intervention, algorithmic management will continue to erode worker autonomy, collective action, and the possibility of democratic oversight in the workplace.

Introduction

In their quest to ruthlessly extract as much as possible from their workers, employers have begun to utilize cybernated tools that are quite literally inhuman in both formation and application. Without expanded protections, the impending transition to hyper-digital workplaces will lead to the widespread erosion of employees' rights. While these tools are often framed as neutral instruments of efficiency, they are frequently deployed to consolidate managerial power, suppress organizing efforts, and obscure the mechanisms by which employment decisions are made. Drawing on existing National Labor Relations Board precedent, legal scholarship, and recent developments in labor policy, the following analysis considers how digital management practices intersect with foundational rights under the National Labor Relations Act and identifies where current protections fall short. Though the technologies may be new, the patterns of power they reproduce are deeply familiar. Boldness, innovation, and disruption cannot be principles reserved solely for techno-oligarchs and their ilk; organizers and the NLRB must create and implement responsive strategies to counter this inchoate crisis.

Jennifer A. Abruzzo, who served as General Counsel to the National Labor Relations Board from 2021 until being ousted in early 2025, was well aware of this exigency. On October 31, 2022, she released a memorandum outlining a swath of issues surfaced or augmented by employers' use of "omnipresent surveillance and other algorithmic-management tools". [1] Abruzzo proposed a number of solutions, all of which were girded by a desire to utilize "well-settled Board principles" to mount a robust and proactive defense of workers' rights. [2] However, the memo was among more than thirty rescinded almost immediately by William Cowen, who President Trump appointed Acting General Counsel of the National Labor Relations Board on February 3, 2025, signaling a sharp shift in the Board's priorities and leaving a vacuum in guidance on these urgent issues. [3]

Screening Out Dissent: Pre-Employment Tools and Anti-Union Animus

As a class, employers demonstrate few qualms about using patently illegal means to squash unionization efforts. Amongst the most common of these unfair labor practices is the illegal discharge of workers for leading or participating in union drives. [4] While Section 8(a)(3) of the National Labor Relations Act states that it is unlawful for an employer "by discrimination in regard to hire or tenure of employment or any terms or condition of employment to encourage or discourage membership in any labor organization", the consequences for doing so are greatly outweighed by the benefits reaped by the company–a reality the PRO Act attempts to address, though it has repeatedly failed to advance to the Senate. [5] The Economic Policy Institute found that in nearly 30% of "NLRB-supervised elections" employers are charged with unlawfully discharging workers. [6] However, a more adroit way to hinder the formation of a union is to make sure that candidates who are likely to "agitate" for better working conditions or the formation of a union do not make it past the interview process. [7]

This is most frequently achieved through the administration of personality tests that prize applicants who demonstrate complacency about their working conditions and a high likelihood to tacitly accept injustice in their workplace. The earliest ancestors of the modern personality test were repurposed screeners for mental illness, leveraged in an attempt to replace the "Yellow Dog Contracts" rendered illegal by the Wagner Act. [8] Over the following decades, these tests were refined into bespoke screeners that could winnow out likely union activists. That there is "little correlation between job performance and such personality tests" indicates that employers' insistence on administering them–over two-thirds of job seekers are estimated to take personality tests each year–is likely driven by a desire not for the best workers but the most complacent ones. [9]

Employers are tenacious and inventive in their pursuit of methods to ascertain whether candidates have union sympathies, and personality tests are now just one method in a battery of tools designed for this purpose. In their article, The Invisible Web at Work, Law professors Richard Bales and Katherine Stone demonstrate that companies are turning to a suite of artificial intelligence-powered tools to sort and

analyze the deluge of employee data they have access to. As previously established, union aversion is one of the implicit standards for determining which candidates are good employees. AI is notorious for being marred by and replicating, or worse yet "amplify[ing]" human biases. [10] As former General Counsel Abruzzo noted in her now-rescinded memo, there is ample precedent indicating that "[e]mployers…violate Section 8(a)(1) if they coercively question employees with personality tests designed to evaluate their propensity to seek union representation." [11] Though her departure and the rescission of this and many related memos have curtailed the official stance of the General Counsel's office on such practices, the underlying precedents remain available for strategic use by unions and advocates. Expert opinion indicates that these algorithmic tests and screeners likely illegally assert anti-union animus through their internal mechanisms; however, these mechanisms would need to be further unveiled to definitively identify clear violations of the NLRA. [12] As will be expounded on later, unions may be best suited to take up this gauntlet given the existent case law regarding the application of Section 8(a)(5) and 8(d) of the Act.

Companies also utilize "algorithmic correlations" found within data culled from internal polls to sniff out the presence of "union sympathies" amongst their employees. [13] While the connection between the use of these internal polls and anti-union discrimination will remain tenuous until information from the aforementioned efforts to better elucidate their inner workings is discovered, there is established precedent that may be useful in finding the practice illegal should their intent to ascertain union sympathies be concretized. One such example is the supplemental decision issued in 1967 for Struksnes Construction Co., Inc,. [14] In this decision, the Board elucidates its stance on polling's role in chilling Section 7 activity, having seen "innumerable cases in which" it was "the prelude to discrimination" and noting that "any attempt by an employer to ascertain employee views and sympathies regarding Unionism…tends to impinge on…Section 7 rights."[15] The Board also clarifies that "Section 8(c)" cannot be construed as "protect[ing] an employer's efforts to discern, through polling or coercive interrogation, the union sentiments of employees." [16] As Nathan Newman notes in his

masterful article for the University of Cincinnati Law Review "[t]he Board and the courts have made clear that "polls that even indirectly elicit information on union sympathies are illegal." [17] [18] [19]

Collective Bargaining as a Bulwark Against Surveillance

Much of the focus here has been on the great mass of workers who face their bosses as atomized entities, that is the 93% of private sector employees without union affiliation. [20] However, it is crucial to note that collective bargaining offers some of the strongest protection against the encroachments of electronic surveillance. In Golden Stevedoring Co. even when the Board finds that an employer "did not tighten the application of disciplinary rules or expand their scope" changing the means by which "disciplinary messages" are disseminated is found to "constitute a significant change in a condition of employment which is a mandatory subject of collective-bargaining." [21] The Board's recognition of the "psychological impact of placing words on paper" on workers derived from the "aura of permanence" that recorded communication holds is particularly useful for developing protective frameworks around electronic management, given the cultural perception that digitally recorded information is perdurable. [22] Similarly beneficial is the Board's finding that the "installation of surveillance cameras is analogous to physical examinations, drug/alcohol testing requirements, and polygraph testing, all of which the Board has found to be mandatory subjects of bargaining." [23] In that same decision the Board states that while the company, Colgate-Palmolive, posits that "bargaining before a hidden camera is actually installed would defeat the very purpose of the camera" the "very existence of secret cameras…is a term and condition of employment, and is thus a legitimate concern for the employees' bargaining representative." [24] This ruling's underlying logic introduces a means for unions to fight the onslaught of digital surveillance by circumscribing a company's supposed right to secrecy in monitoring its employees. This could be particularly efficacious in combating the imposition of some of the most nebulous technologies, such as algorithms.

The Quantified Worker: Surveillance as Management

Putting the speculative matter of these pre-employment screenings and employee polls aside, there are many other means of "scientific management" that descend from some of the most hideous–though lawful at the time of their occurrence–labor practices seen in the United States, including the strategic monitoring of slaves to ensure that the most profit could be wrung from their bodies and the use of firms such as the Pinkertons to surveil and intimidate union workers. [25] Companies now have access to and utilize an assortment of technologies that track and analyze, amongst other things, workers' "microexpressions", proximity to each other, keystrokes, biometric data, and supposed productivity metrics. [26] Each of these inputs is mined, compiled, and run through proprietary algorithmic systems to produce assessments of performance, trustworthiness, or risk; none of which workers are permitted to interrogate or appeal.

These practices transform employees into data points, turning them into what law professor Ifeowa Ajunwa calls a "quantified worker." [27] Just as the early industrialists once used stopwatches to regiment human motion on the assembly line, the contemporary employer uses wearable sensors and automated monitoring to extend managerial power into every second of the workday. The result is not just intensified exploitation, but a qualitative transformation in the employment relationship. Workers are not merely observed but continuously modeled, predicted, and preemptively disciplined. Surveillance today is no longer a supplement to management; it is management.

An Inadequate Standard: The Law's Failure to Keep Pace

As the Board notes in Cannon Electric Company, "an employer cannot discriminate against union adherents without first determining who they are." [28] This axiom highlights surveillance's role in maintaining a power imbalance between workers and their employers. The Board's rulings on surveillance often center on its psychological effect on employees, taking the position that workers ought to be free

from concerns that "members of management are peering over their shoulders, taking note of who is involved in union activities, and in what particular ways." [29] The Board also finds "employer conduct that creates an impression of surveillance to be a violation of Section 8(a)(1) of the Act." [30] Panoptic workplace conditions violate both of these provisions. While "[t]he test for whether an employer unlawfully creates an impression of surveillance is whether under the circumstances, the employee reasonably could conclude from the statement in question that his protected activities are being monitored", former General Counsel Abruzzo took the more aggressive and sensible position that "an employer has presumptively violated Section 8(a)(1) where the employer's surveillance and management practices, viewed as a whole, would tend to interfere with or prevent a reasonable employee from engaging in activity protected by the Act." [31] As a prominent legal scholar Charlotte Garden asserts, the current standard for what constitutes "unlawful surveillance" grants employers the ability "to set the baseline" evinced by cases such as The Broadway, 267 N.L.R.B. 385, 400 (1983) and Metal Industries Inc., 251 N.L.R.B. 1523, 1523 (1980) wherein the Board found that companies did not commit unfair labor practices by surveilling "concerted activity" because the methods in question were established policies before workers began partaking in the activity and were therefore not "out of the ordinary." [32]

As demonstrated here, utilizing standards for ascertaining whether employer surveillance is unlawful developed in an era where surveillance was more limited and labor-intensive and is wholly inadequate in addressing the nightmarish technological impositions of the 21st-century workplace. The extant legal framework was never designed to account for the reach, speed, or opacity of contemporary data capture. All labor issues are connected to larger economic and sociological patterns. In this context, grotesquely invasive privacy violations are widely accepted and presented as an inevitability of the digital sphere despite end-user squeamishness. Employers no longer need to peer over shoulders to monitor workers, they can leverage technology that does so continuously, invisibly, and without recourse. The hegemonic force exerted by technological giants and the companies who stand to benefit from their might primes

large portions of the public for disempowerment in their roles as both producers and consumers. The normalization of grotesquely invasive privacy violations happens under the guise of natural extensions of consumer convenience or efficiency, reflecting broader ideological shifts that prioritize capital accumulation over personal autonomy. Functionally, this means workers are coerced into ceding an enormous amount of power to an abstruse and intangible web of systems. They are rendered knowable to their employers in excruciating detail, while the systems that evaluate and discipline them remain black-boxed and inaccessible. This asymmetry mirrors and intensifies existing power dynamics. It also forecloses one of the most basic preconditions for collective action: the ability to know what is happening, to whom, and why. Digitally surveilled workers will remain double victims of the data-driven economy unless the Board is dogged in its obligation "to adapt the Act to changing patterns of industrial life." [33]

Conclusion

The rescission of GC 23-02, along with more than two dozen memoranda, is more than a mere bureaucratic shift. The new General Counsel's actions amount to a strategic retreat from the Board's duty to evolve alongside the rapidly transforming workplace. This quietly executed deregulatory offensive has devastating implications. It is a probative measure, labor's portion of the "shock and awe" stratagem foundational to the second Trump term.

While the Board is ostensibly impartial, in practice, its slant is highly contingent upon the political comportment of its leadership. This is evinced by the oscillation between pro-labor and anti-labor stances that accompany shifts in presidential administrations. The Board's rulings, priorities, and enforcement decisions are shaped not by an unwavering commitment to the principles enshrined in the NLRA, but by the ideological proclivities of whoever is in power. As a result, the Board's effectiveness is brittle and subject to abrupt reversal. That such sweeping deregulation could occur in near silence, via internal memoranda and without legislative input, underscores how little democratic insulation workers truly have from the whims of political appointees. It

is not just the tools of surveillance that are opaque and unaccountable but the mechanisms of protection themselves.

Endnotes

[1] Jennifer A. Abruzzo, "Electronic Monitoring and Algorithmic Management of Employees Interfering with the Exercise of Section 7 Rights," Memorandum GC 23-02 (National Labor Relations Board, October 31, 2022), 1.

[2] Abruzzo, "Electronic Monitoring and Algorithmic Management," 6.

[3] William B. Cowen, "Rescission of Certain General Counsel Memoranda," Memorandum GC 25-05 (National Labor Relations Board, February 14, 2025).

[4] Celine McNicholas et al., Unlawful: U.S. employers are charged with violating federal law in 41.5% of all union election campaigns, Economic Policy Institute (2019), 2 https://www.epi.org/publication/unlawful-employer-opposition-to-union-election-campaigns/.

[5] Charlotte Garden, Enforcement-Proofing Work Law, 44 Berkeley Journal of Employment and Labor Law (2022).

[6] McNicholas et al., *Unlawful*, 2.

[7] Nathan Newman, UnMarginalizing Workers: How Big Data Drives Lower Wages and How Reframing Labor Law Can Restore Information Equality in the Workplace, SSRN Electronic Journal (2016), 693.

[8] Newman, "UnMarginalizing Workers," 710.

[9] Newman, "UnMarginalizing Workers," 707.

[10] Richard A Bales & Katherine V.W. Stone, The Invisible Web at Work: Artificial Intelligence and Electronic Surveillance in the Workplace, 41 Berkeley Journal of Employment & Labor Law (2020), https://lawcat.berkeley.edu/record/1181483?ln=en., 22.

[11] Abruzzo, "Electronic Monitoring and Algorithmic Management," 4.

[12] Newman, "UnMarginalizing Workers," 710.

[13] Newman, "UnMarginalizing Workers," 739.

[14] Struksnes Construction Co., Inc., 165 NLRB 1062(1967)

[15] Allegheny Ludlum Corporation v. N.L.R.B, 301 F.3d 167 (3d Cir. 2002), 1359.

[16] Allegheny Ludlum Corp., 333 NLRB 734, 740 (2001), 737.

[17] Newman, "UnMarginalizing Workers," 739.

[18] *Midwest Regional Joint Board, Amalgamated Clothing Workers of America v. NLRB, 564 F.2d 434, 438 (D.C. Cir. 1977); see also Burns International Security Services, Inc., 225 N.L.R.B. 271, 274 (1976), enforcement denied, 567 F.2d 945 (10th Cir. 1977).*

[19] *Contractor Services, Inc., 324 N.L.R.B. 1254, 1255 (1997).*

[20] U.S. Bureau of Labor Statistics, "Union Members—2022," news release, January 19, 2023.

[21] Golden Stevedoring Co., Inc., 335 N.L.R.B. 410, 426 (2001).

[22] *Golden Stevedoring, 426.*

[23] Colgate-Palmolive Co., 323 N.L.R.B. 515, 515 (1997).

[24] *Colgate-Palmolive*, 516.

[25] Esperanza Fonseca, Worker Surveillance Is on the Rise, and Has Its Roots in Centuries of Racism, Truthout (2020), https://truthout.org/articles/worker-surveillance-is-on-the-rise-and-has-its-roots-in-centuries-of-racism/ (last visited Dec 17, 2023).

[26] Bales and Stone, "Invisible Web at Work," 12.

[27] Ifeoma Ajunwa, The Quantified Worker: Law and Technology in the Modern Workplace (Cambridge: Cambridge University Press, 2023), 1.

[28] Cannon Electric Co. and Charles H. Warren, 151 N.L.R.B. 1465, 1468 (1965).

[29 Flexsteel Industries, Inc., 311 N.L.R.B. 257, 257 (1993).

[30] Delmas Conley d/b/a Conley Trucking, 349 N.L.R.B. 308, 315 (2007).

[31] Abruzzo, "Electronic Monitoring and Algorithmic Management," 8.

[32] Charlotte Garden, "Labor Organizing in the Age of Surveillance," St. Louis University Law Journal 63 (2018): 62.

[33] NLRB v. J. Weingarten, Inc., 420 U.S. 251, 266 (1975).

Bibliography

Abruzzo, Jennifer A. "Electronic Monitoring and Algorithmic Management of Employees Interfering with the Exercise of Section 7 Rights." Memorandum GC 23-02. National Labor Relations Board. October 31, 2022.

Ajunwa, Ifeoma. The Quantified Worker: Law and Technology in the Modern Workplace. Cambridge: Cambridge University Press, 2023.

Bales, Richard A., and Katherine V. W. Stone. "The Invisible Web at Work: Artificial Intelligence and Electronic Surveillance in the Workplace." Berkeley Journal of Employment and Labor Law 41, no. 1 (2020): 1–61.

Cowen, William B. "Rescission of Certain General Counsel Memoranda." Memorandum GC 25-05. National Labor Relations Board. February 14, 2025.

Fonseca, Esperanza. "Worker Surveillance Is on the Rise, and Has Its Roots in Centuries of Racism." Truthout. December 17, 2020.

Garden, Charlotte. "Enforcement-Proofing Work Law." Berkeley Journal of Employment and Labor Law 44 (2023): 191–250.

Garden, Charlotte. "Labor Organizing in the Age of Surveillance." St. Louis University Law Journal 63 (2018): 55–68.

McNicholas, Celine, Margaret Poydock, Julia Wolfe, Ben Zipperer, Gordon Lafer, and Lola Loustaunau. Unlawful: U.S. Employers Are Charged with Violating Federal Law in 41.5% of All Union Election Campaigns. Economic Policy Institute. December 11, 2019.

Newman, Nathan. "UnMarginalizing Workers: How Big Data Drives Lower Wages and How Reframing Labor Law Can Restore Information Equality in the Workplace." University of Cincinnati Law Review 85, no. 3 (2017): 693–746.

U.S. Bureau of Labor Statistics. "Union Members—2022." News release. January 19, 2023.

The Causal Ligament: Corporate and Agency Law and the Autonomous Weapons Accountability Gap

Lauren E. Garrett

" THE LIFE OF THE LAW HAS NOT BEEN LOGIC:
IT HAS BEEN EXPERIENCE. "

Abstract

When an autonomous weapons system misclassifies civilians as combatants and destroys them, the frameworks most commonly invoked return a verdict that is, beneath their institutional veneer, a verdict of silence: command responsibility requires proof of command knowledge that autonomous targeting dissolves by design, and the policy aspiration of meaningful human control, however prolifically circulated, prescribes no operational threshold and furnishes no forum with anything a plaintiff can take to court. This Article proposes that corporate and agency law, brought to bear alongside IHL rather than in its place, supplies accountability mechanisms the humanitarian framework cannot produce unaided. The principal-agent relationship applies to autonomous weapons deployment with documentary precision; respondeat superior fastens liability without the intent requirements autonomous targeting renders permanently unavailable; and the retained control and inherently dangerous activity exceptions reach manufacturers and developers under existing domestic law, each subject to the qualification that courts have historically resisted applying these doctrines to battlefield sovereign conduct and that the argument for extending them must be earned rather than assumed. Against deploying states, both IHL and corporate tort law encounter the same sovereign immunity wall, and this Article identifies a treaty pathway whose structural design questions existing international instruments have already answered, a pathway made politically conceivable by the commercial pressure that transformed asbestos, tobacco, and pharmaceutical litigation into regulatory reform, though its realization remains contingent on political will this Article cannot supply.

Introduction: The Frameworks Governing Autonomous Weapons Accountability and Why They Fail

"The life of the law has not been logic: it has been experience."[1] Holmes wrote those words in 1881, a man who had bled for the republic at Ball's Bluff and Antietam before he read its law until it yielded, and they illuminate with uncomfortable precision the central failure this Article addresses. The frameworks that presently govern accountability for autonomous weapons are architectures of responsibility erected upon the bedrock assumption of human moral agency, pressed into service against a technology that dissolves that assumption on contact and perpetuated by the very institutions whose discomfort with the resulting gap has not yet risen to the pitch that compels change.

Autonomous weapons systems, weapons capable of selecting and engaging human targets without meaningful human intervention at speeds and through processes their operators cannot follow in real time, present a legal problem of precisely the kind that international humanitarian law was forged to address yet, by reason of its foundational premises, lacks the doctrinal architecture to resolve.[2] The scholarship that has accumulated around that observation is now substantial in both depth and urgency.[3] This Article departs from that literature's prevailing orientation, arguing that corporate and agency law, brought to bear alongside IHL rather than in its place, supplies accountability mechanisms that the humanitarian law framework cannot produce unaided. The argument is offered as a doctrinal framework for a genuinely novel problem: no court has yet applied respondeat superior to an autonomous weapons engagement, no treaty has yet codified the liability architecture proposed in Part VIII, and the case law on which the analogy rests was developed without autonomous weapons in mind. This Article accordingly proceeds with the intellectual seriousness that unexplored legal territory demands, proposing frameworks where doctrine permits and acknowledging uncertainty where it does not, with the aim of advancing the analysis rather than foreclosing it.

The scenario that follows is labeled a hypothetical, but its operative facts have already occurred in recognizable form. A sovereign state deploys an autonomous weapons system into a contested urban environment, charged with identifying and lethally engaging hostile combatants. The system, executing its mission within the bounds of its programmed operational parameters, misclassifies a cluster of civilians as combatants and destroys them. Nearly 50 people perish, and international humanitarian law, upon examination, returns a verdict of silence dressed in the language of doctrine.[4]

Command responsibility doctrine, as enshrined in Articles 86 and 87 of Additional Protocol I, conditions superior liability upon a showing that a commander "knew or had reason to know" that subordinates were perpetrating violations.[5] The standard was forged for a world in which the causal ligament between command and consequence, however attenuated, remained at least traceable, where some human officer could be shown to have known or to have had reason to know that a violation was occurring.[6] A weapons system whose targeting determination is reached in milliseconds through algorithmic processes opaque even to its designers[7] dissolves that ligament entirely, leaving the doctrine without the human object its knowledge requirement demands. One might argue that the constructive knowledge standard captures the deployment decision itself: a state that knowingly deploys a system prone to civilian misclassification arguably "had reason to know" at the moment of procurement, not the moment of the strike. The argument is superficially appealing but structurally deficient. Command responsibility's knowledge element runs to the specific violation being committed by a specific subordinate, not to a class of technological risk accepted at procurement; stretching the doctrine to reach deployment-level risk acceptance would transform "had reason to know" into something indistinguishable from strict liability, a standard the text does not contemplate and the drafting history does not support. The doctrine was not designed to operate at the level of generality autonomous weapons require.[6] The dominant policy alternative offers no more purchase. "Meaningful human control," the phrase inhabiting DoD Directive 3000.09[8] and propagating through an ever-expanding corpus of UN and NGO pronouncements,[9] is a policy aspiration that its frequent

invocation has dressed in the language of legal obligation.[10] The phrase, for all its institutional circulation, prescribes no operational threshold, offers no criterion by which meaningful control is distinguished from its nominal simulacrum in real-time deployment, and provides no retrospective mechanism for determining whether any given engagement complied. A formulation capacious enough to shelter both the soldier who pulled the trigger and the official who approved the targeting algorithm two years earlier is performing the appearance of legal work while the substance of accountability remains absent.

Robert Sparrow gave the problem its most lucid philosophical formulation: when an autonomous weapon perpetrates what would, in any other context, constitute a war crime, "there is no one who can appropriately be held responsible" for this. 11 The literature proceeding from Sparrow's diagnosis is consequential, and this Article departs from it to redirect the inquiry from the philosophical to the doctrinal. Scholars working in AI liability have gestured toward private law as a productive frame, but the specific architecture of corporate and agency doctrine applied to state-deployed lethal autonomous systems, complete with the sovereign immunity problem it must navigate and the manufacturer liability chain it traces, has not been developed.12 IHL is structured to ask whether a specific human being made a legally cognizable decision, a question that autonomous weapons dissolve by removing the human decision-maker at the precise moment that question would apply. Corporate and agency law redirects the inquiry to ground where autonomous weapons offer no such resistance: whether the harm-causing system operated within the scope of authority its principal conferred, a question to which the documentary record of any deployment can, in principle, supply an answer.

The body of law this Article proposes to bring alongside IHL is corporate and agency law. Each doctrine it deploys, reaching from respondeat superior through the independent contractor exceptions to strict liability for abnormally dangerous activities, has governed non-human actors causing harm on behalf of identifiable principals for the better part of two centuries without requiring either the moral agency of a natural person

or the guilty mind of an intentional tortfeasor. The framework is offered as a rigorous doctrinal proposal for genuinely novel territory, and intellectual honesty about that novelty is both a methodological commitment and a condition of the argument's credibility. Where doctrine applies directly, this Article argues the application; where extension is required, the argument for extension is made and its contestability acknowledged; where courts have not yet ruled, the analysis projects from existing principle and says so. The case proceeds along two dimensions, distinct in reach but connected in what they require. Against manufacturers, prime contractors, and AI developers, the corporate doctrine operates under existing domestic law in courts that can act today without sovereign immunity as an obstacle. Against the deploying state, it encounters the same wall IHL approaches from a different direction, and the answer lies in a Protocol whose design questions existing international instruments have already answered, though its ratification remains contingent on political will the doctrine alone cannot supply. These two dimensions are not alternatives. Sustained civil litigation against private defendants creates the commercial incentive structure that makes a sovereign accountability Protocol politically conceivable, the pressure generated by the first that makes the second possible at all. The argument proceeds across eight Parts and a Conclusion. Part I applies the principal-agent framework to autonomous weapons deployment and tests it against its strongest objections, including the institutional resistance courts have demonstrated to applying private law frameworks to battlefield conduct. Parts II through IV develop the liability framework in sequence, establishing respondeat superior, the manufacturer liability chain, and strict liability as independent and overlapping grounds for recovery, each with explicit acknowledgment of the doctrinal distance between established precedent and the proposed application. Part V addresses the ultra vires problem and the foreseeability question it generates for ML-based targeting. Part VI confronts sovereign immunity directly. Part VII examines the incentive structures that have kept the corporate law framework from autonomous weapons accountability, and the mechanism by which private defendant litigation generates conditions for sovereign accountability. Part VIII identifies the elements a

functional AWS Accountability Protocol must incorporate. The Conclusion closes on the questions the framework generates and the doctrine that awaits them.

I. The Anatomy of Delegation: The Principal-Agent Structure of Autonomous Weapons Deployment and Its Doctrinal Application

The Restatement (Third) of Agency defines the agency relation as arising when one person (the principal) manifests assent to another (the agent) to act on the principal's behalf and subject to the principal's control, the agent in turn consenting to so act.13 Three definitional elements obtain: a manifestation of assent by the principal to the agent's acting on its behalf; actual action by the agent in furtherance of the principal's purposes; and subordination of the agent's conduct to the principal's directional authority.14 Each element maps onto the relationship between a deploying state and its autonomous weapons system with a precision that suggests the analogy was always available, had the field chosen to pursue it.

A state or military entity that deploys an autonomous weapons system evinces its assent in the most unequivocal register available to legal analysis. It encodes the system's targeting parameters and calibrates its rules of engagement to prescribed operational criteria; the activation order itself specifies the geographic and strategic environment within which those directives operate. 15 The deployment order is an affirmative grant of authority, its content defined and bounded by the mission guidelines that the principal establishes. The system pursues the principal's military objectives in the principal's operational theater, wielding lethal force that the principal, as the recognized sovereign actor, possesses the legal authority under international law to direct, a relationship of agency that expresses itself with greater fidelity to the Restatement's three elements than most relationships the doctrine was designed to govern.

Agency law has, throughout its long development, imposed no requirement of natural personhood upon the occupant of the agent position.16 Corporations have functioned as agents since the commercial law first required them to; trusts, limited partnerships, and other non-human legal constructs occupy the role with quotidian

regularity, without generating doctrinal difficulty.17 The Restatement's animating concern is with the functional topology of the relationship (who acts for whom, under whose authority, for whose benefit) and not with the ontological category of the actor.13 An autonomous weapons system acts for the state, pursuant to the state's commission, in single-minded pursuit of the state's operational objectives. Whether the system harbors anything a philosopher would call intention is beside the legal point, just as a corporation's "intent" is an imputed legal construct rather than a report on any interior state.

The Restatement's consent requirement merits direct engagement because it is the objection most likely to be pressed. Agency requires that the agent "manifest assent or otherwise consent so to act," and a machine cannot consent.14 The answer does not require resolving any deep question about machine cognition because the consent element is satisfied by the humans who constitute the organizational actor whose conduct is being attributed, not by the system itself. A corporation does not itself "consent" when it acts as an agent; the consent is constructed from the acts of the human officials permitted to bind the institution.16 When a state deploys an autonomous weapons system, human officials throughout the deployment chain, from the program officers who approved the acquisition through the commander’s sanction of its activation to the field operators who activated the system, have each expressed consent on behalf of the representative principal that their authority to bind the institution makes legally operative. The AWS is the instrument through which that institutional consent is executed in the world. The analysis itself is, granted, unremarkable, as it is the same reasoning applied to every corporate agency relationship. The observation that a machine cannot consent simply redescribes the architecture of institutional agency, which has never required consent from the physical medium through which formal decisions are expressed.14

The closest existing authority for applying agency attribution principles to non-juridical automated systems emerges from patent and telecommunications regulatory contexts rather than tort law, a provenance that requires explicit acknowledgment and

does not eliminate the doctrinal distance the AWS argument must cross. In Akamai Technologies, Inc. v. Limelight Networks, Inc., the Federal Circuit en banc expressly derived its standard for attributing distributed performance to a single actor from "general principles of vicarious liability," holding that one who directs or controls the execution of tasks by others, including automated processes, is chargeable with all resulting performance; the court itself acknowledged "vicarious liability is not a perfect analog" in the patent context while nonetheless applying the underlying attribution logic of direction-and-control to conduct executed through distributed non-human intermediaries.93 The Federal Communications Commission reached an analytically parallel conclusion in In re Dish Network, LLC, applying federal common law agency principles, including actual authority, apparent authority, and ratification, to determine when a principal bears liability for violations committed through third-party automated telemarketing systems operating on its behalf.93 Neither precedent involves tort liability, and neither court held that an automated system constitutes a legal agent in the full Restatement sense. What they do collectively establish is that the attribution logic underlying agency law has been applied, by a federal circuit court and a federal regulatory body, to harm-causing conduct executed through automated systems operating within a principal's sanctioned delegation. The AWS argument requires extending that imputation rationale into a tort context against sovereign actors, a step neither Akamai nor Dish Network takes and whose demonstrable novelty this Article does not minimize. The foundation of the argument now rests on something more substantial than the Restatement's silence on natural personhood, on a recorded pattern of courts and regulatory bodies extending agency attribution principles to activity executed through automated systems, a pattern whose basic logic points toward the AWS application even if no tribunal has yet arrived there.

A more rudimentary objection than the consent question also warrants a check. It holds that an autonomous weapons system is a tool rather than an agent, and that no elaboration of the Restatement's personhood analysis changes that categorical fact. A missile is a tool. A land mine is a tool. A hammer is a tool. The objection has intuitive

force, and that force is precisely what makes it worth examining, because the legal distinction between a tool and the instrument of an agency relationship does not depend on the sophistication of the object. It depends on the structure of the authorization relationship within which the object operates. No one encodes a hammer's operational parameters, specifies its engagement criteria in a procurement document subject to civil discovery, retains authority to override or withdraw it during performance, or bears documentary accountability under federal directive for its operational decisions. The deploying state does all of these things with respect to its autonomous weapons systems.94 What places an AWS in the agent position of an agency relationship is the configuration of the deployment relationship itself, not the sophistication of the object. That the system processes information and makes classifications has no bearing on that analysis; a thermostat does both without anyone suggesting it occupies the agent position in an agency relationship. The deployment relationship presents each feature agency doctrine treats as constitutive with a principal conveying assent through inscribed operational parameter specification, a defined scope of authorized conduct bounding the system's operation, human officials at every level of the command chain binding the organizational actor through decisions within their authority, and directional control persisting through rules of engagement and override capacity retained throughout performance.14 The Restatement's agency definition tests the relationship's composition, not the instrument's sophistication.94 Against a missile, that relational arrangement does not obtain, but against an autonomous weapons system deployed under a documented operational mandate by human officials exercising institutional authority, it does. The "mere tool" objection, examined against what the Restatement actually requires, resolves into a question the deployment record can, from the documentary record it creates, answer: whether this system was operating within the scope of authority the deploying state conferred.

Agency law distinguishes between actual authority, meaning authority the principal has expressly or impliedly vested, and apparent authority, which concerns third-party reliance on a principal's representations about an agent's scope.18 For

autonomous weapons systems, actual authority is the operative concept; apparent authority has no purchase where no commercial third parties exist to rely on representations about the system's sanctioned ambit of conduct. The system's operative boundaries are fixed by what the deploying state has actually granted, enumerated in the acquisition and deployment mandate with sufficient precision to render the delegation's scope a matter of archival record. A system whose conduct remains within that delegated ambit acts with actual authority; conduct that exceeds it invokes the ultra vires analysis developed in Part V.

As a practical matter of litigation mechanics, experienced plaintiffs' counsel in cases of this complexity would join every party in the liability chain as a defendant simultaneously, encompassing the deploying state, the prime systems contractor, the AI developer, relevant subcontractors, and, at the furthest terminus of that sequence, the companies whose foundational AI architectures underpin the targeting system's classification logic, relying upon contribution and indemnification proceedings among the named defendants to apportion responsibility with appropriate granularity once the threshold liability question is resolved. The approach is orthodox mass tort strategy, the same practice courts have employed to complex military products litigation for decades.20

The analogy faces objections, and engaging them directly is an exigency of the argument. One line of resistance holds that corporate law is constitutively too transactional to bear the normative weight that armed conflict dictates. The gravity of armed conflict makes the answerability gap more intolerable rather than less consequential; the more serious the harm, the stronger the case for a regime that delivers redress rather than its simulacrum. A second contends that scope of employment is too indeterminate a standard in combat to furnish workable liability rules, yet that indeterminacy pervades every proposed accountability architecture for autonomous weapons, and the agency construct at minimum presents courts with an analytical question, being what the deployment order actually authorized, rather than

the structurally irresolvable one that IHL bequeaths them, concerning what the commander actually knew. A more fundamental challenge merits candid acknowledgment, given that courts assess not only whether an analogy technically maps but whether extending doctrine in this direction is institutionally sound. There is deep-seated, legitimate resistance to applying private law constructs to battlefield sovereign conduct, and this Article does not dismiss it. The framework proposed here requires courts to take a step that no court has yet taken, and the argument for taking it rests on the proposition that the accountability gap it would close is more damaging to the legal order than the doctrinal extension it requires. That proposition is contestable; this Article argues it is correct. The third objection, sovereign immunity, carries sufficient magnitude to warrant its own Part, to which this Article turns in due course.21

II. Respondeat Superior: The Liability Chain and Its Application to Autonomous Systems

Respondeat superior ("let the master answer") imposes vicarious liability upon a principal for harm visited upon third parties by an agent acting within the scope of the agent's authority.[22] The doctrine is among the most venerable in Anglo-American jurisprudence, grounded in a normative foundation both utilitarian and intuitive, holding that those who deploy agents for their own benefit, who define the ambit of those agents' authority, should bear the costs when that authority occasions harm.[2324] The doctrine registers no concern for the principal's intent, the principal's actual knowledge, or whether the principal had any opportunity to intervene.[25] Its operative inquiry resolves to a question asking only whether the agent was performing the principal's work when the harm befell a third party.

Applied to autonomous weapons deployment, respondeat superior narrows the liability question to a determinate examination of whether the system acted within the scope of its delegated mandate at the moment it caused harm. [26] Where the system executed a strike within the geographic boundaries, target classification criteria, and weapons parameters of its programmed engagement envelope, liability fastens to the deploying state without further inquiry into whether the targeting decision was tactically

sound, proportionate, or compliant with any other extrinsic standard. Command responsibility, as the prior Part has established, demands proof that the commander "knew or had reason to know" of subordinates' violations;[5] respondeat superior demands only proof of deployment and proof that the conduct fell within the scope of the delegation, both verifiable in the ordinary case by documentary record rather than judicial inference.[27]

This analytical framework transforms rather than forecloses factual investigation, redirecting the investigative apparatus toward questions amenable to determinate resolution. Rather than probing a commander's subjective knowledge, an inquiry inherently difficult to verify and easily obscured, investigators examine deployment contracts, operational parameter specifications, system testing records, rules of engagement documentation, and behavioral logs of system activity at the relevant time.[28] These are matters of evidentiary record, subject to the ordinary processes of civil discovery.[29] The inquiry shifts from a forensic excavation of subjective mental states into a comparative audit of what the delegation authorized measured against what the system did. Where evidence emerges, independently, that a commander possessed advance knowledge of a specific strike or harbored affirmative intent bearing upon the deployment decision, that evidence does not displace the respondeat superior claim; it augments it with an independent layer of liability that runs alongside the vicarious theory without supplanting it.

The one doctrinal complication of moment is the frolic and detour rule, which denies principal liability for harm occasioned by an agent who has abandoned authorized conduct in favor of purely personal purposes, the servant whose deviation from employment is so complete as to sever the juridical ligament between master and act.[30] An autonomous weapons system that "malfunctions" has not embarked upon a frolic in any sense the doctrine was conceived to address. The system has not forsaken the principal's purposes; it has, at worst, pursued those purposes with a fidelity imperfect enough to produce unintended consequences. The foreseeability examination centers on the risk category as a whole rather than the particular instance. Autonomous

targeting systems, as a class, were recognized to produce erroneous engagement decisions in adversarial environments; they were not merely foreseeable but affirmatively foreseen, memorialized in the deploying state's own testing protocols and acknowledged in the foundational policy instruments governing AWS deployment. 3132 The applicable standard demands only that the general category of harm fall within reasonable contemplation, a threshold that asks nothing of the precise instance. 31 The behavioral volatility of complex AI systems operating in dynamic adversarial environments is a well-established and anticipated foreseeable characteristic of the technology's intrinsic composition, an acknowledged and accepted condition of the technology's operation, recognized across every level of its deployment as intrinsic to the medium rather than symptomatic of any particular defect.[31] Deploying an AI system whose conduct in both known and novel environments cannot be fully specified in advance is, in a morally and legally significant sense, incommensurable with deploying an employee who makes an unexpected personal choice.31 The deploying state knew, or by reason of its own testing and risk documentation should have known, that behavioral unpredictability inhabited the system's operational envelope from the moment of deployment.[32] That antecedent knowledge extinguishes the frolic defense before it can be interposed.

III. The Body The Law Can Reach: Manufacturers, the Delegation Chain, and the Failure of the Boyle Defense

The corporate and agency framework developed in Parts II and III does not require a treaty or legislative action to become operative. For one category of defendants (the manufacturers, prime contractors, and AI developers who build autonomous weapons systems), it applies directly in domestic courts under existing doctrine, without sovereign immunity, because those defendants hold no governmental immunity to assert. The claim against the private defendant represents the immediately cognizable remedy, the relief courts can extend today under law that already exists, without waiting for any state to consent to any accountability mechanism.

How the AWS developer or manufacturer is situated within this analytical framework is a matter of substantial practical consequence. Autonomous weapons development unfolds through extended sequences of delegation, beginning with procurement specifications in government acquisition documents, proceeding through prime systems contractors and AI developers who supply or construct the targeting logic, and reaching component suppliers whose contributions are embedded at every level of the system's architecture. International humanitarian law possesses no conceptual apparatus adequate to navigate the liability implications of so intricate a sequence;9 corporate and agency law offers one that is both developed and, in its essential logic, directly opposite.

The default rule is that principals incur no vicarious liability for the tortious acts of independent contractors, meaning parties who provide services to a principal but exercise their own professional judgment in determining the manner and means of performance, without being subject to the principal's direction in those particulars.[33] Were AWS manufacturers properly characterized as independent contractors, a deploying state might argue that liability exhausts itself at the manufacturer's threshold, leaving no claim to traverse the intervening commercial relationships. That argument encounters two well-developed exceptions that apply with particular force to autonomous weapons procurement.

The retained control exception provides that a principal who maintains meaningful control over the operative details of an independent contractor's work (control over the manner and means of performance, not merely the desired outcome) forfeits the protection of the independent contractor classification.[34] The governing threshold separates control over how a contractor performs from mere specification of what the contractor must achieve; the former triggers the exception, the latter does not. Retained control under Restatement § 56 requires ongoing authority over the methods and execution of performance. In the autonomous weapons procurement context, the relevant control extends well beyond initial performance specifications. States retain ongoing authority over the conditions of actual deployment, determining when and

where the system is activated, establishing the rules of engagement governing its conduct, and holding reserved power to override or terminate it mid-engagement. DoD Directive 3000.09 explicitly requires that program offices maintain documentation of AI lifecycle activities, a continuing supervisory obligation rather than a one-time specification.8,35 This continuing supervisory authority over the conditions of the system's field deployment constitutes retained control over the mode of its performance, extending beyond mere specification of desired outcomes. States that procure systems through commercial off-the-shelf arrangements with less direct developmental involvement may occupy a weaker position on the retained control analysis, though they remain fully exposed to the inherently dangerous activity exception, which requires no showing of control at all.[35] In any commercially meaningful sense, the relationship is better understood as a hierarchically structured enterprise in which the state principal exercises directional authority both at development and through the command sequence that authorizes each deployment, rather than as an arm's-length transaction between a passive buyer and an independent supplier. The retained control exception, applied to this relationship, would sustain direct liability without evident doctrinal difficulty. The same passive/active distinction that courts have begun applying in AI content liability cases, distinguishing between a platform that merely hosts what others produce and a developer that shaped the harm's possibility through training data selection, model architecture, and deployment design, maps onto the retained control analysis with equal precision; the developer who built and configured the targeting logic cannot claim to be a passive instrument that the operator simply pointed at a target.87

The inherently dangerous activity exception furnishes an independent and self-sufficient basis for the deploying state's direct liability to victims.[36] The doctrine holds that where a principal engages a contractor to perform inherently dangerous work, the principal cannot invoke the contractor relationship as a shield against liability to those the dangerous work harms. Restatement (Second) of Torts § 427 makes the hiring party (the deploying state) directly liable to injured third parties regardless of whether the work was formally delegated to a contractor. The exception does not trace liability

upward from manufacturer to state; it prevents the state from escaping direct liability to victims by pointing to the contractor structure. The state that deploys an AWS is the party that put the inherently dangerous activity in motion, and § 427 holds that party liable to the people the activity harms, independent of any question about manufacturer liability in the same procurement structure.[37] A weapons system purpose-built to locate, autonomously classify, and lethally engage human beings presents no serious difficulty of categorization; the label "inherently dangerous" applies with full doctrinal force, and no credible argument exists for excluding it.36 The deploying state accordingly cannot shelter behind the development contract to defeat the third-party victims' direct claim against it.

The government contractor defense that manufacturers will invoke, the three-part test established in *Boyle v. United Technologies Corp.*, rests on a foundational assumption that passes without examination in every case where the defense succeeds, namely that a specification document exists which captures, with sufficient precision, the operative behavior of the system it describes. For deterministic hardware with predictable performance envelopes that assumption holds. A missile's trajectory can be modeled, specified, and approved with engineering precision. Whether an ML targeting system's classification behavior in adversarial environments can satisfy *Boyle*'s first prong (government approval of reasonably precise specifications) is an unresolved question that future litigation will force courts to confront. The better answer, this Article submits, is no, but the argument requires stating with precision what that answer requires.[38] *Boyle*'s first prong is satisfied when the government approved the specific behavior that caused the harm, a requirement that cannot be met by approval of the system's general performance thresholds alone. A state that approved a targeting algorithm two years before a strike has approved training data, model architecture, and performance benchmarks on a training distribution. Approval of the training regime, however thorough, cannot extend to the system's specific classification behavior in the adversarial environment it encounters at deployment, since that behavior emerges from interactions between the trained model and real-world inputs that no specification

document can exhaustively predict. Courts applying *Boyle* will need to determine whether approval of a training regime constitutes approval of emergent deployment behavior; the technology's defining characteristic is precisely behavioral emergence across novel inputs, and the case law has not yet engaged that question with any resolution. The *Boyle* defense for ML-based autonomous weapons systems confronts its most fundamental obstacle in the inadequacy of its governing assumption, an inadequacy this Article advances through argument rather than assumes from existing doctrine.

With the government contractor defense addressed, the aggregate practical implication of the framework comes into view. Liability extends from the deploying state through the prime developer to component manufacturers and, at the furthest terminus, to the companies whose foundational AI architectures animate the targeting logic; the precise apportionment of responsibility among these parties is to be determined through contribution and indemnification proceedings among the defendants themselves.[39] Environmental liability, pharmaceutical liability, and defense contractor liability have been structured and adjudicated on exactly this basis in domestic tort law for decades, without generating the existential doctrinal difficulties that critics of the AWS application invariably predict.[20] What merits scrutiny is the failure rather than the result. The same framework that governs a petrochemical facility's network of subcontractors has, without satisfactory explanation, never been seriously advanced for accountability among the developers of autonomous lethal systems, a gap whose persistence in the scholarly literature is as difficult to account for as its absence from the courtroom.

IV. Without Fault, Without Nerve: Strict Liability for Abnormally Dangerous Activities as the Doctrinal Backstop

Where the principal-agent framework leaves interstices (and no doctrinal framework of finite scope is without them), strict liability for abnormally dangerous activities stands ready to occupy them. The Restatement (Second) of Torts establishes that one who carries on an abnormally dangerous activity bears strict liability for

resulting harm, even if every reasonable precaution has been taken.[40] The doctrine severs, by design, the connection between fault and liability, constructed to operate in those contexts where requiring proof of intent or negligence would immunize the very actors whose choices set the dangerous enterprise in motion.[41] Its operative inquiry carries a simplicity that is itself formidable, asking only whether the activity was abnormally dangerous and whether that danger materialized into harm, with nothing further required of the plaintiff.

The Restatement's six-factor examination for abnormal dangerousness yields consistently supportive results when applied to autonomous weapons deployment, and the analysis is offered as a framework for future adjudication rather than a prediction of how courts will necessarily rule.42 The specific doctrinal mechanism by which courts have historically resisted extending the doctrine to military and government activities warrants direct statement. The sixth factor, which asks whether the activity's value to the community is outweighed by its dangerous attributes, has consistently been resolved in favor of defendants engaged in defense and national security applications, on the ground that the state's interest in maintaining military capability constitutes social value of the highest order. No court has applied the abnormally dangerous activity doctrine to sovereign military weapons deployment. Gerald W. Boston's documented account of the doctrine's trajectory toward "near extinction" through what he calls the negligence barrier, meaning courts' increasing tendency to find that risks can be eliminated by the exercise of reasonable care, identifies exactly the trend that the ML unpredictability argument is designed to circumvent; where behavioral unpredictability in adversarial environments is an irreducible intrinsic feature of the technology rather than a correctable defect, the negligence barrier cannot close, and the Restatement's rationale for strict liability applies in full. 95 With those constraints stated directly, the remaining factors paint a coherent picture. The risk of serious harm is constitutive of the system's purpose, inseparable from its design at every level. The gravity of potential harm admits no gradation; the harm the system inflicts is death, irreversible and beyond remedy. The risk cannot be eliminated through better engineering because behavioral

unpredictability in dynamic environments is an endemic property of the technology itself, arising from the nature of the system rather than any deficiency in its construction.[31] Autonomous weapons deployment does not constitute common usage by any standard the factor contemplates.[43] The equilibrium between social value and dangerous attributes is the factor on which the argument's practical fate most depends, and this Article's claim is the circumscribed one, namely that the specific combination of behavioral unpredictability and irreversibility of harm that characterizes AWS deployment is factually distinguishable from the defense activities courts have previously declined to subject to strict liability, and that courts willing to engage with the distinction rather than the category would find the six-factor calculus materially different from the cases in which sovereign military conduct has escaped the doctrine's reach.

"Even a dog," Holmes wrote with characteristic economy, "distinguishes between being stumbled over and being kicked."[44] The aphorism reaches toward something of enduring jurisprudential significance, observing that even the most elementary legal intuitions have always organized themselves around the actor's relationship to the harm. Strict liability inverts that orientation expressly, forged because certain categories of dangerous enterprise are ones where a fault requirement would systematically insulate the very actors whose choices set them in motion. In the AWS context, requiring proof of intent or knowledge produces the accountability gap; strict liability forecloses it.[45] The syllogism that emerges tends to occasion resistance in proportion to its directness. The deploying state introduced an abnormally dangerous activity into the world; the characteristic danger of that activity materialized into harm; the deploying state is liable. That directness may itself explain the syllogism's absence from the discourse, as it offers no refuge for the sophisticated defendant and no occasion for the elaborate doctrinal maneuvering that complexity invites.

V. The Outer Boundary: Ultra Vires Acts, Foreseeability, and the Machine Learning Problem

A more intricate doctrinal question presents itself when an autonomous weapons system's conduct transgresses the outer limits of its authorized scope rather than merely underperforming within it. Geographic incursion into zones the commission explicitly withheld, hostile classification of persons whose engagement was expressly proscribed, and deployment of munitions outside the sanctioned weapons configuration each represent conduct of this character. The directives that circumscribe the system's authorized conduct constitute the precise boundary of the principal's delegation, separating what the principal permitted from what the principal withheld. These cases engage the ultra vires doctrine, which addresses what follows when an agent crosses that boundary and causes harm to third parties who had no opportunity to verify where it lay.[46]

The ultra vires doctrine has evolved from the strict charter doctrine of the nineteenth century toward a foreseeability-centered inquiry,[47] but the animating question has remained constant, concerning the circumstances under which a principal bears liability for agent conduct that transgresses granted authority and the point at which that transgression severs the juridical connection between principal and harm. The resolution, under modern doctrine, turns on the foreseeability of the deviation at the moment of delegation.[48,52] A principal who authorizes one category of action and discovers its agent performed conduct of a wholly different character may disclaim that performance as ultra vires, but the disclaimer provides no shelter against liability to third parties who lacked any means of verifying the scope of the agent's authority.[49]

A preliminary distinction clarifies the relationship between the frolic doctrine examined in Part II and the ultra vires inquiry developed here. The frolic doctrine asks whether the category of activity falls within the scope of authorized employment; a servant who abandons his employer's errand entirely to pursue personal business has stepped outside that scope, and the employer bears no liability for what follows. The ultra vires doctrine poses a distinct question, namely whether a particular act, taken within the authorized category, exceeded the precise bounds of the authority granted. An autonomous weapons system executing a targeting engagement remains within the

authorized category of activity and therefore presents no frolic; the ultra vires question is whether the specific classification decision exceeded the rules of engagement that bounded its authority. The soldier who shoots a protected person while on patrol remains within his assigned category of activity, doing exactly what he was commissioned to do, yet may nonetheless have acted beyond the precise limits of his authority. The AWS liability framework mirrors this structure, with the frolic inquiry addressing the category question and the ultra vires inquiry addressing the scope question. The two doctrines are complementary rather than contradictory because they operate at different levels of abstraction.[46]

Applied to autonomous weapons systems, the ultra vires inquiry resolves into two sequential questions, the first asking whether the harm-causing conduct fell within the scope of the authority delegated at deployment, and the second asking whether the capacity for the deviation that produced that conduct was foreseeable at the moment authority was granted. The second question carries the greatest consequence for the accountability of modern machine learning systems, and it is there that the inquiry achieves its most distinctive force. A rule-based autonomous system, one whose decision logic can in principle be fully enumerated in a specification document, presents a relatively tractable foreseeability question, asking whether the harmful conduct was contemplated by the rules as programmed. A machine learning system relates to this model altogether differently, since its operational envelope emerges from the interaction of training data, network composition, and the distribution of environmental inputs the training process represented, a distribution that cannot encompass the full range of adversarial conditions the system will encounter in deployment.[50] No specification document of any length or technical sophistication can fully characterize what an ML-based targeting system will do when it encounters an adversarial operational environment engineered to exploit the vulnerabilities of its classification logic in environments that military adversaries have every incentive to construct.[51] The state that deployed a machine learning system for autonomous targeting chose, with the awareness that carries legal consequence, to commission a system whose behavioral

envelope could not be fully specified in advance, gaining in exchange the tactical advantages that adaptive, input-sensitive systems offer over their rule-bound predecessors. That deliberate election forecloses the ultra vires defense by rendering the entire class of the system's emergent behavioral possibilities, including its capacity for unauthorized deviation, foreseeable as the ineluctable category of risk that attends the adoption of a technology inherently characterized by unpredictability in novel conditions.

The ultra vires defense remains available, then, but only in narrow circumstances, requiring that a system's deviation fall demonstrably outside any reasonable characterization of its behavioral envelope and that the deploying state had no basis, at the moment of authorization, for anticipating that the deviation was within the range of the system's performance capabilities.[52] The extensively recorded unpredictability of AI systems in novel adversarial environments, combined with the fact that state actors authorizing AWS deployment have access, as a matter of regulatory requirement and institutional practice, to testing data, red team assessments, and failure mode analyses characterizing the system's known behavioral limitations,[53] means that this standard will, in practice, rarely be satisfied. A principal that reads its own risk assessments before authorizing deployment has already foreclosed any subsequent claim of ignorance. The defense of unforeseeability cannot be sustained by a party whose own documentation establishes that the risk was, in fact, foreseen. This framework is not, it bears emphasis, without limiting principles. Not every AWS output is foreseeable in a legally operative sense; the foreseeability analysis is bounded by what the deploying state actually knew or had reason to know from its own documentation and testing regimes. A system whose conduct in a truly unprecedented environment falls outside any documented failure mode may present a case where the defense survives. The framework's claim is the more modest one, namely that the category of civilian misclassification in adversarial environments is, given the accumulated AI risk literature and mandatory pre-deployment testing requirements, almost never truly unforeseeable to a state that has satisfied its own regulatory obligations before deployment.

International humanitarian law and the corporate accountability framework developed in this Article arrive at the same structural obstacle from different directions, and that convergence warrants explicit statement before examining either in detail. IHL fails at the outset, as individual victims have no standing to bring civil claims against states in the forums where IHL operates.9 The ICJ adjudicates state-to-state disputes exclusively, the ICC cannot reach nationals of non-member states without Security Council referrals those states can block, and command responsibility under Articles 86 and 87 of Additional Protocol I requires proof of command knowledge that autonomous weapons systematically eliminate.96 The corporate framework encounters its own obstacle at the opposite end, supplying the doctrinal basis for civil liability but meeting sovereign immunity before reaching the state defendant. In both cases the victim is left without a forum and without recourse; the harm was wrongful, the doctrine exists to address it, and the defendant's governmental status alone forecloses relief. That parallel failure is a structural feature of the international legal order, endemic to its architecture rather than peculiar to either body of law. Its most important implication concerns the sovereign immunity obstacle, whose existence does not constitute an argument against the corporate tort framework, since IHL encounters the same barrier; the obstacle's persistence is the structural problem this Part undertakes to address.

VI. The Same Wall: *Sovereign Immunity, Its Parallel Failure Across Both Frameworks, and the Treaty Pathway*

The framework constructed across Parts I through V is rooted in domestic private law doctrine, in the accumulated jurisprudence of Anglo-American courts and in the Restatement's systematic codification of that jurisprudence. The objection this foundation invites is obvious in its formulation and serious in its implications, as states that deploy weapons systems are sovereign entities, and sovereign immunity interposes between them and the private law liability the framework would otherwise impose with a legal force that corporate actors cannot claim and that no corresponding doctrine neutralizes.[54] The objection is well-founded, though on examination it proves less categorically disqualifying than its initial presentation suggests, and the pathway that

existing international law instruments have constructed around its most obstructive features is considerably more navigable than the existing autonomous weapons accountability literature has had occasion to acknowledge.

In the United States, the Foreign Sovereign Immunities Act establishes as the governing default rule that foreign states are immune from the jurisdiction of American courts, subject to carefully delimited exceptions whose boundaries courts have construed with consistent fidelity to the statute's evident design.[55] The three exceptions most proximate to autonomous weapons accountability, the commercial activity exception, the tortious act exception, and the terrorism exception, each fail on their own terms. The commercial activity exception requires the gravamen of the claim to be grounded in activity private parties characteristically conduct; weapons deployment is a sovereign function, not a commercial one.[5657] The tortious act exception reaches only torts occurring on American soil, which AWS strikes on foreign civilians do not satisfy.[58] The terrorism exception applies only to states designated as terrorism sponsors through a politically driven statutory process entirely unrelated to AWS capability, and the states that actually develop advanced autonomous weapons are not on that list.[59] The FSIA, as written, immunizes foreign states from AWS-related civil litigation in American courts with a breadth that is structural rather than inadvertent. The purpose of this analysis is to establish the precise reason why the treaty pathway developed in Part VIII is legally necessary rather than merely aspirational; no domestic route around sovereign immunity exists under current FSIA doctrine for battlefield claims of this character, and IHL and corporate tort law arrive at the same barrier by different paths, a barrier that only a Protocol whose terms include a defined immunity waiver can dismantle.

At the international level, the argument that violations of peremptory norms (*jus cogens*) override sovereign immunity has been advanced, litigated, and, for now, rejected. In 2004, the Italian Supreme Court held in *Ferrini v. Federal Republic of Germany* that *jus cogens* violations strip a state of its immunity.[60] The International Court of Justice rejected that reasoning years later in 2012, holding by twelve votes to three that no *jus cogens* exception to sovereign immunity exists under customary international

law.[61] The ICJ's ruling has not ended the debate. Three judges did dissent.[62] Several European domestic courts continued to develop the *jus cogens* exception in the decade after *Jurisdictional Immunities*,[63] and scholars and international legal bodies have argued that the combination of systematic *jus cogens* violations and the absolute denial of civil remedy generates pressure toward a customary law obligation of access to justice.[64] The trajectory of the norm is still in motion, approaching but not yet at the point of recognizing that sovereign immunity cannot function as a permanent shield against accountability for the gravest violations of international law.

Whether the complete structural foreclosure of civil remedy for documented wrongful harm, denied solely because of the defendant's governmental status, should remain entirely beyond judicial examination is a question the existing scholarship has largely deferred. The prevailing assumption holds that immunity is a matter of legislative and executive creation: courts interpret existing waivers but cannot scrutinize the decision not to waive, and separation of powers concerns insulate foreign policy choices from judicial review. That assumption deserves more sustained scrutiny than it has received in the autonomous weapons context. Where a legal regime produces the systematic denial of any civil remedy for harm that would generate liability if caused by any private party, courts in constitutional orders committed to judicial remedy as a fundamental right may have independent grounds to examine whether that denial is constitutionally and internationally compelled, or merely customary. The European Court of Human Rights confronted the tension directly in *Al-Adsani v. United Kingdom* in 2001.[88] The Grand Chamber held that sovereign immunity generally prevailed over the applicant's civil torture claims, though the majority's acknowledgment that Article 6's right of access to a court was engaged, even if outweighed, established a principle of consequence: the complete denial of any judicial remedy raises human rights questions that immunity doctrine cannot simply foreclose without examination. Eight judges dissented on the ground that *jus cogens* violations should prevail over immunity, a position whose force has grown rather than diminished in the years since the decision. The Italian Constitutional Court drew the harder line in Judgment No. 238/2014,

declining to enforce the ICJ's *Jurisdictional Immunities* ruling on the ground that its domestic constitutional order independently requires access to a judicial remedy for victims of grave violations, a conclusion reached through constitutional rather than international law reasoning, and therefore not controlled by the ICJ's 2012 holding.[63] Neither court dismantled sovereign immunity. Both established that a constitutional commitment to judicial remedy is an independent constraint on the circumstances in which immunity can produce complete denial of access to courts. In the United States, the due process guarantee of the Fifth Amendment and the obligations accepted under the International Covenant on Civil and Political Rights, which requires effective remedies for rights violations recognized therein[89] furnish analogous grounds for examining whether blanket immunity applied to AWS casualties is constitutionally and internationally compelled, or merely unexamined. Courts adjudicating manufacturer claims arising from the same autonomous weapons incidents are not required to treat the parallel immunity of the deploying state as entirely beyond judicial consideration. The trajectory of constitutional and human rights law on the completeness of the immunity barrier is toward scrutiny, not deference, when the denial of remedy is total and the harm is grave.

The question of what international law can actually deliver as remedy, as distinct from what it articulates as aspiration, was given its most durable formulation four centuries ago: covenants without the sword are words and nothing more.[65] Nothing in the subsequent development of international legal institutions has fundamentally altered that assessment where the parties in question are the permanent members of the Security Council, the states with the greatest stakes and investments in autonomous weapons development and the most reliable means of blocking any enforcement mechanism directed at themselves.

Hobbes' observation bears on the international law framework governing autonomous weapons accountability with a precision that the years since 1651 have done nothing to diminish. The International Court of Justice holds the authority to adjudicate disputes between states that have consented to its jurisdiction, though it

possesses no enforcement mechanism of its own, relying as it is upon referral to a Security Council whose five permanent members hold individual veto authority over any coercive response.[6696] The International Criminal Court commands jurisdiction over individual war crimes perpetrators with a theoretical amplitude that encompasses the command responsibility of those who deploy AWS, though the United States, Russia, and China are not parties to the Rome Statute, and jurisdiction over their nationals requires a Security Council referral that each of them can individually block.[67] The Security Council, when animated by unanimity, possesses the authority to impose sanctions, create compensation mechanisms, and authorize enforcement measures of considerable force, and unanimity requires the concurrence of the five permanent members, and the states most likely to deploy autonomous weapons in circumstances generating accountability questions are the five permanent members.[68] International law's primary contribution in the AWS context is therefore normative rather than remedial, shaping state conduct through reputational pressure and the gradual accretion of customary international law,[69] rather than through tribunals delivering enforceable verdicts against powerful states.

This assessment is offered as candor rather than resignation on the matter. It describes the present architecture of international law not to foreclose the possibility of accountability but to define with precision what a viable accountability pathway must be designed to circumnavigate. Three existing international instruments furnish structural precedents for such a pathway, each demonstrating through its actual operation that the design problem is tractable.

Two of the three instruments address the same core problem from complementary angles, each exploring how to generate civil remedies for sovereign wrongdoing without requiring one state to sue another. The Convention Against Torture established that states can commit by treaty to providing such remedies within their own legal systems; Article 14 imposes the affirmative obligation to "ensure in its legal system that the victim of torture obtains redress and has an enforceable right to fair and adequate compensation."[70] An AWS accountability instrument modeled on Article 14

would require signatory states to create domestic civil remedies for autonomous weapons casualties, generating accountability through national courts exercising jurisdiction over their own states and sidestepping the FSIA problem without purporting to resolve it. The International Oil Pollution Compensation Fund proceeds differently: the underlying Civil Liability Convention imposes strict liability on shipowners, while the Fund provides supplemental compensation financed by cargo interests when CLC limits are exceeded, giving individual claimants direct access to compensation through an administrative mechanism that bypasses sovereign immunity entirely.[71] Together, these two instruments establish the foundational vocabulary for an AWS accountability protocol, where they combine state-level civil remedy obligations on one side with a pooled fund that converts sovereign immunity from a jurisdictional obstacle into a contribution question on the other.

A third precedent is the UN Compensation Commission, created after Iraq's invasion of Kuwait and established as an instrument of reparative justice under Security Council authorization.[72] The UNCC processed nearly 2.7 million submitted claims across the full span of its operations and awarded approximately $52.4 billion in compensation to some 1.5 million successful claimants, a figure that demonstrates the institutional machinery for mass individual claims administration works at scale..[73] The political precondition that brought the UNCC into existence, Security Council unanimity, renders the model unavailable for accountability proceedings involving P5 states under present geopolitical conditions. Its operational design is sound, its track record documented, and nothing in how it functioned depends upon the Security Council authorization that is the source of its current inapplicability; the Commission's procedural framework can be transposed into a treaty-based mechanism of comparable scope that derives its authority from the consent of state parties rather than from the Security Council's coercive mandate.

The convergent pathway these three models delineate points toward a Protocol whose structural design questions each of these instruments has, in analogous form, already answered. The essential innovation is a defined and consensual immunity

waiver, the mechanism by which the wall that both IHL and corporate tort law encounter is, for signatory states, converted from an absolute barrier into a contained concession. States limit their own immunity by consent with notable regularity in bilateral investment treaties, commercial arbitration agreements, and specialized international conventions;[74] the concession a Protocol would ask of signatory states is structurally identical in kind, distinguished only by its subject matter. A Protocol incorporating a narrowly circumscribed waiver for claims channeled through its designated commission would furnish what current international law conspicuously withholds: a tribunal with actual jurisdiction, a liability standard untethered from the intent requirements that autonomous weapons render permanently unavailable, and a genuine avenue of individual access to remedy. The waiver would be confined to autonomous weapons deployments, operative only before the Protocol's Commission, leaving intact a signatory state's immunity in all other proceedings and making accountability the price of admission to that category of warfare.

VII. The Accountability Theater: Incentive Structures, Regulatory Capture, and the Commercial Pressure That Changes Them

The framework developed across Parts I through VI requires no doctrinal invention. Agency law is among the most ancient and pervasively taught bodies of doctrine in the Anglo-American legal tradition, its foundations predating the republic whose legal system inherited them.[75] Strict liability for abnormally dangerous activities has occupied its place in the Restatement since 1977, applied to blasting operations and chemical storage and nuclear facilities without generating the existential doctrinal controversies that its AWS application is apparently expected to produce.[40] The question of why this well-developed framework has been withheld from the autonomous weapons accountability answerability discourse is worth examining; though this Article treats it as secondary to the affirmative argument, it illuminates the political economy of international legal norm formation with a clarity that the affirmative argument alone cannot provide.

The CCW process stands as the primary exhibit for the prosecution: more than a decade of annual deliberation, no binding instrument of any description, and a set of "guiding principles" whose drafters took the precaution of stating explicitly that they operated "without prejudice to the result of future discussions," a formulation that evacuates whatever normative content the principles might otherwise have claimed.[76] The EU AI Act, celebrated as the world's first comprehensive AI governance instrument, carves out national security applications so broadly that the carve-out effectively swallows the rule for virtually all military autonomous systems.[77] DoD Directive 3000.09 demands "appropriate levels of human judgment over the use of force" without specifying what renders judgment appropriate rather than nominal, what threshold of human engagement the standard requires, or what retrospective mechanism would permit a court or investigator to determine whether the standard had been satisfied in a given deployment.[8] Surveyed together, these instruments share a common architecture, offering the rhetoric of accountability stripped of its operative mechanism, language that a plaintiff cannot present to a court and a court cannot apply to a defendant's conduct. Holmes articulated in 1897 the criterion for distinguishing legal instruments from their simulacra: "the prophecies of what the courts will do in fact, and nothing more pretentious, are what I mean by the law."[78] Applied to this body of soft-law instruments, the test resolves the question cleanly: a plaintiff cannot present DoD Directive 3000.09 to any court and obtain relief, and the CCW's guiding principles support no cognizable claim in any tribunal with enforcement authority. Under Holmes's definition, these instruments are the performance of law rather than its substance.

Sparrow argued that the responsibility gap constitutes a fundamental moral failure, not merely a practical deficiency, since any just war requires that someone bear responsibility for deaths that occur in its prosecution, a condition that autonomous weapons deployment cannot satisfy. [79] The institutional response to that argument, across the decade and a half since the CCW process commenced its deliberations, has been to produce frameworks that engage Sparrow's moral charge at the level of rhetoric rather than law, instruments that acknowledge the problem's gravity while declining to

create the mechanisms that would address it. The pattern is, upon reflection, precisely what the incentive structure of the relevant institutions would lead a dispassionate observer to anticipate.80 The states most energetically engaged in shaping the CCW process are the same states most materially invested in preserving unencumbered operational freedom in autonomous weapons deployment.[80] The five permanent members of the Security Council include the world's three leading autonomous weapons developers alongside two other states with significant and advancing programs, and every one of them commands individual veto authority over Security Council action of any legally binding consequence,[8168] while simultaneously occupying seats at the CCW's deliberative table. The soft-law outcomes that process has generated require no elaborate explanation; they are the predictable product of a regulatory negotiation whose most consequential participants are simultaneously the parties against whom the resulting norms would operate.[82] The observation is structural, and it describes how international norm formation works when the regulated parties possess both the incentive and the institutional authority to shape the regulatory outcome in their favor. Those institutions will not close the distance between the law as it reads and the law as it functions; that closing, if it comes, will come through courts applying doctrine in cases where doctrine reaches and through plaintiffs suing manufacturers and developers in domestic proceedings where sovereign immunity does not apply, through mechanisms designed with enough structural specificity to navigate the sovereign immunity problem rather than sufficient diplomatic tact to avoid it.

The mechanism by which sustained civil litigation against private defendants creates conditions for sovereign ratification is not speculative. American legal and regulatory history has produced it, with documented specificity, across multiple industries that faced sustained civil liability exposure for product-related harm. Johns-Manville Corporation filed for bankruptcy protection in 1982 under the accumulated weight of asbestos litigation, at that moment the largest corporate bankruptcy in American history, after internal documents established that its executives had known for decades that asbestos exposure caused mesothelioma and had suppressed that

knowledge. The bankruptcy did not close the litigation. Claims against successor entities and co-defendants accelerated through the 1980s and 1990s, consuming more than $70 billion in corporate assets across the industry and driving over sixty companies into insolvency. The industry's strategic response was active lobbying for the FAIR Act, a proposed federal compensation trust fund that would have removed all asbestos claims from the tort system and replaced them with a defined administrative scheme.[90] The industry that had opposed any accountability frameworks for decades sought one, because a defined and capped liability arrangement is commercially preferable to unlimited catastrophic exposure. The tobacco industry followed the same arc on a longer timeline. State attorneys general litigation in the 1990s, which produced the 1998 Master Settlement Agreement committing the four major manufacturers to an estimated $206 billion in payments over twenty-five years, had been resisted for years on the grounds that tort liability for the consequences of a legal product was legally unprecedented.[91] Within a decade of the settlement, those same manufacturers were engaged with federal regulators on the Family Smoking Prevention and Tobacco Control Act of 2009, which gave the FDA comprehensive authority over tobacco products. Sustained litigation had converted opponents of federal oversight into architects of its design. The opioid litigation produced the same conversion at greater speed: manufacturer settlements exceeding $50 billion, an attempted immunity-purchase arrangement by the Sackler family that the Supreme Court rejected in *Harrington v. Purdue Pharma*, 603 U.S. 204 (2024), on the grounds that it improperly shielded non-debtor third parties from civil liability, and regulatory reforms incorporated into the SUPPORT for Patients and Communities Act of 2018, reforms those same companies had previously opposed.[92] In each instance the operative mechanism was identical: sustained civil exposure in courts that could reach the private defendant, unconstrained by any immunity to assert, created the commercial incentive to prefer a defined accountability regime over unlimited litigation risk. The autonomous weapons manufacturer occupies the identical structural position. It builds a system. The system causes harm through processes the manufacturer designed, trained, and deployed with

documented knowledge of their risk characteristics. The corporate liability chain reaches the manufacturer under the retained control and inherently dangerous activity exceptions, and the *Boyle* defense fails for the reasons Part III establishes. When courts apply the corporate liability chain consistently and at scale, the commercially rational response will mirror the response Johns-Manville, Altria, and Purdue Pharma eventually reached: active support for a defined accountability regime. The AWS Accountability Protocol whose elements this Article identifies in Part VIII is calibrated to be precisely that regime, the one a commercially rational manufacturer under sustained litigation pressure would prefer to open-ended catastrophic exposure. Private defendant litigation compensates victims in the near term and builds the political constituency that makes a sovereign accountability Protocol ratifiable.

A qualification on the mechanism is warranted. In each of the historical cases cited, the sovereign, meaning the states and the federal government, was on the plaintiff side, not the defendant side. Manufacturer defendants faced open-ended civil exposure and sought regulatory resolution. The AWS context differs in one structural aspect in that the deploying state is both a sovereign that could ratify a Protocol and the party whose immunity the Protocol would waive. History does not yet offer an example of litigation pressure on private defendants causing powerful sovereign states to accept treaty-based civil liability for their own military conduct. That gap in the evidence base is real. This Article's claim is more limited and more defensible, namely that sustained domestic litigation against manufacturers under the framework developed in Parts II through V creates commercial incentives among the private-sector participants in the AWS enterprise that, over time, shift the political economy of treaty ratification for the sovereign accountability mechanism. Manufacturers who face open-ended liability without a defined Protocol to cap it become rational advocates for the regulatory clarity the Protocol provides. The historical analogy holds at the level of private-defendant incentive structure, not at the level of sovereign liability design. The sovereign accountability component remains aspirational pending the political will that Part VII identifies as the constraint; what the domestic litigation mechanism demonstrates is

that the commercial constituency for the Protocol's ratification exists and is capable of being mobilized.

VIII. The Instrument: Proposed Provisions for an AWS Accountability Protocol

Legal scholarship that identifies a doctrinal deficiency without proposing what would remedy it has completed only half its proper undertaking, the diagnostic half, which however accomplished remains insufficient to discharge the discipline's obligation to constructive analysis.[83] What the framework developed in this Article requires at the international level is a Protocol, not a first draft of one but a clear identification of the structural elements such a Protocol would need to incorporate and a demonstration that each is achievable under existing international law precedent. The design questions are answerable, and the analogous instruments are already in operation. This Part addresses what a functional AWS Accountability Protocol must contain, how each element functions, and where the difficult cases arise, a blueprint whose architecture is already validated, even if its construction awaits political will.[84]

i. Liability Standard

The liability provision is the Protocol's doctrinal foundation, and its architecture follows directly from the principal-agent analysis developed in this Article. A state that deploys an autonomous weapons system is strictly liable for harm caused to civilian persons when that system operates within the scope of its authorized mandate, with liability attaching upon proof of deployment and proof that the harm-causing conduct fell within those bounds, with no showing of intent, knowledge, or negligence required. Strict liability is the only standard adequate to the AWS context; any fault-based threshold restores the knowledge and intent requirements that autonomous weapons systematically eliminate.45 The liability chain extends jointly and severally through the development and deployment sequence (prime contractor, AI developer, component suppliers) with contribution proceedings among defendants allocating responsibility according to each party's degree of control over the system's behavior. The covered system's definition is the provision's most demanding drafting question; any Protocol

would need to establish the threshold of autonomous operation that brings a system within its scope, almost certainly keyed to the degree of independence from real-time human targeting decisions rather than to technical architecture, which varies and can be manipulated by definitional framing. A state that characterized its system as merely semi-autonomous to avoid Protocol coverage would bear the burden of demonstrating that human judgment operated at the moment of target selection in a manner that meaningfully distinguished the deployment from covered autonomous operation. Two definitional gaps compound the coverage question. The term 'civilian persons' requires elaboration since civilian status in armed conflict is contested enough to encompass questions of hors de combat status, persons directly participating in hostilities whose combatant classification is contested, and the treatment of persons in genuinely ambiguous circumstances, rendering an undefined term vulnerable to litigation in virtually every significant claim. The Protocol's formulation should adopt the IHL threshold as a floor while specifying that the Commission applies it independently of any IHL adjudication. The causation requirement presents a distinct challenge in environments where multiple forces deploy lethal systems simultaneously; a claimant in a contested urban environment must establish that the autonomous system, and not conventional weapons or other actors, caused the harm. A Protocol adequate to this challenge would establish a preponderance-of-the-evidence burden with allocation allocation rules accounting for the information asymmetry between claimant and deploying state (the state has access to system behavioral logs and deployment records that the claimant cannot independently obtain, which is exactly why the Commission's evidentiary compulsion authority is essential to the claim process rather than incidental to it.

ii. Individual Standing and Access to Remedy

Individual access to remedy is among the Protocol's most demanding design requirements, and the provisions must engage honestly with the conditions under which claims will actually arise. The population of potential claimants after a mass casualty autonomous weapons strike includes people in active conflict zones, without legal

representation, without documentation establishing their identities or the harm they suffered, and often without knowledge that any accountability mechanism exists. Individual filing requirements drawn from commercial arbitration practice would defeat most claims before they reached any substantive review. The UNCC addressed this directly through category-based claim processing, consolidating similarly situated claimants, then adopting standardized evidentiary requirements for each category and allowing representative filings where individual claimants could not practicably appear. A Protocol adequate to the AWS context would need a comparable collective mechanism, with expedited processing for incidents above a defined casualty threshold, standing for recognized human rights organizations to file representative claims on behalf of identified victim populations, and provisions for state parties to submit claims on behalf of their nationals where individual filing is impracticable.[73] The limitation period, meaning the time within which claims must be filed after a qualifying incident, presents a secondary but significant drafting question. Conflict environments frequently make immediate filing impossible, and evidence of system behavior and deployment parameters may be destroyed if not preserved quickly. A Protocol should include both a tolling provision where claimants lacked reasonable access to filing mechanisms and an interim measures authority allowing the Commission to order preservation of operational logs and system behavior data before merits review begins.

iii. Sovereign Immunity Waiver

The sovereign immunity waiver is the Protocol's architecturally decisive provision, the mechanism by which the wall that both IHL and corporate tort law encounter is, for signatory states, converted from an absolute barrier into a defined and consensual carve-out. The precedent is well-established in that states waive immunity through treaty in bilateral investment agreements and commercial arbitration frameworks routinely, accepting the jurisdiction of specified tribunals for specified categories of dispute while preserving immunity everywhere else. A Protocol waiver would follow the same structure, limited to claims arising from autonomous weapons deployments, operative only before the Protocol's Claims Commission, and explicitly preserving

immunity in domestic court proceedings of other states. The narrowness of the waiver is its political virtue: a state signing the Protocol does not expose itself to general civil jurisdiction; it accepts a precisely delimited accountability obligation in exchange for the legitimacy of participating in the regime that governs the category of warfare it has chosen to enter. The waiver's limitation to Commission proceedings also means that private defendant litigation under domestic law proceeds independently, without requiring nor foreclosing Protocol claims against the deploying state, though the Protocol should specify the preclusion rules that apply where a claimant has obtained a domestic judgment against a private defendant arising from the same incident, to prevent double recovery while preserving access to Commission proceedings for claims the domestic judgment did not fully resolve. The waiver provision must additionally address two structural contingencies, being, a withdrawal clause specifying that a state party's withdrawal from the Protocol does not affect jurisdiction over claims arising during membership, with a savings clause preserving pending proceedings; and a non-derogation provision barring state parties from contracting around Protocol obligations through bilateral agreements with manufacturers or developers.[74]

iv. Claims Commission Structure.

The Claims Commission requires three elements that a Protocol's establishment cannot supply unaided, though existing practice illuminates each. Its jurisdictional foundation would draw on the Protocol's liability provision and the ILC Articles on State Responsibility's attribution rules, with the principal-agent framework of this Article informing the interpretation of both.84 Decisions would be final and binding on state parties and enforceable in their domestic courts as if they were domestic judgments, an enforcement mechanism that the ICSID Convention already employs successfully for investment arbitration awards. The Commission's evidentiary authority presents the more demanding structural question: it must have compulsory power to require production of deployment contracts, operational parameter specifications, system testing records, and behavioral logs. Without documentary compulsion, the evidentiary foundation that makes the respondeat superior analysis workable, namely the audit of

delegation rather than the excavation of mental states, is practically inaccessible to claimants. A criminal referral provision completes the design where Commission proceedings produce evidence of deliberate targeting of civilians, the Commission should have authority to refer that evidence to the ICC or to competent domestic prosecutors.67 The Protocol is a civil accountability instrument, but its proceedings should not operate as a shield against criminal liability where the evidence warrants it.[85]

v. Compensation Fund.

The compensation fund resolves the sovereign immunity obstacle for claimants who cannot practicably pursue Commission proceedings against a non-complying state party, and it creates the financial mechanism that makes the Protocol's incentive architecture coherent. States contribute annually in proportion to the scale and frequency of their autonomous weapons deployments, as reported under the Protocol's transparency obligation. Where a responsible state party fails to satisfy a Commission award within a specified period, the Fund satisfies the award and pursues reimbursement from the state party through inter-state proceedings. The IOPCF operates on exactly this model, and its operational track record of over 150 incidents and some £752 million in compensation paid establishes that the mechanism functions at scale. [71] The contribution calculus is the fund's most significant political question, because the formula determines the relationship between deployment scale and financial obligation. A formula calibrated only to incident frequency would reward states that deploy extensively but carefully; a formula calibrated to deployment scale creates incentives to underreport the scope of authorized operations.86 The IOPCF's model, contributions proportional to quantities of covered substance received in member states, provides the governing logic, in that contribution is tied to the activity that generates the risk, not to the incidents that materialize from it.

vi. Transparency and Verification.

The provisions outlined above address what a functional Protocol must contain, and each finds its precedent in the operational experience of international instruments that have resolved comparable problems in analogous contexts. The liability provision

derives from the doctrinal analysis of Parts I through V. The claims mechanism draws on the UNCC's demonstrated capacity for mass individual claim administration. The immunity waiver is grounded in the established practice of consensual immunity limitation in investment and arbitration treaties. The Commission's evidentiary authority follows from the documentary audit structure that makes respondeat superior viable. The fund mechanism is modeled on the IOPCF's operational design. The transparency and verification obligations complete the architecture, and one of their functions is structural being that state parties would be required to report the scale and frequency of their autonomous weapons deployments to a designated body on a defined schedule, and it is precisely this reporting obligation that generates the deployment data on which the contribution formula in section v is calculated. A formula tied to activity that generates risk rather than incidents that materialize from it requires a reliable record of that activity, and the transparency provision is what produces that record. The broader drafting questions that transparency and verification obligations present, including verification mechanisms, inspection rights, and the treatment of classified operational data, are ones that existing arms control practice has addressed in comparable contexts, though a thorough treatment of that precedent exceeds the scope of this Article. None of these elements requires doctrinal invention. The Protocol whose construction they describe awaits only the political decision to build it.[85]

Conclusion

What pervades autonomous weapons law is a gap, and a gap of a precise and consequential kind. The instruments needed to hold deploying states and manufacturers answerable for harm caused by autonomous lethal systems have existed for centuries, embedded in legal traditions older than the states that now deploy those systems. Respondeat superior,[22] the principal-agent framework,[13] the retained control and inherently dangerous activity exceptions,[34] and strict liability for abnormally dangerous activities[40] together constitute an accountability edifice of considerable reach and doctrinal coherence, one that operates where IHL cannot, that requires no proof of intent, knowledge, or any other mental state element that AWS deployment

renders unavailable, and that has governed non-human actors causing harm on behalf of identifiable principals since long before the first autonomous weapons program entered development.

The deficit is one of application, compounded by institutional will and the demonstrable novelty of the territory. The principal-agent framework has been withheld from autonomous weapons accountability not solely because powerful states prefer the gap, though they undeniably do, but also because the analogy requires courts to take a step no court has yet taken and because legitimate institutional resistance to applying private law frameworks to battlefield sovereign conduct is a real constraint on the argument's practical force, not a rhetorical obstacle to be dismissed. This Article's claim is more limited and more defensible than the claim that courts will inevitably adopt the proposed framework; it holds that the framework is doctrinally coherent, that the analogical extensions it requires are grounded in verifiable structural equivalences rather than mere resemblance, and that the accountability gap it would close is more damaging to the legal order than the doctrinal extension it requires. That claim is contestable. It is also, this Article argues, correct. The soft-law apparatus that has proliferated in the accountability vacuum reflects not merely the political economy of international legal norm formation but also the demonstrable difficulty of the problem, a difficulty that honest engagement with the framework's limits does not diminish but illuminates.[76]

Sovereign immunity, examined with rigor rather than reverence, proves considerably narrower in its actual disqualifying force than its invocation in the literature suggests.[54] States possess the sovereign authority to circumscribe their own immunity by consent, and exercise that authority with routine regularity in the commercial and investment contexts where the economic advantages of consenting to dispute resolution outweigh the costs of accepting a defined category of potential liability.[74] An AWS Accountability Protocol that synthesizes the Convention Against Torture's civil remedy obligation,[70] the International Oil Pollution Compensation Fund's pooled compensation mechanism,[71] and a specific, cabined immunity waiver for proceedings

before the Claims Commission, together asking of states a modest concession, namely submitting, within a defined and limited compass, to the jurisdiction of a tribunal constituted by their own consent. That concession is common practice in commercial arbitration; extending it to autonomous weapons' accountability is a political decision dressed as a legal impossibility.

The framework operates in sequence across two channels that are causally connected rather than merely parallel. The first is domestic: the principal-agent framework reaches manufacturers and developers under existing law, without sovereign immunity, in courts that function today. That litigation produces civil exposure that is real, sustained, and as the asbestos, tobacco, and opioid histories each demonstrate, commercially transformative. The second is international: that exposure creates the incentive structure that makes a Protocol ratifiable, because manufacturers facing open-ended domestic liability become rational advocates for the defined accountability regime the Protocol provides, exactly as industries in those earlier cases became advocates for the regulatory settlements they had previously opposed. The domestic channel compensates victims in the near term and builds the political constituency that makes the international track conceivable; the international track is the only mechanism that reaches the deploying state and closes the immunity barrier the domestic track cannot cross. Neither track alone closes the accountability gap. Together, the sequence does: domestic litigation makes the Protocol commercially necessary to the industry that can advocate for its ratification, and the Protocol makes the deploying state answerable to the victims the domestic track cannot reach. That sequence is the argument this Article advances, and the doctrine at each step already exists.

What the adoption of this framework would generate is a reorientation in the questions courts are asked to answer, shifting from the permanently contested terrain of moral philosophy into the more tractable domain of doctrinal analysis. With what degree of specificity must a deploying state define an autonomous weapons system's authorized conduct to establish the outer boundary of its delegated authority, [48] At what

threshold of documented behavioral unpredictability does a machine learn and at what threshold of documented behavioral unpredictability does a machine learning system's deviation from anticipated conduct become sufficiently foreseeable to extinguish an ultra vires defense at the moment of deployment authorization? [52] These are questions that the legal discipline knows how to address, because they are variations, beneath the novel technological substrate, of questions courts have been resolving for two centuries in cases involving the full range of non-human actors who cause harm while operating within the scope of another's authorization.

Whether states will prove willing to acknowledge that they are subject to the same logic is a political question, beyond legal scholarship to resolve, though legal scholarship can at least ensure the doctrine is ready when the political will arrives. The causal ligament that autonomous weapons dissolve, the law already knows how to restore. The doctrine awaits them, whenever they are ready.

Notes

1. Oliver Wendell Holmes Jr., The Common Law 1 (1881). Holmes's aphorism opens his treatise on the development of Anglo-American common law through practical experience. His prediction theory of law appears in The Path of the Law, 10 Harv. L. Rev. 457, 461 (1897), applied in Part VII of this Article to assess the soft-law frameworks governing autonomous weapons. This Article employs Holmes's analytical framework - that law develops through experience rather than logical derivation, and that legal rules should be tested by their practical consequences - without endorsing his moral skepticism. For the most sustained critique of Holmes's jurisprudence as a system that stripped law of moral content, see Albert W. Alschuler, Law Without Values: The Life, Work, and Legacy of Justice Holmes (2000), reviewed by Phillip E. Johnson, Law Without Values: The Life, Work, and Legacy of Justice Holmes, First Things (June 2001), https://firstthings.com/law-without-values-the-life-work-and-legacy-of-justice-holmes/ (arguing that Holmes's empiricism, taken to its logical conclusion, resolves into a social Darwinism that identifies law only with power and has no resources for distinguishing better from worse legal rules). The AWS accountability gap this Article addresses is precisely the kind of practical failure - a rule whose operation produces a result the legal system cannot justify to its own beneficiaries - that Holmes's experience-based method would treat as the signal for doctrinal development, whatever one concludes about the deeper normative foundations of that method.

2. Autonomous weapons systems are variously defined as weapons that can select and engage targets without meaningful human intervention. See Human Rights Watch & Int'l Human Rights Clinic, Losing Humanity: The Case Against Killer Robots 2-4 (2012) (providing a foundational taxonomy distinguishing fully autonomous, semi-autonomous, and human-supervised systems). The definitional ambiguity is legally significant: accountability mechanisms can be evaded by classifying a system as "semi-autonomous." This Article argues the principal-agent framework applies across the spectrum because the liability analysis turns on the scope of delegation, not the degree of machine independence.

3. The IHL and AWS literature includes Robert Sparrow, Killer Robots, 24 J. Applied Phil. 62 (2007); Kenneth Payne, I, Warbot: The Dawn of Artificially Intelligent Conflict (2021); Human Rights Watch, Stopping Killer Robots: Country Positions on Banning Fully Autonomous Weapons and Retaining Human Control (2020); ICRC, Autonomous

Weapon Systems: Implications of Increasing Autonomy in the Critical Functions of Weapons (2016). This Article departs from that literature's near-universal reliance on IHL as the primary analytical framework.

4. The hypothetical is constructed but its operative facts are drawn from documented incidents. See United Nations Panel of Experts on Libya, Final Report of the Panel of Experts Established Pursuant to Resolution 1973, S/2021/229, para. 63 (2021) (documenting what appears to be the first instance of an autonomous weapons system programmed to attack targets without requiring data connectivity, involving a Kargu-2 rotary-wing attack drone).

5. Protocol Additional to the Geneva Conventions of 12 August 1949, and Relating to the Protection of Victims of International Armed Conflicts arts. 86-87, June 8, 1977, 1125 U.N.T.S. 3 [hereinafter API]. Article 86(2) provides that a superior is responsible for a violation committed by a subordinate if they “knew, or had information which should have enabled him to conclude in the circumstances at the time" that a breach was occurring. Article 87 requires commanders to prevent violations and initiate disciplinary or penal action. The doctrinal scope and limitations of command responsibility have been actively litigated in international criminal tribunals. See Prosecutor v. Hadžihasanović, Alagić & Kubura, Case No. IT-01-47-AR72, Decision on Interlocutory Appeal Challenging Jurisdiction in Relation to Command Responsibility (ICTY App. Ch., July 16, 2003) (addressing two fundamental limitations on the command responsibility doctrine: (1) its application to superiors in non-international armed conflict, and (2) whether a superior bears responsibility for acts committed by subordinates before he assumed command, with the Appeals Chamber confirming that the knowledge element runs to each specific violation, not merely to the general character of a subordinate's conduct). The Hadžihasanović decision illustrates that the knowledge requirement's conditions precedent are themselves contested and multiply qualified - conditions that autonomous targeting eliminates entirely by removing the human decision-maker from the precise moment the knowledge element would otherwise apply.

6. See Jean-Marie Henckaerts & Louise Doswald-Beck, Customary International Humanitarian Law vol. 1, Rule 153, at 558-63 (2005) Analysis of Rule 153 ("Commanders and other superiors are criminally responsible for war crimes committed by their subordinates if they knew, or had reason to know, that the subordinates were about to commit or were committing such crimes and did not take

all necessary and reasonable measures in their power to prevent their commission, or if such crimes had been committed, to punish the persons responsible."). The Henckaerts & Doswald-Beck commentary.

7. ICRC, Autonomous Weapon Systems: Implications of Increasing Autonomy in the Critical Functions of Weapons (Geneva: ICRC, 2016), 10-11, 16-21 (discussing the technical characteristics of autonomous weapon systems, including the role of speed in targeting decision cycles, and the implications for human control and IHL compliance).

8. U.S. Dep't of Defense, Directive 3000.09: Autonomy in Weapons Systems (Nov. 21, 2012, updated Jan. 25, 2023). The Directive requires "appropriate levels of human judgment over the use of force" without specifying what "appropriate" means operationally, what level of human engagement satisfies the standard, or how compliance would be measured or enforced after the fact. For the official announcement of the January 2023 update and description of its revisions, see U.S. Dep't of Defense, Press Release, DoD Announces Update to DoD Directive 3000.09, "Autonomy in Weapon Systems" (Jan. 25, 2023), https://www.defense.gov/News/Releases/Release/Article/3278076/dod-announces-update-to-dod-directive-300009-autonomy-in-weapon-systems/ (describing policy revisions to address the rapidly evolving landscape of AI-enabled autonomous systems while retaining the “appropriate levels of human judgment standard” without defining operational thresholds).

9. See, e.g., ICRC, Autonomous Weapon Systems: Implications of Increasing Autonomy in the Critical Functions of Weapons 33-35 (2016); Human Rights Watch & Int'l Human Rights Clinic, Mind the Gap: The Lack of Accountability for Killer Robots 6-9 (2015); Special Rapporteur on Extrajudicial, Summary or Arbitrary Executions, Report on Lethal Autonomous Robotics and the Protection of Life, U.N. Doc. A/HRC/23/47 (Apr. 9, 2013) (Christof Heyns).

10. See generally Martti Koskenniemi, Between Apology and Utopia: The Politics of International Law, in Martti Koskenniemi, The Politics of International Law 35, 38, 42 (Hart Publishing 2011) (analyzing how international legal discourse produces instruments that appear binding while systematically avoiding substantive obligation, and that where law tracks only the effective interests of states it becomes "an apology for the interests of the powerful"); see also Anthea Roberts, Is International Law

International? 231–253 (2017) (documenting the tendency of international legal instruments produced by powerful state coalitions to preserve those states' freedom of action).

11. Robert Sparrow, Killer Robots, 24 J. Applied Phil. (2007). This Article agrees with Sparrow's diagnosis and departs from his prescription: Sparrow argues for prohibition; this Article argues that the principal-agent liability framework provides a workable accountability mechanism without requiring it.

12. Ryan Calo, Robotics and the Lessons of Cyberlaw, 103 Calif. L. Rev. 513-63 (2015). (arguing that robotics has a distinct set of essential qualities, embodiment, emergence, and social valence, that will generate legal challenges categorically different from those posed by the Internet, and that law must develop a new theoretical and doctrinal infrastructure to address them; the emergence quality, under which robotic systems accomplish tasks in ways that cannot be anticipated in advance, is the dimension most relevant to the causal attribution problem this Article addresses). His analysis focuses on domestic deployment of commercial robotic systems and does not engage with state actors, sovereign immunity, international humanitarian law, or the treaty-based accountability mechanisms that autonomous weapons contexts require.). Jack M. Balkin, The Path of Robotics Law, 6 Calif. L. Rev. Circuit 45-60 (2015) (arguing against essentialist approaches to robotics law and identifying two central challenges: how to allocate rights and duties among human beings when robots and AI entities cause harm or create benefits, and how to address the "substitution effect" - the contextual, unstable practice of substituting robots for humans across a range of functions). Balkin's framework operates at a higher level of abstraction than the principal-agent liability architecture this Article develops, and does not engage with the sovereign immunity, international humanitarian law, or state-actor dimensions that autonomous weapons deployment requires.). Woodrow Barfield & Ugo Pagallo eds., Research Handbook on the Law of Artificial Intelligence (2018) (collecting twenty-five contributions addressing the legal challenges posed by increasingly autonomous AI systems across constitutional law, intellectual property, private law, criminal law, and corporate law, and reflecting the broad scholarly consensus that existing legal frameworks designed for human actors require fundamental reorientation to address systems that operate with a degree of autonomy that makes conventional attribution of intent, fault, and causation problematic). The volume illustrates both the breadth of the doctrinal challenge and the absence of a

worked architecture specifically addressing state deployment of lethal autonomous systems - the sovereign immunity problem, the manufacturer liability chain, and the treaty-based accountability pathway that autonomous weapons contexts require remain outside its frame.

13. Restatement (Third) of Agency § 1.01 (Am. Law Inst. 2006). The Restatement defines the agency relationship as arising when "one person (a 'principal') manifests assent to another person (an 'agent') to act on the principal's behalf and subject to the principal's control, and the agent manifests assent or otherwise consents so to act."

14. Id. § 1.01 cmt. c (identifying three definitional elements of an agency relationship: mutual consent between principal and agent that the agent will act on the principal's behalf; the agent's subjection to the principal's control; and the agent's power to affect the legal rights and duties of the principal).

15. U.S. Dep't of Defense, Instruction 5000.02: Operation of the Adaptive Acquisition Framework (Jan. 23, 2020) (establishing that acquisition programs for defense systems require specification of operational requirements, performance parameters, and test and evaluation criteria, all of which constitute the principal's manifestation of what the system is authorized to do).

16. See Restatement (Second) of Agency § 1(1) (Am. Law Inst. 1958) ("Agency is the fiduciary relation which results from the manifestation of consent by one person to another that the other shall act on his behalf and subject to his control, and consent by the other so to act."). The Second Restatement's definition, like the Third's, is silent on the requirement of natural personhood in the agent. Corporate agents have been recognized since the earliest development of agency doctrine.

17. See Dodge v. Ford Motor Co., 170 N.W. 668 (Mich. 1919) (recognizing that corporations act through agents whose conduct binds the corporate principal). The absence of any natural-personhood requirement for either principal or agent in the agency relationship is implicit in the entire structure of corporate law: a corporation is a juridical person that acts exclusively through human and institutional agents. See Restatement (Third) of Agency § 1.01 cmt. e (Am. Law Inst. 2006) (confirming that the agent need not be a natural person and that organizational actors routinely occupy the agent position).

18. Restatement (Third) of Agency §§ 2.01-2.03 (Am. Law Inst. 2006). Section 2.01 and 2.02 provide that actual authority exists when "the agent reasonably believes, in accordance with the principal's manifestations to the agent, that the principal wishes the agent so to act." Section 2.03 addresses apparent authority arising from a third party's reasonable belief. For AWS, actual authority is the operative concept because the relevant relationship is between the principal and the system, not between the system and a third party relying on representations.

19. Restatement (Third) of Agency § 7.07 cmt. b (Am. Law Inst. 2006) (conduct is within the scope of employment when it is "of the kind [the employee] is employed to perform" and is "actuated, at least in part, by a purpose to serve" the employer). For AWS, the question is whether the system's targeting conduct was of the kind the deployment authorized, a question answered by the deployment order and operational parameters, both of which are documentary.

20. In practice, plaintiffs' counsel would name all parties in the chain simultaneously, relying on the contribution and indemnification framework to allocate responsibility among defendants after liability is established. This is standard mass tort strategy. See generally In re Agent Orange Prod. Liab. Litig., 635 F.2d 987 (2d Cir. 1980) (illustrating multi-defendant allocation in a complex military products case); United States v. Bestfoods, 524 U.S. 51, 64-65 (1998) (addressing parent corporation liability through subsidiary, confirming that the corporate liability chain does not terminate merely because intermediate entities are formally separate); Fed. R. Civ. P. 14 (permitting defendants to implead parties who may be liable for contribution or indemnification).

21. The transactional-framework and scope-of-employment objections are the author's own formulation of the strongest available resistance to the doctrinal move this Article makes. The sovereign immunity objection is grounded in Jurisdictional Immunities of the State (Ger. v. Italy: Greece intervening), Judgment, 2012 I.C.J. Rep. 99 (Feb. 3), which establishes that customary international law bars domestic civil jurisdiction over a foreign state's acts of armed forces even where serious violations of international humanitarian law are alleged. For the broader IHL accountability landscape from which these objections arise, see Michael N. Schmitt & Jeffrey S. Thurnher, "Out of the Loop": Autonomous Weapon Systems and the Law of Armed Conflict, 4 Harv. Nat'l Sec. J. 231 (2013) (analyzing the accountability mechanisms available under existing law of armed conflict for autonomous weapons deployment); Kenneth Anderson & Matthew Waxman, Law and Ethics for Autonomous Weapon

Systems: Why a Ban Won't Work and How the Laws of War Can (Hoover Institution, 2013) (arguing that existing IHL frameworks are sufficiently robust to address autonomous weapons without a treaty ban, and identifying the accountability gap as a central challenge to that position).

22. Respondeat superior's application in the employment discrimination context is analyzed in Faragher v. City of Boca Raton, 524 U.S. 775, 792 (1998), where the Court applied agency principles to Title VII hostile work environment liability. The foundational common law statement of the doctrine appears in Prosser and Keeton on Torts § 69, at 499 (5th ed. 1984), and traces to Turbervell v. Stamp, 1 Ld. Raym. 264, 91 Eng. Rep. 1072 (1697).

23. W. Page Keeton et al., Prosser and Keeton on the Law of Torts, 5th ed. (St. Paul: West Publishing Co., 1984), 499-500. The doctrine of respondeat superior makes the master liable for the torts of a servant committed in the course of employment. As Keeton explains, the historical foundation of the rule shifted over centuries - through the sixteenth century it was considered that the master should not be liable for the servant's torts at all, but the rule was found to be "far too narrow to fit the expanding complications of commerce and industry," and the law evolved through what amounted to "the fiction of a command to the servant," finally abandoning that fiction in favor of a more straightforward principle of enterprise liability. The basis of the doctrine, as the treatise summarizes it, is that the master has selected the servant, set the whole thing in motion, and is therefore responsible for what has happened - he has selected the servant and sought to profit by it, and it is just that he rather than the innocent injured plaintiff should bear the loss.

24. See Restatement (Third) of Agency § 2.04 (Am. Law Inst. 2006) ("An employer is subject to liability for torts committed by employees while acting within the scope of their employment."). The employer's liability is vicarious and does not depend on any fault on the employer's part, only on the scope of the employment relationship. See also Meyer v. Holley, 537 U.S. 280, 285 (2003) (reaffirming that respondeat superior makes principals "directly liable" for agents' authorized conduct regardless of the principal's knowledge or intent).

25. Keeton et al., Prosser and Keeton on the Law of Torts, 502-503. The scope of employment is "a highly indefinite phrase" whose very vagueness has been of value in permitting a desirable degree of flexibility in decisions, covering acts so "closely

connected with what the servant is employed to do, and so fairly and reasonably incidental to it, regarded as methods..." of carrying out the purposes of the employment that liability attaches, while excluding acts constituting a substantial departure - what the treatise distinguishes as a "frolic" rather than a mere "detour."

26. See Restatement (Third) of Agency § 7.07(1) (Am. Law Inst. 2006) ("An employer is subject to vicarious liability for a tort committed by an employee acting within the scope of employment.".

27. Compare API art. 86(2) (requiring that a superior "knew, or had information which should have enabled him to conclude in the circumstances at the time" that a violation was occurring) with Restatement (Third) of Agency § 7.07(1) (requiring only that the agent "act[ed] within the scope of employment," with no knowledge or intent requirement for the principal). See also Fed. R. Civ. P. 26(b)(1) (making discoverable "any nonprivileged matter that is relevant to any party's claim or defense," including deployment contracts, operational parameter specifications, and system behavior logs).

28. Deployment contracts and operational documentation are subject to discovery in civil litigation notwithstanding classification concerns, subject to the procedures established under the Classified Information Procedures Act, 18 U.S.C. app. 3 §§ 1-16 (2018), and Fed. R. Civ. P. 26(b)(1). For the comparable evidentiary framework in defense contractor litigation, see In re "Agent Orange" Prod. Liab. Litig., 96 F.R.D. 578 (E.D.N.Y. 1983). This decision established a framework for producing sensitive government-held documents in complex contractor litigation under protective order, demonstrating that classification and confidentiality concerns do not categorically bar civil discovery of military specifications and contractor communications.

29. Fed. R. Civ. P. 26(b)(1) (permitting discovery of "any nonprivileged matter that is relevant to any party's claim or defense and proportional to the needs of the case, considering... the importance of the discovery in resolving the issues"). Government contracts, system performance documentation, and testing records are routinely produced in defense contractor litigation under Fed. R. Civ. P. 26, subject to applicable privilege and national security restrictions.

30. The frolic and detour doctrine originates in Joel v. Morison (1834), 6 C. & P. 501; 172 E.R. 1338 (K.B.) (Parke B.), where Baron Parke held that a master could not be

liable for a servant acting "on a frolic of his own." See also Keeton et al., Prosser and Keeton on the Law of Torts, 504-05. (surveying the scope of the frolic doctrine and identifying factors distinguishing a frolic (which breaks the employment relationship) from a mere detour (which does not).

31. National Institute of Standards and Technology, Artificial Intelligence Risk Management Framework (AI RMF 1.0) 6, 38 (2023) [hereinafter NIST, AI RMF 1.0] (identifying "inherent uncertainties in AI systems" as a core risk characteristic, and noting the "inability to predict or detect the side effects of AI-based systems beyond statistical measures" as a feature distinguishing AI risk from traditional software risk).See also Stuart Russell & Peter Norvig, Artificial Intelligence: A Modern Approach 651-750 (4th ed. 2020) (explaining how reinforcement learning systems develop behaviors through environmental interaction that cannot be predicted from inspection of the training objective alone).

32. U.S. Department of Defense, Directive 3000.09, Autonomy in Weapon Systems (January 25, 2023), §§ 3.a, 4.d.(4). The directive requires that systems undergo "rigorous hardware and software V&V and realistic system developmental and operational T&E, including analysis of unanticipated emergent behavior," and that "adequate training, TTPs, and doctrine are available... to understand the functioning, capabilities, and limitations of the system's autonomy in realistic operational conditions" - requirements that, if satisfied, establish the deploying principal's knowledge of the system's behavioral characteristics, including its known limitations and failure modes.

33. Restatement (Third) of Agency § 7.07(3)(a) and cmt. b (Am. Law Inst. 2006) (defining employee as an agent whose principal controls or has the right to control the manner and means of the agent's performance of work, and explaining that "[w]hen an agent is not an employee, the principal lacks the right to control the manner and means of the agent's physical conduct in how work is performed"). This control-based distinction establishes the default rule that a principal is not vicariously liable for the torts of an independent contractor, subject to exceptions including the retained control exception and the nondelegable duty doctrine applicable to inherently dangerous activities.

34. The retained control exception is stated in Restatement (Third) of Torts: Physical & Emotional Harm § 56 (Am. Law Inst. 2012) and its predecessor Restatement (Second)

of Torts § 414 (Am. Law Inst. 1965): a principal who retains control over the operative details of an independent contractor's work is subject to direct liability for physical harm caused by the contractor's negligent failure to exercise reasonable care in carrying out that work. See Carney v. Union Pacific Railroad Co., 2016 IL 118984, 77 N.E.3d 1 (Ill. 2016) (providing authoritative analysis of § 414, confirming that the retained control exception imposes direct liability on the principal for its own negligence, and holding that "[t]here must be such a retention of a right of supervision that the contractor is not entirely free to do the work in his own way" for liability to attach under the exception). See also Hooker v. Department of Transportation, 27 Cal.4th 198, 210 (2002) (refining the retained control standard and holding that a hirer is not liable to an employee of an independent contractor merely because the hirer retained control over safety conditions at a worksite, but that a hirer is liable insofar as the hirer's exercise of retained control affirmatively contributed to the employee's injuries - establishing that retained control alone is insufficient and that the principal's actual exercise of that control must have causally contributed to the harm).

35. U.S. Dep't of Defense, Instruction 5000.02, supra note 15. See also U.S. Dep't of Defense, Directive on Responsible Artificial Intelligence (May 26, 2021) (establishing that DoD retains "oversight and accountability" for AI systems deployed by DoD components and that program offices must "maintain appropriate documentation of AI lifecycle activities," establishing the evidentiary record for retained control analysis).

36. The peculiar risk exception is stated in Restatement (Third) of Torts: Physical & Emotional Harm § 59 (Am. Law Inst. 2012), which consolidates and replaces Restatement (Second) of Torts §§ 416 and 427 (Am. Law Inst. 1965). Under § 59, an actor who hires an independent contractor for an activity that the actor knows or should know poses a peculiar risk is subject to vicarious liability for physical harm when the independent contractor is negligent as to the peculiar risk. A peculiar risk is one that, if reasonable care is not taken, produces a risk that differs in kind from the types of risk that are usual in the community. See Pusey v. Bator, 762 N.E.2d 968, 975 (Ohio 2002) (holding employer vicariously liable for armed security contractor's negligence under the inherently dangerous work exception, a doctrine the Third Restatement now addresses under the peculiar risk framework of § 59).

37. See Restatement (Second) of Torts §§ 416, 427 (Am. Law Inst. 1965); Restatement (Third) of Torts: Physical & Emotional Harm § 59 (Am. Law Inst. 2012). The principle that inherently dangerous work cannot be delegated so as to relieve the

principal of liability applies to both direct and vicarious liability claims. See Restatement (Third) of Torts: Physical & Emotional Harm § 55 cmt. a (Am. Law Inst. 2012) ("Section 413 subjects the hirer to liability for his own negligence if, in the face of a peculiar risk, the hirer fails to provide for precautions in the contract or fails to use reasonable care for the taking of such precautions. This rule is subsumed within the hirer's liability for negligence under § 55."); Saiz v. Belen School District, 827 P.2d 102, 110 (N.M. 1992) (holding that §§ 416 and 427 "represent different formulations of the same principle: the employer remains liable for injuries resulting from dangers the employer should have anticipated at the time the employer entered into the contract").

38. For the government contractor defense and its limits, see Boyle v. United Technologies Corp., 487 U.S. 500 (1988) (three-part test: (1) government approval of reasonably precise specifications; (2) conformance to those specifications; and (3) disclosure of known dangers). Whether an AI system's behavioral parameters constitute "reasonably precise specifications" under Boyle's meaning is an unresolved question future litigation will force courts to confront. See also Trevino v. General Dynamics Corp., 865 F.2d 1474, 1481 (5th Cir. 1989) (holding that where the government approves only imprecise or general guidelines, discretion over important design choices remains with the contractor and the defense is unavailable, because the defense requires that the government - not the contractor - be the primary agent of the relevant design decision; a mere rubber stamp by a federal procurement officer does not constitute approval sufficient to invoke the defense).

39. See United States v. Bestfoods, 524 U.S. 51, 64-65 (1998) (holding that a corporate parent may be held directly liable as an operator of a subsidiary's facility where the parent itself manages, directs, or conducts operations related to the harm - distinct from derivative liability through veil-piercing - establishing that corporate structure does not insulate a parent from liability for its own operational involvement in a subsidiary's activities). For the attribution problem that arises when liability must be allocated across multiple defendants in complex military products litigation, see In re "Agent Orange" Prod. Liab. Litig., 818 F.2d 187, 189 (2d Cir. 1987) (noting that no plaintiff could establish which manufacturer's product caused their injury because the government mixed products from multiple suppliers and stored them in unlabeled containers, rendering individual attribution impossible).

40. Restatement (Second) of Torts § 519 (Am. Law Inst. 1977) provides that "one who carries on an abnormally dangerous activity is subject to liability for harm to the

person, land or chattels of another resulting from the activity, although he has exercised the utmost care to prevent the harm." The doctrine traces to Rylands v. Fletcher, L.R. 3 H.L. 330 (1868) (H.L.) (establishing strict liability for non-natural use of land causing harm to neighboring property). See also In re Hanford Nuclear Rsrv. Litig., 534 F.3d 986, 1004-05 (9th Cir. 2008) (applying § 520 factors to hold that operation of a government nuclear weapons facility constituted an abnormally dangerous activity warranting strict liability, notwithstanding defendants' compliance with government direction and exercise of reasonable care, because the gravity of potential harm and the uncommon nature of the activity placed it squarely within the doctrine).

41. See Restatement (Second) of Torts § 519 cmt. d (Am. Law Inst. 1977) ("It is founded upon a policy of the law that imposes upon anyone who for his own purposes creates an abnormal risk of harm to his neighbors, the responsibility of relieving against that harm when it does in fact occur. The defendant's enterprise, in other words, is required to pay its way by compensating for the harm it causes, because of its special, abnormal and dangerous character."). See also W. Page Keeton et al., Prosser and Keeton on Torts § 78, at 556 (5th ed. 1984) (explaining that where an enterpriser deliberately engages in an activity that is highly dangerous even when reasonable care is exercised and that is not the kind commonly engaged in, "such intentional exposure of another to great danger, however socially desirable the activity, can generally be regarded as a sound basis on which to allocate the risk of loss to the person or entity engaging in that ultra-hazardous and abnormally dangerous activity").

42. Restatement (Second) of Torts § 520 (Am. Law Inst. 1977). The six factors are: (a) existence of a high degree of risk of some harm to the person, land or chattels of others; (b) likelihood that the harm that results from it will be great; (c) inability to eliminate the risk by the exercise of reasonable care; (d) extent to which the activity is not a matter of common usage; (e) inappropriateness of the activity to the place where it is carried on; and (f) extent to which its value to the community is outweighed by its dangerous attributes.

43. No civilian or commercial entity has developed or deployed autonomous targeting systems comparable to military AWS programs. The activity fails the "common usage" factor under id. § 520(d), which asks whether the activity is not a matter of common usage - a factor satisfied when the activity is not "customarily carried on by the great mass of mankind or by many people in the community" but rather is "carried on by only a comparatively small number of persons." Id. cmt. i. The development and

deployment of lethal autonomous weapons systems belongs exclusively to a narrow class of state military actors, placing it well outside any cognizable definition of common usage under the Restatement framework.

44. Holmes Jr., Oliver Wendell. The Common Law. Boston: Little, Brown, 1881, 3.

45. Restatement (Second) of Torts § 519 cmt. d (Am. Law Inst. 1977); Richard A. Posner, "A Theory of Negligence," Journal of Legal Studies 1 (1972): 75-76 ("This is a serious limitation of the negligence system as a method of optimizing the allocation of resources to safety. Yet the courts did not brush the problem under the rug entirely. They carved an important exception to the standard of negligence for ultrahazardous activities, such as blasting. Those are by definition activities where unavoidable accident costs are great, and therefore where one is most likely to find that an alternative method of achieving the same result (digging instead of blasting) is cheaper when unavoidable accident costs are taken into account. A rule of strict liability - the rule applied to activities classified as ultrahazardous - compels them to be taken into account."). In the AWS context, requiring proof of intent or knowledge systematically fails to internalize the costs of lethal autonomous deployment - the precise condition that the Restatement's strict liability rule is designed to address.

46. Restatement (Third) of Agency § 8.09 (Am. Law Inst. 2006) ("An agent has a duty to take action only within the scope of the agent's actual authority" and "a duty to comply with all lawful instructions received from the principal and persons designated by the principal concerning the agent's actions on behalf of the principal"). The ultra vires analysis in the agency law context is thus whether the agent's conduct exceeded the scope of the granted authority, a question determined by reference to the original delegation, not by the agent's subjective intent in taking the action.

47. The ultra vires doctrine's evolution in corporate law is traced in Ashbury Railway Carriage & Iron Co. v. Riche, L.R. 7 H.L. 653 (1875) (establishing strict ultra vires in English company law, holding that acts beyond a corporate charter were absolutely void). Modern American corporate law has largely moved away from the strict void-act consequence, treating unauthorized acts as voidable rather than void and shifting the inquiry to scope of authority. See Restatement (Third) of Agency § 8.09 (Am. Law Inst. 2006) (agent has duty not to act outside scope of authority granted by principal; principal's liability to third parties for unauthorized acts turns on whether apparent authority or ratification supplies a basis for attributing the act); see also Model

Business Corp. Act § 3.04 (2002) (eliminating the void-act consequence of strict ultra vires while preserving the actionability of authorized-scope-of-authority analysis).

48. Restatement (Third) of Agency § 8.09 cmt. b (Am. Law Inst. 2006) ("the boundary of an agent's rightful action is the scope of the agent's actual authority"); id. § 2.01 (actual authority exists only when "the agent reasonably believes, in accordance with the principal's manifestations to the agent, that the principal wishes the agent so to act"). The ultra vires question is therefore one of actual delegation, not of the agent's subjective construction of the scope of permission granted.

49. Restatement (Third) of Agency § 6.11(1) (Am. Law Inst. 2006) (providing that a principal's liability to third parties for an agent's representations turns on whether the agent acted with actual or apparent authority, and that a third party's right to protection depends on the absence of notice that the agent's conduct was unauthorized). The principal cannot invoke its own failure to clearly define the scope of the agent's authority as a shield against liability to third parties who had no means to verify those limits. Applied to AWS, the civilians who are injured by the system's conduct have no ability to verify the parameters of the deployment order; the deploying state's failure to prevent the harm cannot be excused by pointing to the system's exceedance of its own authorization.

50. See Russell & Norvig, supra note 31, at 712-15 (explaining that machine learning systems learn behavioral policies from training data and that their behavior in novel environments (environments not represented in the training distribution) cannot be predicted from the training objective alone).

51. National Institute of Standards and Technology, Artificial Intelligence Risk Management Framework (AI RMF 1.0) 6, 38 (2023) [hereinafter NIST, AI RMF 1.0] (noting that "measuring AI risks in a laboratory or a controlled environment may yield important insights pre-deployment, these measurements may differ from risks that emerge in operational, real-world settings," and identifying as a distinguishing feature of AI risk the "inability to predict or detect the side effects of AI-based systems beyond statistical measures," as well as the tendency of training datasets to "become detached from their original and intended context or may become stale or outdated relative to deployment context"). See also Daniel S. Hoadley & Nathan J. Lucas, Artificial Intelligence and National Security, Congressional Research Service Report R45178, at 29, 33 (Apr. 26, 2018) (noting that sensitivity to training data creates

"domain adaptability" failures when AI systems are deployed in settings that differ from their training environment, and documenting adversarial manipulation of image classification systems that caused AI to misidentify objects in ways undetectable through standard observation, including causing a system to misidentify a stop sign as a speed limit sign and machine guns as a helicopter).

52. Restatement (Third) of Agency § 7.07 cmt. b (Am. Law Inst. 2006) (stating that under the respondeat superior foreseeability framework, conduct falls outside the scope of employment only when it is "so unusual or startling that it seems unfair to include the loss resulting from it in the employer's business costs"). The unpredictability of AI systems in adversarial environments is neither unusual nor startling given the documented state of the technology at the time of deployment authorization. See also W. Page Keeton et al., Prosser and Keeton on Torts § 70, at 504-05 (5th ed. 1984) (foreseeability in the respondeat superior context asks whether the general type of harm was foreseeable in light of the nature of the employment, not whether the specific act was predicted).

53. U.S. Dep't of Defense, Directive 3000.09, supra note 8, §§ 3, 4.1.c(5), 4.1.d(3) (requiring rigorous verification, validation, and testing "including analysis of unanticipated emergent behavior" before deployment, and mandating that plans assess system reliability and suitability "under realistic conditions, including possible adversary actions" both before formal development and again before fielding).

54. Restatement (Fourth) of Foreign Relations Law § 451 (Am. Law Inst. 2018, updated 2024) (restating the general rule that states enjoy immunity from jurisdiction in the courts of other states, subject to exceptions). The immunity of states from domestic court jurisdiction is a fundamental principle of the international legal order, rooted in the sovereign equality of states recognized in U.N. Charter art. 2(1). See also The Schooner Exchange v. McFaddon, 11 U.S. (7 Cranch) 116, 136 (1812) (Marshall, C.J.) (establishing sovereign immunity in U.S. law on the basis that the equal dignity and independence of sovereigns precludes one nation's courts from exercising jurisdiction over another).

55. Foreign Sovereign Immunities Act of 1976, 28 U.S.C. §§ 1602-1611 (2018). The FSIA codified the shift from absolute to restrictive immunity, limiting immunity to sovereign acts (jure imperii) and withdrawing it for commercial acts (jure gestionis).

See Republic of Austria v. Altmann, 541 U.S. 677, 690 (2004) (tracing the history of sovereign immunity doctrine in U.S. law).

56. The two principal FSIA exceptions relevant to AWS accountability are the commercial activity exception, 28 U.S.C. § 1605(a)(2), and the tortious act exception, § 1605(a)(5). See Verlinden B.V. v. Cent. Bank of Nigeria, 461 U.S. 480, 488-89 (1983) (holding that the FSIA codifies the restrictive theory of sovereign immunity as substantive federal law and that subject matter jurisdiction in every action against a foreign sovereign depends on the existence of one of the Act's specified exceptions). See also Argentine Republic v. Amerada Hess Shipping Corp., 488 U.S. 428, 434-35 (1989) (establishing that the FSIA provides the sole and exclusive basis for obtaining jurisdiction over foreign states in U.S. courts).

57. See Saudi Arabia v. Nelson, 507 U.S. 349, 358 (1993) (holding that the commercial activity exception requires that the gravamen of the claim be based on a commercial activity of the foreign state, not merely that a commercial activity is somehow involved). Weapons deployment is a sovereign function (jure imperii) rather than a commercial activity within the meaning of § 1605(a)(2).

58. 28 U.S.C. § 1605(a)(5) (2018). The tortious act exception requires that the tort "occur[] in the United States," which AWS strikes on foreign civilians do not satisfy. See Argentine Republic v. Amerada Hess Shipping Corp., 488 U.S. 428, 439-40 (1989) (holding that the tortious act exception does not apply to acts taken on the high seas or in foreign territory).

59. 28 U.S.C. § 1605A (2018). The terrorism exception applies only to states designated as state sponsors of terrorism by the Secretary of State pursuant to applicable export control and foreign assistance statutes, a designation made on political grounds entirely unrelated to autonomous weapons development or deployment. See 28 U.S.C. § 1605A(h)(6) (defining "state sponsor of terrorism" by reference to applicable export control designations). As of 2024, the designated states are Cuba, Iran, North Korea, and Syria, none of which are the primary developers or deployers of advanced autonomous weapons systems.

60. Ferrini v. Federal Republic of Germany, Cass., Sez. Un., 11 Mar. 2004, n. 5044 (It.) (holding that German sovereign immunity did not bar Italian tort claims for Nazi-era

deportation and forced labor, reasoning that jus cogens violations strip a state of its immunity).

61. Jurisdictional Immunities of the State (Ger. v. Italy: Greece intervening), Judgment, 2012 I.C.J. Rep. 99, paras. 91-97 (Feb. 3). The majority held by twelve votes to three that no jus cogens exception to sovereign immunity exists under customary international law, even for grave violations of peremptory norms.

62. See Dissenting Opinion of Judge Cançado Trindade, Jurisdictional Immunities of the State, 2012 I.C.J. at 179 (arguing the majority's reasoning constitutes a denial of justice); Dissenting Opinion of Judge Yusuf, id. at 208 (arguing peremptory norms should prevail over procedural immunity rules when in conflict); Dissenting Opinion of Judge Gaja, id. at 222 (questioning the majority's categorical exclusion of a jus cogens exception).

63. See Judgment No. 238/2014, Corte Costituzionale (It.) (2014) (the Italian Constitutional Court declining to give effect to the ICJ's judgment insofar as it required dismissal of claims arising from crimes against humanity, holding that the Italian Constitution's fundamental principles prevented recognition of a rule of customary international law immunizing such acts); see also Sévrine Knuchel, State Immunity and the Promise of Jus Cogens, 9 Nw. J. Int'l Hum. Rts. 149 (2011).

64. See Sévrine Knuchel, State Immunity and the Promise of Jus Cogens, 9 Nw. J. Int'l Hum. Rts. 149 (2011) (arguing that peremptory norms function as a driver of customary law evolution toward an access to justice obligation for victims of jus cogens violations); Committee Against Torture, General Comment No. 3, U.N. Doc. CAT/C/GC/3 (Dec. 13, 2012) (addressing states' obligation to provide civil remedies for serious violations of the Convention); Draft Articles on Responsibility of States for Internationally Wrongful Acts arts. 40-41, in Report of the International Law Commission, U.N. GAOR, 56th Sess., Supp. No. 10, U.N. Doc. A/56/10 (2001) (establishing that serious breaches of peremptory norms generate obligations on all states, including the obligation not to recognize as lawful any situation created by such a breach). See also Philippa Webb, Human Rights and the Immunities of State Officials, in Erika De Wet & Jure Vidmar eds., Hierarchy in International Law: The Place of Human Rights 114 (2012) (drawing on the case law of twenty-four jurisdictions to examine the norm conflict between human rights accountability and sovereign

immunity, and concluding that existing case law does not yet evidence the emergence of a human rights-based hierarchy within international law).

65. Thomas Hobbes, Leviathan pt. II, ch. 17 (1651) ("Covenants, without the sword, are but words, and of no strength to secure a man at all."); id. pt. I, ch. 14 ("the bonds of words are too weak to bridle men's ambition, avarice, anger, and other passions, without the fear of some coercive power"). Hobbes wrote about the state of nature among individuals; international law operates in something disturbingly close to it where the parties are sovereign states that also happen to be permanent members of the Security Council.

66. Legality of the Threat or Use of Nuclear Weapons, Advisory Opinion, 1996 I.C.J. Rep. 226 (July 8); see also Military and Paramilitary Activities in and Against Nicaragua (Nicar. v. U.S.), Judgment on Merits, 1986 I.C.J. Rep. 14 (June 27) (demonstrating that even binding ICJ judgments may go unenforced when a permanent Security Council member exercises its veto to block enforcement measures); Congo, Ghana, Madagascar, Trinidad and Tobago and United Arab Emirates, Draft Resolution, U.N. Doc. S/18250 (July 31, 1986) (draft resolution calling for full compliance with the Nicaragua judgment, vetoed by the United States by a vote of 11-1-3).

67. Rome Statute of the International Criminal Court arts. 5, 25, 28, July 17, 1998, 2187 U.N.T.S. 3. Article 5 limits the Court's jurisdiction to "the most serious crimes of concern to the international community as a whole," enumerated as genocide, crimes against humanity, war crimes, and the crime of aggression. Article 25 establishes that a person who commits a crime within the Court's jurisdiction "shall be individually responsible and liable for punishment," including persons who order, solicit, induce, aid, abet, or otherwise assist in such crimes. Article 28 imposes command responsibility on any military commander who "knew or, owing to the circumstances at the time, should have known that the forces were committing or about to commit such crimes" and "failed to take all necessary and reasonable measures within his or her power to prevent or repress their commission." Command responsibility under Article 28 is theoretically applicable to commanders who deploy AWS that commit war crimes, but the practical constraints are substantial: the United States, China, and Russia have not ratified the Rome Statute, placing their nationals outside the Court's reach absent a Security Council referral - which each of those states can veto under U.N. Charter art. 27(3).

68. U.N. Charter art. 27(3) ("Decisions of the Security Council on all other matters shall be made by an affirmative vote of nine members including the concurring votes of the permanent members"). The five permanent members - the United States, Russia, China, the United Kingdom, and France - may individually block any Security Council resolution creating binding accountability mechanisms for autonomous weapons deployment, as any such resolution would constitute a decision on a substantive matter requiring their concurring votes.

69. Harold Hongju Koh, "Why Do Nations Obey International Law?," Yale Law Journal 106 (1997): 2599, 2603 (introducing the concept of "transnational legal process"); see generally id. (arguing that international law operates primarily through repeated cycles of interaction, interpretation, and internalization by which global norms penetrate domestic legal systems, rather than through centralized enforcement); Thomas M. Franck, "Legitimacy in the International System," American Journal of International Law 82 (1988): 705, 706 (arguing that states comply with international law primarily when they perceive rules as legitimate - as having "come into being in accordance with right process" - not when threatened with sanctions); Statute of the International Court of Justice art. 38(1)(b), June 26, 1945, 59 Stat. 1055, 33 U.N.T.S. 993 (defining customary international law as "a general practice accepted as law").

70. Convention Against Torture and Other Cruel, Inhuman or Degrading Treatment or Punishment art. 14, December 10, 1984, 1465 U.N.T.S. 85 (requiring each state party to "ensure in its legal system that the victim of an act of torture obtains redress and has an enforceable right to fair and adequate compensation, including the means for as full rehabilitation as possible"); see also Committee Against Torture, General Comment No. 3, Implementation of Article 14 by States Parties, U.N. Doc. CAT/C/GC/3 (December 13, 2012) (interpreting Article 14 to require states to investigate violations, prosecute alleged perpetrators, and provide reparation for victims, and confirming that a "person should be considered a victim regardless of whether the perpetrator of the violation is identified, apprehended, prosecuted or convicted" - extending obligations beyond state officials to non-state actors where the state failed to exercise due diligence).

71. International Convention on Civil Liability for Oil Pollution Damage, November 29, 1969, 973 U.N.T.S. 3; Protocol of 1992 to Amend the International Convention on the Establishment of an International Fund for Compensation for Oil Pollution Damage of 1971, November 27, 1992, 1953 U.N.T.S. 330. See International Oil Pollution

Compensation Funds, Annual Report 2022, 8-10 (2023) (reporting that the 1992 Fund and its predecessor the 1971 Fund have together been involved in over 150 incidents worldwide and have paid some £752 million in compensation since 1978).

72. S.C. Res. 687, paras. 16, 18, U.N. Doc. S/RES/687 (April 3, 1991) (paragraph 16 reaffirming Iraq's liability under international law for direct loss, damage, environmental damage, and injury to foreign governments, nationals, and corporations resulting from its unlawful invasion and occupation of Kuwait; paragraph 18 deciding to create a compensation fund and establishing a Commission to administer it).

73. United Nations Compensation Commission, Final Report of the Governing Council to the Security Council on the Work of the Commission, U.N. Doc. S/2022/104, paras. 2, 30, 49 (February 14, 2022) (reporting that the Commission resolved nearly 2.7 million submitted claims with an asserted value of over $352 billion, awarded $52.4 billion in compensation to approximately 1.5 million successful claimants, and completed its final payment in January 2022).

74. The principle that states may voluntarily waive sovereign immunity is foundational to international practice. See Verlinden B.V. v. Cent. Bank of Nigeria, 461 U.S. 480, 489 (1983) (recognizing explicit and implied waiver as FSIA exceptions under 28 U.S.C. § 1605(a)(1)). States regularly consent to arbitration through bilateral investment treaties and investment chapters in trade agreements, constituting a defined waiver of immunity for that category of dispute. See Convention on the Settlement of Investment Disputes between States and Nationals of Other States arts. 25-26, March 18, 1965, 575 U.N.T.S. 159 (ICSID Convention) (providing that jurisdiction extends to disputes which the parties "consent in writing to submit to the Centre" and that such consent, once given, may not be withdrawn unilaterally, and that the Preamble expressly declares that no state is obligated to submit any particular dispute by mere ratification, confirming that jurisdiction arises from affirmative, voluntary consent rather than treaty membership alone).

75. See Holmes, supra note 1, at 1 (characterizing common law development as proceeding through the accretion of experience rather than logical derivation).

76. Convention on Certain Conventional Weapons, Group of Governmental Experts on Emerging Technologies in the Area of Lethal Autonomous Weapons Systems, Report of the 2019 Session, U.N. Doc. CCW/GGE.1/2019/3, annex IV (Sept. 25, 2019). The CCW

Group of Governmental Experts has met annually since 2014 without producing a binding instrument. The eleven guiding principles adopted in 2019 are explicitly stated to be "without prejudice to the result of future discussions." The states declining to support a binding instrument are, without exception, those with the most advanced autonomous weapons development programs — a coincidence the existing literature does not treat as coincidental. See Paul Scharre, Army of None: Autonomous Weapons and the Future of War 349 (W.W. Norton & Co. 2018) (observing that states supporting a ban "are not major military powers" and that leading military powers are actively developing autonomous weapons capabilities). See also Campaign to Stop Killer Robots, Country Views on Killer Robots (July 7, 2020) (identifying Australia, France, Israel, Republic of Korea, Russia, Turkey, the United States, and the United Kingdom as states expressing firm opposition to negotiating a new treaty on fully autonomous weapons).

77. Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 (Artificial Intelligence Act), art. 2(3), 2024 O.J. (L) 1689. Article 2(3) provides that the Act "does not apply to AI systems where and in so far they are placed on the market, put into service, or used with or without modification exclusively for military, defence or national security purposes, regardless of the type of entity carrying out those activities." Recital 24 confirms that the exclusion extends to any entity, public or private, carrying out such activities, and identifies public international law as "the more appropriate legal framework for the regulation of AI systems in the context of the use of lethal force." The carve-out is broad enough to exclude virtually all autonomous weapons system applications from the Act's risk-classification and prohibited-practices provisions.

78. Holmes, The Path of the Law, supra note 1, at 461. The prophecy theory maps cleanly onto the enforceability deficits of soft-law governance frameworks. As Chinkin observes, much of the substantive content of soft law is "subjective and discretionary" and therefore "unsuited to adjudication" — its proper domain is the creation of expectations and frameworks for negotiation, not the generation of cognizable claims. C. M. Chinkin, The Challenge of Soft Law: Development and Change in International Law, 38 Int'l & Comp. L.Q. 850, 862 (1989). The instruments surveyed in this section operate precisely within that domain: they create expectations without generating obligations, and they organize discourse without supplying the "quality of legal consequences" that Chinkin identifies as a prerequisite for any instrument that

purports to regulate breach. Id. at 859. Under Holmes's test, an instrument that cannot found a cause of action is not law in any sense the prophecy theory recognizes.

79. Sparrow, supra note 11, at 67, 74–75. Sparrow argues that deployment of autonomous weapons systems is unethical because no candidate locus of responsibility, whether programmer, commanding officer, or machine, can bear just accountability for the deaths they cause, and that this gap violates a fundamental condition of jus in bello: "it is a necessary condition for fighting a just war, under the principle of jus in bello, that someone can be justly held responsible for deaths that occur in the course of the war." Id. at 67. This Article agrees with Sparrow's diagnosis and departs from his prescription: Sparrow concludes that the accountability gap compels prohibition; this Article argues that a principal-agent liability framework provides a workable accountability mechanism without requiring it.

80. See George J. Stigler, The Theory of Economic Regulation, 2 Bell J. Econ. & Mgmt. Sci. 3, 3 (1971) (originating the regulatory capture thesis: "as a rule, regulation is acquired by the industry and is designed and operated primarily for its benefit"). The international law analog, where the states being regulated are also the states doing the regulating, may produce more acute capture than in domestic contexts, because there is no separation between the legislative, regulatory, and subject classes. See also Koskenniemi, supra note 10, at 38, 42 (observing that international legal discourse must rely on "essentially contested — political — principles to justify outcomes to international disputes," and that where law tracks only the interests of effective states, it "becomes an apology for the interests of the powerful").

81. See Stockholm International Peace Research Institute, SIPRI Yearbook 2023: Armaments, Disarmament and International Security 17–19 (2023) (documenting that a handful of states continued to oppose even a two-tiered regulatory approach to autonomous weapons systems in the CCW process as of 2022). See also Scharre, supra note 76, at 349 (observing that leading military powers, including the United States, Russia, and China, are actively developing autonomous weapons capabilities while opposing binding regulatory instruments); Campaign to Stop Killer Robots, Country Views on Killer Robots, supra note 76 (documenting the positions of the United States, Russia, China, the United Kingdom, and France opposing binding legal instruments in the CCW process).

82. See Stigler, supra note 80, at 3, 10–12 (arguing that regulated industries systematically shape regulatory processes to serve their own interests, and that the political system structurally favors organized minorities over diffuse majorities).

83. See Felix S. Cohen, Transcendental Nonsense and the Functional Approach, 35 Colum. L. Rev. 809, 833-34 (1935) (arguing that legal scholarship's function is not merely to describe the law but to evaluate existing legal rules against functional criteria and propose alternatives when those rules fail). Applied here: identifying the accountability gap without specifying the architecture that would close it would leave the Article at the descriptive level that Sparrow and others have already occupied. The proposed treaty provisions attempt to move from diagnosis to design.

84. The proposed AWS Accountability Protocol draws its structural models from the Convention Against Torture's civil remedy requirement, supra note 70; the IOPCF's strict liability and pooled compensation mechanism, supra note 71; and the UNCC's individual claims commission model, supra note 72. The Protocol's immunity waiver provision follows the jurisdictional consent model of bilateral investment treaties and ICSID arbitration clauses, supra note 74. See also ILC Articles on State Responsibility, arts. 4–8, G.A. Res. 56/83, annex (Dec. 12, 2001) (providing the attribution rules that the Claims Commission would apply to determine which state's deployment caused the harm at issue). The viability of bilateral UN-state agreements as the legal basis for international accountability mechanisms operating outside Security Council mandatory authority is established by the jurisprudence of the Special Court for Sierra Leone. See Prosecutor v. Gbao, Case No. SCSL-2004-15-PT, Decision on Preliminary Motion on the Invalidity of the Agreement Between the United Nations and the Government of Sierra Leone on the Establishment of the Special Court for Sierra Leone (Special Ct. Sierra Leone, Appeals Chamber, May 25, 2004), paras. 5, 8–10 (holding that states may consent through bilateral treaty to international criminal accountability mechanisms whose legal basis is independent of Security Council Chapter VII mandatory authority, and that amnesty provisions do not irrevocably extinguish sovereign prosecution capacity under international law, a principle directly applicable to the AWS Protocol's voluntary immunity waiver design).

85. The model for treaty commentary supplementing spare operative provisions is well-established in international law. See, e.g., Commentary on the Draft Articles on Responsibility of States for Internationally Wrongful Acts, in Report of the International Law Commission, U.N. Doc. A/56/10 (2001) (providing article-by-article commentary

elaborating the ILC Articles' meaning, application to hard cases, and relationship to existing authority). The AWS Accountability Protocol's Commentary would perform the same function for the provisions proposed in Part VIII.

86. See Int'l Oil Pollution Compensation Funds, Annual Report 2022, at 34-35 (2023), supra note 71 (documenting the IOPCF's contribution formula, which calculates each contributor's annual levy based on the quantity of contributing oil received in the prior calendar year, thereby tying the contribution obligation directly to the extent of participation in the risk-creating activity). The proposed Protocol's analogous formula would tie state contributions to the scale and frequency of autonomous weapons deployments reported under the transparency provision, discussed infra Part VIII.vi.

87. 47 U.S.C. § 230(c)(1). Section 230 immunizes providers and users of interactive computer services from liability for content provided by another information content provider. Courts examining whether AI systems that generate harmful output fall outside this protection have focused on the passive/active distinction — specifically whether the developer's own design choices, rather than third-party content, are the source of the alleged harm. See Lemmon v. Snap, Inc., 995 F.3d 1085, 1091-93 (9th Cir. 2021) (holding that § 230 did not immunize Snap from negligent design liability because the plaintiffs' claim treated Snap as a product designer whose own architecture created the harm rather than as a publisher of third-party content, and that Snap "could have satisfied its alleged obligation ... without altering the content that Snapchat's users generate"); see also Andersen v. Stability AI Ltd., 700 F. Supp. 3d 853, 860-64 (N.D. Cal. 2023) (sustaining direct copyright infringement claim against AI image generator on the ground that the company actively shaped output through training data selection and model design rather than functioning as a passive conduit for third-party content); Doe v. GitHub, Inc., 672 F. Supp. 3d 837, 857-58 (N.D. Cal. 2023) (sustaining DMCA claims against AI code-generation tool on the ground that defendants intentionally designed the system to handle licensed training data in specific ways, treating developer design choices as active conduct). Whether this active-producer characterization removes AI-generated content from § 230 protection has not been definitively resolved at the circuit level; the Supreme Court granted certiorari in Gonzalez v. Google LLC, 598 U.S. 617 (2023), specifically to address the scope of § 230's passive/active distinction but declined to reach the question after finding the underlying complaint failed to state a claim on independent grounds.

88. Al-Adsani v. United Kingdom, App. No. 35763/97 [2001] ECHR 752 (Grand Chamber, Nov. 21, 2001), 34 Eur. H.R. Rep. 273. The Grand Chamber held nine to eight that sovereign immunity prevailed over the applicant's civil torture claims, finding no crystallized rule of international law subordinating state immunity to jus cogens violations in civil proceedings, while acknowledging that Article 6's right of access to a court was engaged and required proportionality analysis rather than automatic deference to immunity. See id. ¶¶ 49, 53-67. The joint dissent of Judges Rozakis and Caflisch, joined by Judges Wildhaber, Costa, Cabral Barreto, and Vajić, along with the separate dissents of Judges Ferrari Bravo and Loucaides, argued that jus cogens norms, as hierarchically superior rules of international law, automatically lift the procedural bar of state immunity when the two conflict, rendering the criminal/civil distinction adopted by the majority jurisprudentially incoherent. See Joint Dissenting Opinion of Judges Rozakis and Caflisch. See also Sévrine Knuchel, State Immunity and the Promise of Jus Cogens, 9 Nw. J. Int'l Hum. Rts. 149 (2011) (examining the normative hierarchy argument advanced by the Al-Adsani dissenters and arguing that jus cogens norms carry sufficient hierarchical force to override the procedural bar of state immunity in civil proceedings, a position whose logical force has been widely acknowledged in subsequent scholarship even by those who ultimately disagree with its conclusion); Lorna McGregor, State Immunity and Human Rights: Is There a Future after Germany v. Italy?, 11 J. Int'l Crim. Just. 125 (2012).

89. International Covenant on Civil and Political Rights art. 2(3), Dec. 16, 1966, 999 U.N.T.S. 171 (ratified by the United States June 8, 1992, subject to reservations, understandings, and declarations). Article 2(3) requires each state party to "ensure that any person whose rights or freedoms as herein recognized are violated shall have an effective remedy, notwithstanding that the violation has been committed by persons acting in an official capacity." The right to an effective remedy is elaborated in Human Rights Committee, General Comment No. 31, U.N. Doc. CCPR/C/21/Rev.1/Add.13 (May 26, 2004), ¶¶ 15–16, which requires states parties to ensure accessible and effective remedies for Covenant violations and specifies that reparation, including compensation where harm has occurred, is a component of the remedy obligation without which Article 2(3) is not discharged. None of the five U.S. substantive reservations addresses Article 2(3) or the remedy obligation; the United States qualified compensation rights only as applied to Articles 9(5) and 14(6), leaving the general remedy obligation of Article 2(3) unqualified by reservation or understanding. The U.S. declaration that Articles 1 through 27 are non-self-executing

limits direct judicial enforcement as a cause of action in domestic courts but does not extinguish the international obligation, which remains relevant to the constitutional and international law examination the text describes.

90. The Manville Corporation filed for Chapter 11 bankruptcy on August 26, 1982, citing estimated asbestos liability of approximately $1.9 billion. In re Johns-Manville Corp., 36 B.R. 727, 734-35 (Bankr. S.D.N.Y. 1984). The FAIR Act of 2005, S. 852, 109th Cong. (2005), introduced by Senator Arlen Specter, would have established a fund of approximately $136 billion - $90 billion from defendant participants and $46 billion from insurer participants - to compensate asbestos victims and would have channeled all future claims through an administrative process, removing them from the tort system. The bill failed to pass but represented a sustained industry effort to replace open-ended tort liability with a defined and capped administrative scheme, a response directly produced by the scale of litigation exposure. For the $70 billion figure, see Stephen J. Carroll et al., Asbestos Litigation, xxv (RAND Corporation, 2005) (estimating total asbestos litigation costs paid through 2002 at approximately $70 billion, with additional significant costs thereafter).

91. Master Settlement Agreement, Nov. 23, 1998 (settling claims of 46 states, the District of Columbia, and several territories against Philip Morris Incorporated, R.J. Reynolds Tobacco Company, Brown & Williamson Tobacco Corporation, and Lorillard Tobacco Company, with additional amounts to Mississippi, Florida, Texas, and Minnesota under separate earlier agreements; projected payments of approximately $206 billion over the first twenty-five years of the agreement, with annual payments continuing in perpetuity thereafter subject to volume and other adjustments). Family Smoking Prevention and Tobacco Control Act, Pub. L. No. 111-31, div. A, 123 Stat. 1776 (2009) (granting the FDA authority to regulate the manufacture, distribution, and marketing of tobacco products). For the industry's historical litigation posture and its evolution through the state attorney general actions, see generally Richard Kluger, Ashes to Ashes: America's Hundred-Year Cigarette War, the Public Health, and the Unabashed Triumph of Philip Morris (Alfred A. Knopf, 1996).

92. The SUPPORT for Patients and Communities Act, Pub. L. No. 115-271, 132 Stat. 3894 (2018), represented Congress's most comprehensive legislative response to the opioid crisis, addressing prevention, treatment, and recovery across Medicare, Medicaid, and federal drug enforcement frameworks. Total settlement amounts paid or committed by opioid manufacturers, distributors, and pharmacy chains as of 2024

exceed $50 billion, drawn from settlements with companies including Johnson & Johnson, the three major pharmaceutical distributors, and the national pharmacy chains. The Sackler family's proposed contribution of between $5.5 and $6 billion to the Purdue Pharma bankruptcy estate was structured as the price of releases from personal civil liability. The Supreme Court rejected that arrangement in Harrington v. Purdue Pharma L.P., 603 U.S. 204 (2024), holding that a bankruptcy plan cannot grant non-debtor third parties releases from civil liability without the consent of affected claimants, thereby leaving manufacturers without the nonconsensual immunity-purchase option that the lower courts had permitted. See In re Purdue Pharma L.P., No. 19-23649 (Bankr. S.D.N.Y.). For the pattern of litigation-induced regulatory reform across multiple industries, see generally David Rosenberg, The Causal Connection in Mass Exposure Cases: A "Public Law" Vision of the Tort System, 97 Harv. L. Rev. 849 (1984) (developing the theory that mass tort litigation performs a regulatory function when administrative agencies fail to act, and that sustained litigation exposure creates incentives for regulated industries to seek legislative and regulatory resolution).

93. The closest available authority for applying agency attribution principles to non-juridical automated systems appears in patent and regulatory contexts. See Akamai Techs., Inc. v. Limelight Networks, Inc., 797 F.3d 1020, 1022-23 & n.2 (Fed. Cir. 2015) (en banc) (deriving "direction or control" standard expressly from "general principles of vicarious liability" and holding that "an actor is liable for infringement under § 271(a) if it acts through an agent (applying traditional agency principles)"; court acknowledged in footnote that "vicarious liability is not a perfect analog" in the patent context but applied the underlying attribution logic of direction-and-control to conduct executed through third-party actors performing steps within a principal's defined operational framework); In re Dish Network, LLC, 28 FCC Rcd. 6574, 6582-84 (2013) (FCC applying federal common law agency principles, including formal agency, apparent authority, and ratification, to determine when a principal bears liability for TCPA violations committed by third-party telemarketers operating on its behalf; concluding that a seller may be held vicariously liable when it directs or controls the conduct of telemarketers acting within the scope of its authorization). Neither case involves tort liability, and neither holds that an automated system constitutes a "legal agent" in the full Restatement sense. Both apply the attribution logic of agency law to harm-causing conduct executed through third-party actors operating within a principal's delegated authorization, and that structural logic, principal liability for instrumentalized conduct within the scope of authorization, supplies the closest available predicate for the AWS

accountability argument advanced here, leaving its specifically tort and sovereign dimensions as genuine doctrinal extensions that no court has yet made.

94. The instrument/agent distinction is implicit throughout the Restatement's treatment of agency. See Restatement (Third) of Agency § 1.01 cmt. c (Am. Law Inst. 2006) (identifying the constitutive elements of agency as consent, action on the principal's behalf, and subordination to principal's control; the commentary focuses throughout on the relational structure rather than the ontological category of the actor). The distinction between a tool and the instrument of an agency relationship has practical significance in defense procurement contexts: DoD Directive 3000.09 requires that autonomous weapons systems be deployed pursuant to documented operational parameters specifying engagement criteria, authorized zones, and weapons configurations, a documentation requirement that has no analog for conventional munitions. See U.S. Dep't of Defense, Directive 3000.09, Autonomous Weapon Systems §§ 1.2, 4 (Jan. 25, 2023). A missile has no documented scope of authorized conduct subject to civil discovery; an AWS deployment does. That documentary asymmetry is precisely what agency law's relational analysis can reach and why the "mere tool" objection, however intuitive, does not survive contact with what the doctrine actually requires.

95. Gerald W. Boston, Strict Liability for Abnormally Dangerous Activity: The Negligence Barrier, 36 San Diego L. Rev. 597 (1999). Boston documents the judicial erosion of strict liability for abnormally dangerous activities through what he terms the "negligence barrier," identifying courts' increasing tendency to resolve the Restatement's third factor — inability to eliminate risk through reasonable care — in favor of defendants by finding that reasonable precautions could have prevented the harm. His analysis identifies this trend as the primary doctrinal mechanism by which strict liability has approached functional extinction in contexts where its underlying rationale remains fully applicable, a trajectory this Article argues the irreducible behavioral unpredictability of ML-based autonomous weapons systems is positioned to disrupt.

96. Statute of the International Court of Justice art. 34(1), June 26, 1945, 59 Stat. 1055, 33 U.N.T.S. 993 ("Only states may be parties in cases before the Court."). Article 34(1) establishes the foundational jurisdictional limitation of the ICJ, confining access to the Court to sovereign states and thereby excluding individuals, corporations, and non-governmental entities from direct standing. In the autonomous weapons

accountability context, this limitation means that victims of AWS-caused harm have no direct avenue of redress before the principal international adjudicatory body, reinforcing the structural remedial gap this Article addresses through the corporate and agency law framework.