NORTH CAROLINA JOURNAL OF LAW & TECHNOLOGY
VOLUME 23, ISSUE 1: OCTOBER 2021

PREDICT AND SUSPECT: THE EMERGENCE OF ARTIFICIAL LEGAL MEANING

Daniel Maggen*

Recent theoretical writings on the possibility that algorithms would someday be able to create law have delayed algorithmic law-making—and the need to decide on its legitimacy—to some future time in which algorithms would be able to replace human lawmakers. This Article argues that such discussions risk essentializing an anthropomorphic image of the algorithmic lawmaker as a unified decision-maker and divert attention away from algorithmic systems that are already performing functions that, together, have a profound effect on legal implementation, interpretation, and development. Adding to the rich scholarship of the distortive effects of algorithmic systems, this Article suggests that state-of-the-art algorithms capable of limited legal analysis can have the effect of preventing legal development. Such algorithm-induced ossification, this Article argues, raises questions of legitimacy that are no less consequential than those raised by some futuristic algorithms that can actively create norms.

To demonstrate this point, this Article puts forward a hypothetical example of algorithms performing limited legal analysis to assist healthcare professionals in reporting suspected child maltreatment. Already in use are systems performing risk analysis to aid child protective services in screening maltreatment

* Visiting Fellow, Information Society Project at Yale Law School; Lecturer, Yale University. This Article benefited immensely in the various stages of its development from conversations with Gilad Abiri, Jack Balkin, Joshua Fairfield, and Christina Spiesel, as well as from presentations at the Information Society Project Workshop and the Law & Technology Virtual Workshop at The Nebraska Governance & Technology Center. I also want to thank the editorial board of the North Carolina Journal of Law & Technology for their fantastic work and, in particular, Anna Comer, Meredith Doswell, Alexandra Farquhar, Amy Jones, Gabrielle Supak, and Alec Suttle for their excellent comments and suggestions. All remaining mistakes are my own.

reports. Drawing on the example of algorithms increasingly used today in social media content moderation, this Article suggests that similar systems could be used for flagging cases that show signs of suspected abuse. Accordingly, such assistive systems, this Article argues, will likely cement the prevailing legal meaning of maltreatment. As mandated child-abuse reporters increasingly rely on such systems, the result would be the absence of legal evolution, inhibiting changes to contentious elements in the legal definition of "reportable suspicion," including, for example, the scope of acceptable physical disciplining. Together with the familiar effect of existing systems, the effect of this hypothetical algorithmic system could have a profound impact on the path of the law regarding child maltreatment, equivalent in its significance to the effect that autonomous algorithmic adjudication would have.

TABLE OF CONTENTS

I. INTRODUCTION 69
II. THOUGHT EXPERIMENTS 72
A. The Three-Pronged Hypothetical 73
B. AI Lawmaking Now 75
III. LEGAL TECHNOLOGIES 77
A. Traditional Algorithms 78
B. Machine Learning Algorithms 81
C. Big Data 85
IV. AUXILIARY SYSTEMS AND THE MEANING OF THE NORM 88
A. Implementation and Distortion 89
B. Algorithms and Interpretation 95
V. ASSISTIVE SYSTEMS AND LEGAL CHANGE 97
A. Assistive Systems 97
B. Preventing Legal Change 102
VI. PREDICTION AND SUSPICION: THE CASE OF MANDATORY REPORTING 105
A. Mandatory Reporting 106
B. Auxiliary Reporting Systems 109

C. The Need for Algorithmic Assistance *111*
D. Assistive Reporting Systems *114*
E. Creating the Legal Meaning of Maltreatment *119*
VII. CONCLUSION 122

"The judges of normality are present everywhere."[1]

I. INTRODUCTION

Theorizing about the legal meaning of artificial intelligence often involves thought experiments. Scholars want to keep ahead of the curve lest society finds in retrospect that it surrendered its legal fate to algorithmic overlords; being ready entails drawing on present experiences in order to prepare for future developments before those developments take place.[2] However, projecting the legal debate to some future time can have the adverse effect of obfuscating minute contemporary development by stressing more apparent future changes, so that by the time the hypothetical is made possible, the debate will have already been concluded.

This problem often happens in the debate on the meaning of decision-making algorithms in the legal realm. In a number of influential pieces, Lawrence Solum and Eugene Volokh use thought experiments to offer illuminating discussions on the legitimacy of algorithmic norm-setting—meaning the use of computer systems to autonomously produce, through legislation and adjudication, the norms by which human beings live.[3] Both Solum and Volokh suggest that, in the non-immediate future, machine learning systems will attain the functional capacity to create norms at a level that at least matches human capabilities, and that, by virtue of their computational superiority, society should favorably, or at least seriously, consider substituting algorithmic for human norm-

[1] MICHEL FOUCAULT, DISCIPLINE AND PUNISH 304 (Alan Sheridan trans., Vintage Books 2d ed. 1995) (1977).

[2] *See* Lawrence B. Solum, *Artificial Meaning*, 89 WASH. L. REV. 69, 77 (2014).

[3] *See id.*; Lawrence B. Solum, *Artificially Intelligent Law*, 1 BIOLAW J. 53 (2019); Eugene Volokh, *Chief Justice Robots*, 68 DUKE L.J. 1135 (2019).

setting.[4] Rather than ignore the myriad of reasons to resist algorithmic law-making, these thought experiments are meant to generate debate on the nature of law, legal interpretation, and legitimacy.[5] However, by postponing the discussion to a future time when algorithms could replace humans as norm-setters, these discussions can desensitize society to the fact that the effects Solum and Volokh discuss are already taking place. As a society, humans are on the cusp of a world affected by artificial legal meaning, and any delay in deciding on its legitimacy can have lasting effects.

The thought experiments Solum and Volokh offer discuss the rise of algorithmic law-making progress in three general stages. In the first stage, which has already become a regular part of contemporary legal reality, algorithmic systems perform *auxiliary* functions that support human decision-making.[6] In the second, which is materializing before our eyes, these systems perform *assistive* functions, comparable to those of human agents but subject to human discretion with regard to the decision itself, particularly with respect to matters of accountability.[7] Finally, in the third step,[8] algorithmic decision-making becomes *autonomous*, in the sense that the system itself makes the normative decisions in question, with no effective human involvement.[9] Solum and Volokh concede that progressing from the second to the third stage raises considerable normative questions but argue that ultimately, no inherent reasons exist to suggest that society avoid taking this step: the legitimacy of

---

[4] *See* Solum, *supra* note 3, at 62.

[5] *See id.* at 62; Solum, *supra* note 2, at 85; Volokh, *supra* note 3, at 1137.

[6] *See* Solum, *supra* note 3, at 53–54; Volokh, *supra* note 3, at 1149. For a discussion on the state of such systems, see *infra* Part IV.

[7] *See* Solum, *supra* note 3, at 54. As discussed below, Volokh does not seem to clearly distinguish between the first and second stages. Volokh, *supra* note 3, at 1154. For a discussion on such systems, see *infra* Part V(A).

[8] This three-stage progression assumes that no additional step follows in which these algorithm systems decide that an ideal normative world would involve ridding the world of human beings. *See* NICK BOSTROM, SUPERINTELLIGENCE 140–54 (2014).

[9] *See* Solum, *supra* note 2, at 75; Solum, *supra* note 3, at 54; Volokh, *supra* note 3, at 1142.

autonomous algorithmic law-making should depend solely on the systems' competencies relative to comparable human adjudicators.[10]

Setting aside the third step's prudence and legitimacy, this Article argues that the move from the second to the third stage is a mirage; the normative implications of algorithmic decision-making are already apparent in the combination of the first two stages. This Article further suggests that delaying the time of scrutiny to the third stage can be based on the thought that creating legal meaning necessarily involves changing legal norms. However, it is now patently clear: auxiliary algorithmic systems, such as those used to provide legal decision-makers with risk predictions, have a considerable effect on the meaning of the legal categories in which those systems operate.[11] Less noticed, however, has been the more profound effect that assistive systems can have on legal development by shaping the legal narratives made available to decision-makers. For example, by determining which cases are brought before human adjudicators, such assistive systems inversely participate in law-making by hindering the law's natural development. Even though such systems do not (yet) generate novel legal paradigms, relying on their assistance effectively means ceding human control over legal development to these assistive systems, limiting the law's future course to those legal classifications that informed the systems' creation.

To demonstrate this point, this Article offers a thought experiment of its own, involving the use of algorithmic systems that assist in the mandated reporting of child maltreatment. Such systems already operate in the auxiliary stage, aiding child protective services to respond to complaints by predicting the level of risk involved.[12] The thought experiment suggests taking such systems to the next level by designing machine learning systems to assist mandated reporters in determining which cases give rise to "reportable suspicion"—meaning a suspicion of child abuse that is

---

[10] *See* Solum, *supra* note 3, at 62; Volokh, *supra* note 3, at 1138.
[11] *See infra* Part IV.
[12] *See infra* Part VI(B).

sufficiently probable, involves serious harm, and does not fall under the acceptable physical disciplining exception, as discussed below.

This Article suggests that, even in the assistive stage (the second stage), algorithmic systems can produce legal meaning by preventing the natural development of reportable maltreatment's meaning. Algorithmic systems have this effect by constricting reporters' decisions to those that meet the algorithm's definition of maltreatment, which is in turn tied to the legal paradigms that informed the algorithm's training process. By insulating the meaning of "maltreatment" from social changes, such as those concerning the legal implications of physical disciplining, these algorithmic legal decision-making systems effectively determine the path of law.

II. THOUGHT EXPERIMENTS

Until not so long ago, the idea of artificially intelligent adjudication was invoked mainly as a thought exercise to tell something about the nature of legal adjudication and its connection to human agency.[13] In recent years, as computer software has become progressively better at emulating human decision-making, legal scholarship has shifted to seriously discuss the legitimacy of using adjudicative algorithms.[14] Still, since contemporary technology is quite far from producing algorithms capable of comprehensive legal analysis, any talk of judicial algorithms remains hypothetical. Nevertheless, as opposed to past creative exercises' discussion of artificial adjudication, today's thought experiments are meant to lay the groundwork for the possibility that, someday soon, computers will be capable of successfully emulating

---

[13] *See* Joshua P. Davis, *Law Without Mind: AI, Ethics, and Jurisprudence*, 55 CAL. W. L. REV. 165, 181 (2018); Lawrence B. Solum, *Legal Personhood for Artificial Intelligences*, 70 N.C. L. REV. 1231, 1231 (1992); Cass R. Sunstein, *Of Artificial Intelligence and Legal Reasoning*, 18 PUB. L. & LEGAL THEORY WORKING PAPERS 1, 1–4 (2001).

[14] *See* Davis, *supra* note 13, at 171–72; Kiel Brennan-Marquez & Stephen E. Henderson, *Artificial Intelligence and Role-Reversible Judgment*, 109 J. CRIM. L. & CRIMINOLOGY 137, 143 (2019).

humans' ability to pass judgment. A successful contemporary thought exercise, therefore, will not only teach something about the law, but will also help the legal community anticipate this potential artificial development.[15]

Unfortunately, these thought experiments' futuristic perspective can also prepare the legal community for the wrong thing; as debate ensues in preparation for the rise of robo-judges, scholars and commentators can become oblivious to the fact that, for all intents and purposes, algorithms have already taken the helm of adjudication. Such mental preparations are analogous to preparing society for the age of vacuuming robots and leaving it unmindful to the reality in which algorithms have taken over the task of vacuuming by inhabiting the vacuum cleaner itself. In other words, just like the mental image of an android using a manual vacuum cleaner, thought experiments that fixate on the judicial decision-maker's agency, discretion, and creativity risk essentializing the image of the autonomous, norm-generating judge or justice. All the while, this fixation neglects the fact that adjudication can involve discrete functions, and that these functions can effectively, if not essentially, be taken over by computer algorithms without some attention-grabbing judicial usurpation. To prepare for such a creeping takeover, the legal community must focus its attention not on the future development of judicial software, but on what algorithms are already doing now, as these algorithms slowly but steadily shape increasing portions of the legal landscape in their image.

A. The Three-Pronged Hypothetical

The thought experiments Solum and Volokh present offer insightful discussions that illuminate the nature of artificial adjudication but, at the same time, also risk obscuring the exigency of the discussions. Any attempt to briefly survey these beautifully crafted hypotheticals would do them terrible wrong; still, it can be generally said that despite their differences, the two models offered in this Article follow the progression of legally-minded algorithms

[15] Solum, *supra* note 3, at 62; Volokh, *supra* note 3, at 1138.

from (1) systems that have been in use for some time now, to (2) emerging and near-future use cases based on state-of-the-art machine learning technology, and lastly, (3) to hypothetical future usages based on not-yet-existing technology. The first two stages of this progression can be described as tracking legal algorithms from *auxiliary* systems that operate in the service of broader legal tasks to *assistive* systems that aid legal decision-makers in adjudication by offering limited forms of legal analysis.[16] The latter systems, actual or hypothetical, are capable of some form of legal reasoning but lack the ability to see the big picture, which is an essential part of any legal analysis. The thought experiments in this Article, however, focus their attention on the third stage, involving autonomous *adjudicative* systems, animated by still nonexistent technology.[17]

In these scenarios, the algorithm's advancement from the auxiliary to the assistive stage is incremental, following the quantitative evolution of its capabilities.[18] Although the human decision-maker increasingly relies on the software's assistance, the final say remains in human hands.[19] In contrast, the move from the second to the third stage in these scenarios is abrupt, highly visible, and immediately consequential, occurring the moment that human decision-makers are taken out of the picture and the algorithms take full control of the process.[20] Volokh describes this move as the "AI promotion,"[21] suggesting that it would be a "startling step," but one worth taking.[22] In a similar vein, Solum writes that taking this step "would surely be controversial" and would raise considerable questions of legitimacy.[23]

---

[16] Volokh, *supra* note 3, at 1149.
[17] *See* Solum, *supra* note 2, at 85; Solum, *supra* note 3, at 55; *id.* at 1137, 1146–47.
[18] *See* Solum, *supra* note 3, at 54; Volokh, *supra* note 3, at 1146–47.
[19] *See* Solum, *supra* note 3, at 53–54.
[20] *See id.* at 54; Solum, *supra* note 2, at 74.
[21] *See* Volokh, *supra* note 3, at 1142, 1156–77.
[22] *See id.* at 1142–43.
[23] Solum, *supra* note 3, at 58–59.

Both discussions intend to draw attention to the impending rise of algorithmic adjudication and greet this rise with a calculated, pragmatic approach.[24] This past-present-future structure has the effect of erroneously affording the discussion surrounding the legitimacy of this consequential shift ample time to assess its implications and relative worth. In truth, the time for deliberation is now, as the discrete functions discussed can come together in piecemeal and have an effect that is comparable to that of the third stage, even in the absence of some clear artificial entity capable of displacing human operators *in toto*. The remainder of this Article seeks to illustrate this point by offering a thought experiment that takes place in the second stage, demonstrating that even assistive systems can reach effective law-making status.

B. Artificial Lawmaking Now

In focusing on the disruptiveness and legitimacy of the third, adjudicative stage of legal algorithm's development, Solum and Volokh's thought experiments make two implicit assumptions that the thought experiment discussed in this Article questions. The first assumption is that the progress from the first, auxiliary stage to the second, assistive stage is mainly quantitative, naturally occurring as the algorithm becomes progressively better at what it does and is thus accorded greater responsibility. This notion of incremental progress conceals the fact that the pattern of development algorithms have followed from the past to the present involves not just greater accuracy but also a qualitative leap, as algorithms progressed from systems only capable of offering *factual* analysis to enhanced systems that can emulate normative decision-making, albeit limited in breadth.[25] Focusing on the hypothetical, highly-visible step of replacing humans with algorithmic decision-makers can overshadow this less visible but just as consequential shift from merely factual to normative algorithms. If this shift is indeed the case, then asking whether algorithmic takeover is legitimate is

[24] *See* Solum, *supra* note 2, at 85; *id.* at 55, 58; Volokh, *supra* note 3, at 1137.

[25] *See infra* Part V. Volokh also introduces this distinction, although he seems to identify both factual and legal analysis in which algorithms are capable of performing robust legal functions. *See* Volokh, *supra* note 3, at 1154–56.

beside the point; instead, what needs to be asked is whether the effects of normative algorithms on legal decision-making are congruent with legitimate law-making.

Second, and relatedly, these hypotheticals assume that the change symbolizing the rise of algorithmic adjudicators involves algorithms acquiring some essential quality that is synonymous with adjudication *qua* the creation of legal meaning. For Solum, this transformative step entails that the algorithm exhibits three qualities: (1) the ability to generate legal norms, (2) the ability to apply the norms generated, and (3) the ability to modify those norms in response to varying factual conditions.[26] Volokh likewise sees norm-creation as the quintessential quality of adjudication, although Volokh suggests that this quality can be measured according to the algorithm's ability to create persuasive legal arguments.[27] As a result, and unlike Solum, Volokh includes in the third stage, not just the image of the algorithmic law-maker, but also the image of the algorithmic attorney—both involving a meaningful sense of judicial agency and are thus equally futuristic.[28] Despite this difference, for both authors, the real normative discussion begins when algorithms acquire some distinct capability that captures the essence of adjudication.

The thought experiment below opposes these two assumptions by suggesting that existing or near-future systems that are still in the assistive stage have already largely acquired the capacities that put the systems in a position to effectively create legal meaning. Computer algorithms are already extensively used to produce first-stage auxiliary systems, aiding decision-makers in making legal decisions by providing them with relevant factual information.[29] Likewise, social media platforms today extensively employ machine learning algorithms in second-stage assistive functions to flag prohibited materials for further human scrutiny—at times basing

---

[26] *See* Solum, *supra* note 2, at 75; Solum, *supra* note 3, at 57.
[27] *See* Volokh, *supra* note 3, at 1182–84.
[28] *See id.* at 1146–47.
[29] *See infra* Part IV.

their determinations on legal classifications, such as non-consent and terrorism.[30] This Article's thought experiment examines the possibility that similar technologies will be used in the near future to assist healthcare mandated reporters in determining whether their patients' cases mandate reporting. Additionally, this Article suggests that reporters' probable use of such systems will likely make the reporters reliant on the system's judgment to determine which cases should undergo further scrutiny. As the thought experiment demonstrates, the growing reliance on algorithmic systems can have minute and unfelt influences, accumulating into a systemic effect that can fulfill the conditions that Solum suggests are the marks of law-making; as the algorithms transform the meaning of existing norms, determine these norms' implementation, and (inversely) affect the course of legal development. Although little differentiates this hypothetical system from the one already in use in social media, the proposed example more clearly illustrates the effects such systems can have on the development of artificial legal meaning.

III. LEGAL TECHNOLOGIES

Although the age of robotic judges is still far off, the legal domain is already accustomed to the use of "artificial intelligence," meaning computer systems capable of performing tasks that would be indicative of intelligence when performed by human beings.[31] Unlike the anticipated—and feared—artificial general intelligence, contemporary artificially intelligent systems perform very specific functions in limited settings and to narrowly defined effects. Accordingly, any discussion of such systems must be firmly rooted

---

[30] *See* DAVID FREEMAN ENGSTROM ET AL., GOVERNMENT BY ALGORITHM: ARTIFICIAL INTELLIGENCE IN FEDERAL ADMINISTRATIVE AGENCIES 19 (2020); David Lehr & Paul Ohm, *Playing with the Data: What Legal Scholars Should Learn About Machine Learning*, 51 U.C. DAVIS L. REV. 653, 676 (2017); Daniel Maggen, *Law In, Law Out: Legalistic Filter Bubbles and the Algorithmic Prevention of Nonconsensual Pornography*, 43 CARDOZO L. REV. (forthcoming 2022).

[31] TOSHINORI MUNAKATA, FUNDAMENTALS OF THE NEW ARTIFICIAL INTELLIGENCE 1 (David Gries & Fred B. Schneider eds., 2d ed. 2008). *But see* NILS J. NILSSON, THE QUEST FOR ARTIFICIAL INTELLIGENCE 483 (2010).

in the specific tasks the systems perform and how the systems are designed to perform those tasks, as well as in the general environment that shapes the systems' design and operation. Even though it is unnecessary to know the exact details of either of those conditions, the general principles that animate artificially intelligent systems and their basic premises set the limits for these systems' use cases and determine the dynamics such systems impose on those who use them.

A. Traditional Algorithms

Until somewhat recently, legal algorithms, and artificial intelligence more generally, relied almost exclusively on manually-created algorithms.[32] Crafting such algorithms was primarily an exercise in formal logic representation that involved creating models of the desired tasks and transforming the tasks into programmable if-then-else rules.[33] A familiar example of manually-created algorithms are "expert systems": computer algorithms that transform subject-matter expertise into formal-logic models put into a user-friendly computer software.[34] Creating such systems is as much a product of subject-matter expertise as it is of coding; unlike machine learning's automated pattern-seeking approach, expert systems heavily rely on human know-how to painstakingly shoehorn knowledge into precise rules and definitions.[35]

An expert system that has been the focus of considerable scholarly interest is one that leverages statistical expertise to produce algorithms capable of offering risk predictions.[36] With the advent of the age of "big data," statistics have shown great promise

---

[32] *See* Frank Pasquale, *A Rule of Persons, Not Machines: The Limits of Legal Automation*, 87 GEO. WASH. L. REV. 1, 29, 48, 50–51 (2019).

[33] *See* Kenneth A. Bamberger, *Technologies of Compliance: Risk and Regulation in a Digital Age*, 88 TEX. L. REV. 669, 688 (2010).

[34] STUART J. RUSSELL & PETER NORVIG, ARTIFICIAL INTELLIGENCE A MODERN APPROACH 22–24 (Stuart Russell & Peter Norvig eds., 4th ed. 2021).

[35] *See* Harry Surden, *Artificial Intelligence and Law: An Overview*, 35 GA. ST. U. L. REV. 1305, 1317 (2019).

[36] *See* Ric Simmons, *Big Data, Machine Judges, and the Legitimacy of the Criminal Justice System*, 52 U.C. DAVIS L. REV. 1067, 1072–74 (2018).

in plotting correlations with impressive precision, at times obviating the need to explain the causal connection between reasons and consequences.[37] With sufficient data, regression analysis can, for instance, be used to demonstrate the useful connection between different independent variables and a dependent variable that presumably ensues from the independent variables, even when the connection is inexplicable.[38] Statistical analysis has thus been used to quantify the relation between various criminogenic variables—factors that experts identify as associated with crime—and the occurrence of criminal acts, ostensibly demonstrating the probability of crime taking place whenever a set of indicators is observed.[39] The statistical model that maps the probabilistic connection between the indicators and the occurrence of crime can then be translated into an algorithm that provides a "risk score" for any given case.[40] In other words, the heavy lifting in risk-prediction expert systems is done by the statistical analysis, and transforming the statistically-produced models into algorithms that animate user-friendly software can provide legal decision-makers with a "mechanical statistician" to be used whenever a legal need arises.

The use of expert systems of this sort has become most dominant in bail decisions,[41] sentencing,[42] and "predictive policing"—a technology that has helped law enforcement agencies manage their

---

[37] VIKTOR MAYER-SCHÖNBERGER & KENNETH CUKIER, BIG DATA: A REVOLUTION THAT WILL TRANSFORM HOW WE LIVE, WORK, AND THINK 50–72 (2013).

[38] Frank Pasquale & Glyn Cashwell, *Prediction, Persuasion, and the Jurisprudence of Behaviorism*, 68 U. TORONTO L.J. 63, 73 (2018).

[39] For discussions of the mechanism of risk prediction, *see* Andrew Guthrie Ferguson, *Big Data and Predictive Reasonable Suspicion*, 163 U. PA. L. REV. 327, 370 (2015); *see also* Elizabeth E. Joh, *The New Surveillance Discretion: Automated Suspicion, Big Data, and Policing*, 10 HARV. L. & POL'Y REV. 15, 16 (2016); Michael L. Rich, *Machine Learning, Automated Suspicion Algorithms, and the Fourth Amendment*, 164 U. PA. L. REV. 871, 877 (2016).

[40] *See* Sandra G. Mayson, *Dangerous Defendants*, 127 YALE L.J. 490, 509 (2017).

[41] *See* John Logan Koepke & David G. Robinson, *Danger Ahead: Risk Assessment and the Future of Bail Reform*, 93 WASH. L. REV. 1725, 1757–65 (2018).

[42] *See* Jessica M. Eaglin, *Constructing Recidivism Risk*, 67 EMORY L.J. 59, 69–72 (2017).

limited surveillance resources more efficiently.[43] Although these applications, increasingly being implemented in all jurisdictions,[44] do not directly replace human discretion, these expert systems are nonetheless viewed as revolutionary police force multipliers that greatly expand law enforcement agencies' reach and formalize adjudication.[45] Some see this increased implementation as a generally welcome development;[46] however, others highlight its pitfalls and the need to develop appropriate frameworks of accountability and oversight in order to meet this expansion of the states' powers.[47]

Despite the widespread adoption of risk-prediction systems, legal expert systems never really gained traction and are generally falling out of style.[48] Using formal logic to represent subject-matter expertise can be a formidable task, both difficult and expensive. Furthermore, expert-produced models are inherently limited in their ability to produce actionable legal insight.[49] For example, a law enforcement professional's ability to observe a situation and determine whether it is "suspicious" involves the intuitive weighing of numerous considerations—an intricate process that can hardly be translated into straightforward, or even complex, logic-based rules.[50] Although legal and other professional decisions often follow general principles that can (at least in theory) be modeled by subject-matter experts and transformed into logic-based algorithms, any significant

---

[43] For discussions of predictive policing, *see* ENGSTROM ET AL., *supra* note 30, at 17; Simmons, *supra* note 36, at 1069–70.

[44] *See* Simmons, *supra* note 36, at 1072.

[45] *See* Joh, *supra* note 39, at 19.

[46] *See* Cary Coglianese & David Lehr, *Regulating by Robot: Administrative Decision Making in the Machine-Learning Era*, 105 GEO. L.J. 1147, 1161 (2017) (noting the growing usage of algorithms in the legal context).

[47] *See* Simmons, *supra* note 36, at 1075–77.

[48] *See* Pasquale, *supra* note 32, at 48.

[49] *See* Surden, *supra* note 35, at 1309.

[50] *See* Jason Millar & Ian Kerr, *Delegation, Relinquishment, and Responsibility: The Prospect of Expert Robots*, *in* ROBOT L. 102 (Ryan Calo et al. eds., 2015).

attempt to even come close to capturing the essence of legal judgment in this way would inevitably grow to gargantuan size.[51]

B. Machine Learning Algorithms

As computer scientists pushed forward in the quest for human-level functionality, the solution to this seemingly insurmountable challenge came from a shift to machine learning.[52] Without needlessly going into details beyond those that will be of use below, the difference between machine learning and logic-based, manually-created algorithms can be illustrated by the attempt to model human language.[53] Early efforts in natural language processing, based on the manual transformation of linguistic rules of syntax and grammar into formal instructions, were initially successful in "teaching" computers to understand language from the inside, so to speak.[54] However, as has quickly become evident, any practical use of such top-down systems would involve, not only the immense task of modeling the complex structures that create natural languages, but also representing in formal terms the vast amounts of worldly knowledge any ordinary use of language relies upon.[55] While some researchers still soldier on,[56] most have shifted from meticulously modeling language and representing knowledge, to devising methods for algorithms to automatically discern patterns of language usage from provided examples, often guided by linguistic expertise.[57] Machine learning did away with the need to represent human knowledge manually, instead focusing on the "datafied" expressions of any form of knowledge to create algorithms that can

---

[51] This challenge has become most evident in the field of natural language processing, essential for legal analysis. *See* NILSSON, *supra* note 31, at 103–21.

[52] *See id.* at 398–425.

[53] *See id.* at 431.

[54] *See id.* at 103–21.

[55] *See id.* at 354–61.

[56] A notable example is the Cyc knowledge representation platform. *See* CYCORP, https://www.cyc.com [https://perma.cc/48YY-6AL2].

[57] *See* NILSSON, *supra* note 31, at 431–36.

emulate knowledge-based decisions by observing the patterns past decisions have left in their wake.[58]

Like more traditional forms of data mining, machine learning is reliant on bifurcated datasets of input and output data: the output data being the end result of the sought-after function and the input data standing for more or less everything else in the datasets.[59] In the oft-used example of spam classification, the datasets include past spam classifications as the output data and all other information about the emails as the input data.[60] Deciding what data to use as output data is therefore synonymous with determining the essence of the algorithm.[61]

Once the training set is prepared, machine learning stands for various methods of learning from the datasets the relationship between the input and output data.[62] The learning algorithm is iteratively "trained" on a dataset—each iteration bringing the algorithm closer to creating a new algorithm that best represents a process or "function" that produced patterns in the data.[63] Like the process used by expert statisticians, a machine learning algorithm is meant to model the connection between the independent and dependent variables in the training set.[64] In *supervised* learning, the learning algorithm primarily creates this model by starting from relatively random configurations of the input/output relationship and measuring each configuration's *fitness*, meaning its congruence with this relationship in the training data.[65] Measuring each iteration's fitness and bringing that measurement to bear on the emerging model—a process commonly referred to as minimizing its

---

[58] *See* Coglianese & Lehr, *supra* note 47, at 1167; MAYER-SCHÖNBERGER & CUKIER, *supra* note 37, at 171–200.

[59] *See* JOHN D. KELLEHER, DEEP LEARNING 26 (2019); Lehr & Ohm, *supra* note 30, at 677.

[60] *See* Harry Surden, *Machine Learning and Law*, 89 WASH. L. REV. 87, 91 (2014).

[61] *See* Lehr & Ohm, *supra* note 30, at 673–74.

[62] *See* KELLEHER, *supra* note 59, at 185–230.

[63] *See id.* at 6–12.

[64] *See* Pasquale & Cashwell, *supra* note 38, at 73.

[65] *See id.* at 13; Deven R. Desai & Joshua A. Kroll, *Trust but Verify: A Guide to Algorithms and the Law*, 31 HARV. J.L. & TECH. 1, 28 (2017).

"objective function"—is the crux of supervised machine learning, making the original output variable used in the datasets the immutable touchstone of everything the ensuing algorithm could possibly do in the future.[66] Absent designer intervention, nothing irreducible to the original output variable—the legal classification in this Article's hypothetical case—would be of meaning to the resulting algorithm.[67]

The main difference between the two approaches, apart from speed and efficiency, is that machine learning, especially in its advanced forms, is not restricted to modeling a relatively limited and predetermined number of variables, or "features," connected through the sought-after "function" to the dependent variable.[68] Machine learning can, for instance, theoretically analyze a training set in its entirety to create a holistic model of criminality (at least as far as criminality is represented in the data), connecting the occurrence of criminal behavior and every bit of information that correlates to it.[69] Although the effort to conserve computational resources and avoid "overfitting" the data commonly leads designers to reduce the number of features modeled by the learning algorithm, the algorithm can nonetheless far surpass human efforts by creating "hyper-dimensional" models out of a large number of features, including variables that would not intuitively strike experts as relevant.[70] Furthermore, state-of-the-art machine learning methods, especially those that involve *deep* learning, include stages of mathematical abstraction that can effectively extract from the training set implicit features that are irreducible to semantic meaning.[71]

---

[66] *See* KELLEHER, *supra* note 59, at 21–22.

[67] *See* Alexander Campolo & Kate Crawford, *Enchanted Determinism: Power without Responsibility in Artificial Intelligence*, 6 ENGAGING SCI. TECH. & SOC'Y 1, 10–12 (2020).

[68] *See* ETHEM ALPAYDIN, INTRODUCTION TO MACHINE LEARNING 2 (3d ed. 2014).

[69] This theoretical analysis does not mean that such a holistic model would be useful or even feasible. In fact, much effort is put into reducing the number of features that learning algorithms consider in order to conserve computational efforts and avoid overfitting. *See* IAN GOODFELLOW ET AL., DEEP LEARNING 417 (2016).

[70] *See* RUSSELL & NORVIG, *supra* note 34, at 751–54.

[71] *See* GOODFELLOW ET AL., *supra* note 69, at 498.

These "deep learning" abilities have made possible the creation of algorithms that can successfully model human capabilities even in areas dominated by human intuition and imagination that not long ago were considered impervious to computer emulation.[72] Provided a sufficiently large number of *labeled* examples, supervised learning has proven to be surprisingly apt at emulating such tasks, not only replacing logic-based systems, but also far outpacing them.[73] Thus, using the large multilanguage depository of digitized books at its disposal, Google, a leader in the field of machine learning, was capable of producing surprisingly reasonable automatic translations, a task that seemed almost unimaginable not long ago: Google trained its algorithms to discern patterns from parallel passages found in books published in different languages.[74] In other cases, "unsupervised machine learning" (a group of methods used to train algorithms on *unlabeled* examples) and "reinforcement learning" (a group of methods that uses positive or negative feedback to direct the learning process) have proved even more remarkable by virtue of their ability to go beyond the constraints of human labeling.[75]

These advancements allow machine learning to take on the previously unimaginable task of emulating genuine legal analysis—

---

[72] The most striking recent example comes from GPT-3, the latest natural language processing algorithm, used to write entries for the New York Times' *Modern Love* section. *See* Cade Metz, *When A.I. Falls in Love*, N.Y. TIMES (Nov. 24, 2020), https://www.nytimes.com/2020/11/24/science/artificial-intelligence-gpt3-writing-love.html [https://perma.cc/FKF6-JHLR].

[73] Data scientists use various methods to come up with labeled datasets. *See* KELLEHER, *supra* note 59 (discussing the use of available labeled databases), B. W. Silverman & M. C. Jones, *E. Fix and J.L. Hodges (1951): An Important Contribution to Nonparametric Discriminant Analysis and Density Estimation*, 57 INT'L STAT. REV. 233 (1989) (discussing the use of mathematical methods for deducing missing labels); Kate Crawford & Vladan Joler, *Anatomy of an AI System*, A.I. NOW INST. & SHARE LAB (2018), https://anatomyof.ai [https://perma.cc/PN3K-X5DK] (discussing the use of cheap human labor to manually label examples).

[74] *See* MAYER-SCHÖNBERGER & CUKIER, *supra* note 37, at 85.

[75] *See* RUSSELL & NORVIG, *supra* note 34, at 651–53.

at least to some level.[76] Legal decisions, like determining whether something constitutes reportable suspicion, are exercises in open-ended and subtle classification. The intuitive determinations that lie at the heart of such judgment cannot be reduced to a small number of logic-based rules or the interaction between a limited number of variables.[77] Unlike traditional algorithms, advanced machine learning methods have proven surprisingly adept at emulating intricate human abilities by extracting from relevant datasets the kind of intuitive meanings and connections that cannot be expressed by formal, articulable rules.[78] This novel skill does not, however, mean that the road is open for full-blown algorithmic legal analysis. Genuine adjudication entails not just correctly implementing legal rules but also intimate knowledge of the legal domain and a keen understanding of the environment in which the decision is made, components that cannot be easily extracted from legal databases.[79]

C. Big Data

The evolution of machine learning goes hand in hand with the rise of big data. Machine learning has existed since the middle of the previous century, with various methods being used to automatically express data patterns as algorithms.[80] However, initial advances in the field were short-lived because, for this ability to be useful, machine learning must rely on large quantities of relevant, accessible, and analyzable data.[81] The falling costs of data storage

---

[76] *See* Ryan Calo, *Artificial Intelligence Policy: A Primer and Roadmap*, 51 U.C. Davis L. REV. 399, 405 (2017); Danielle Ensign et al., *Runaway Feedback Loops in Predictive Policing*, 81 PROC. MACH. LEARNING RSCH. 1, 1 (2018); Surden, *supra* note 60, at 94.

[77] *See* Hani Nouman & Ravit Alfandari, *Identifying Children Suspected for Maltreatment: The Assessment Process Taken by Healthcare Professionals Working in Community Healthcare Services*, 113 CHILD. & YOUTH SERVS. REV. 1, 2 (2020).

[78] *See id.* at 6; Lehr & Ohm, *supra* note 30, at 678 (discussing convolutional neural networks).

[79] *See id.*; Volokh, *supra* note 3, at 1159.

[80] *See generally* PEDRO DOMINGOS, THE MASTER ALGORITHM (2015) (discussing the dominant approaches to machine learning).

[81] *See* Coglianese & Lehr, *supra* note 47, at 1164–65; Lehr & Ohm, *supra* note 30, at 677.

and computation, as well as advancements in data analysis and learning methods, have transformed machine learning and other forms of statistical analysis into primary ways of gleaning knowledge directly from data, potentially disposing of the need to acquire insights from meticulous theorizing and expertise.[82]

These developments have altered the purpose for which algorithms are used from mainly modeling established connections between independent and dependent variables—say, the numerical relation between an applicant's income and the risk of loan default, to use a familiar example—to uncovering previously unknown patterns representing insights hidden in the data.[83] The ability to model the subterranean forces and hidden influences buried in the data has led to the development of "data mining": the practice of using data obtained for one purpose (or even without a purpose, such as collecting the "data exhaust" individuals passively emit) to reveal meaningful, even intimate, insights into the phenomena manifest in the data.[84] Data has thus become a resource, both valuable and endlessly exploitable; once information undergoes datafication, making it suitable intake for big data algorithms, the produced data can be endlessly reused and repurposed, even for previously unanticipated uses.[85] Although datafication can provide immense benefits, these developments also produce numerous deleterious effects.[86]

In turn, the ballooning of data necessitates the use of algorithms to keep data under control. Big data, in this sense, underlies both the

---

[82] *See* MAYER-SCHÖNBERGER & CUKIER, *supra* note 37, at 123–49.

[83] *See id.* at 19–31.

[84] *See id.* at 111–15; HUAN LIU & HIROSHI MOTODA, FEATURE EXTRACTION, CONSTRUCTION AND SELECTION 3–5 (1998).

[85] *See* Joh, *supra* note 39, at 20.

[86] *See generally* NICHOLAS CARR, THE GLASS CAGE 183–210 (2014) (discussing the risks of increasing dependance on automation); CATHY O'NEIL, WEAPONS OF MATH DESTRUCTION (2016) (discussing the dangers of algorithmic decision-making); JOSEPH TUROW, THE DAILY YOU (2011) (discussing the negative effects of algorithmic-based advertising); SHOSHANA ZUBOFF, THE AGE OF SURVEILLANCE CAPITALISM (2019) (discussing corporate use of predictive analytics).

need for automated screening to make data manageable and also the technology that responds to this need by putting in place algorithmic gatekeepers.[87] The availability of big data creates an opportunity that soon becomes a necessity: its vastness buttresses the ability to shift from understanding complex processes to modeling them—immediately opening the floodgates to a sea of precious knowledge that can be handled only by embracing this paradigm shift and adopting its products.[88] The advent of big data thus puts machine learning algorithms in a position to determine what information is seen by human users, creating a sizable winnowing effect on what data passes through the algorithmic gatekeepers and into human hands.[89] Often, the sheer scale of data makes unassisted human decisions impossible; and many times, the breadth of data is itself the product of algorithms.[90] The massive amounts of data gathered on social media, for instance, are largely the result of the algorithms that control their conveyance: without algorithms to search, analyze, and use these data, this information would not be publicly available.[91] The data on YouTube's servers, for instance, is "big" by virtue of its enormous size, but YouTube would not have swollen to this size without the algorithms that facilitate the matchmaking between videos and viewers.[92]

At other times, big data facilitates the creation of machine learning algorithms that make assisted decision-making a cost-effective alternative to unmediated human decisions. Even when a task does not involve handling massive amounts of data, such as the screening of patient information, the use of machine learning algorithms can free time and resources that can be of better use in making the clinical decision itself, making algorithms appear to be

---

[87] *See* MAYER-SCHÖNBERGER & CUKIER, *supra* note 37, at 7.

[88] *See id.* at 50–72.

[89] *See* Bamberger, *supra* note 33, at 707; Maayan Perel & Niva Elkin-Koren, *Black Box Tinkering: Beyond Disclosure in Algorithmic Enforcement*, 69 FLA. L. REV. 181, 185 (2017).

[90] MAYER-SCHÖNBERGER & CUKIER, *supra* note 37, at 73–97.

[91] *See, e.g.*, Desai & Kroll, *supra* note 65, at 51.

[92] *See Search and Discovery on YouTube*, YOUTUBE CREATOR ACADEMY, https://creatoracademy.youtube.com/page/lesson/discovery#strategies-zippy-link-1 [https://perma.cc/UX7A-DUZC].

attractive alternatives to human labor. And in yet another category of cases, which will mainly be left out of this Article's discussion, the relevant data are themselves the products of algorithms. Algorithms can, for instance, be proactively used to "mine" publicly available data, video surveillance, and even commercially available data exhaust, moving from a stance of passively monitoring freely provided information to proactive surveillance in search of actionable data.[93] In either of these use cases and for these different reasons, algorithms can come to dominate the data stream by determining what information merits human attention.[94] Even when the target data are available for manual inspection, they can become "subconscious" knowledge negotiated by an algorithmic superego.[95]

IV. AUXILIARY SYSTEMS AND THE MEANING OF THE NORM

In the three-prong framework offered by Solum and Volokh, the first stage, that of auxiliary legal systems, is far from hypothetical. Algorithms embedded in various measurement and analysis devices have long informed legal decisions.[96] Legal decisions are often grounded in empirical data that are increasingly the products of algorithms, from biological, physical, and chemical analysis to facial recognition and other advanced forms of data collection and processing.[97]

Advanced and transformative as these usages may be, the algorithms that animate these usages can appear to be inert tools, with only minor substantive legal effects.[98] Nevertheless, the profound effects that even such auxiliary algorithmic systems can have on the implementation and interpretation of a legal norm have

---

[93] *See, e.g.*, Calo, *supra* note 76, at 421; Desai & Kroll, *supra* note 65, at 50–51; Joh, *supra* note 39, at 28; Perel & Elkin-Koren, *supra* note 89, at 185.

[94] *See* Surden, *supra* note 35, at 1326.

[95] *See* Tarleton Gillespie, *The Relevance of Algorithms*, *in* MEDIA TECHNOLOGIES: ESSAYS ON COMMUNICATION, MATERIALITY, AND SOCIETY 167, 175–76 (Tarleton Gillespie et al. eds., 2014).

[96] *See* Andrea Roth, *Machine Testimony*, 126 YALE L.J. 1972, 1975–77 (2017).

[97] *See id.*

[98] *See id.* at 2001.

become familiar themes in legal scholarship.[99] Although these systems are incapable of actively engaging with the law's normative meaning, hidden biases and invisible design choices can nonetheless affect the legal norm's practical meaning by determining the enforcement patterns of private and public agents.[100]

A. Implementation and Distortion

As many legal scholars note, algorithms are prone to producing skewed factual findings, leading to biased applications of legal norms.[101] Such distortions can result either from inaccurate specifications, meaning the incorrect translation of the legal task into the specific requirement defining the algorithm's function, or from the faulty implementation of specifications in the algorithm's design.[102] The gap between the abstract legal norm, the system's express specifications, and the algorithm's implementation of either is filled with design choices that are fertile ground for distortions

---

[99] *See* Ben Green & Yiling Chen, *Algorithmic Risk Assessments Can Alter Human Decision-Making Processes in High-Stakes Government Contexts*, 5 PROC. ACM HUM.-COMPUT. INTERACTION, 418:1, 418:1 (2021); Lehr & Ohm, *supra* note 30, at 678; Perel & Elkin-Koren, *supra* note 89, at 185; Joshua A. Kroll et al., *Accountable Algorithms*, 165 U. PA. L. REV. 633, 633 (2017); Roth, *supra* note 96, at 2021–22; Surden, *supra* note 60, at 101; Surden, *supra* note 35.

[100] For discussion of legal algorithms in general, *see* Coglianese & Lehr, *supra* note 47, at 1170; ENGSTROM ET AL., *supra* note 30, at 16, 22; Perel & Elkin-Koren, *supra* note 89, at 183. For discussion of algorithm use by state agencies, *see, e.g.*, RASHIDA RICHARDSON ET AL., LITIGATING ALGORITHMS 2019 US REPORT: NEW CHALLENGES TO GOVERNMENT USE OF ALGORITHMIC DECISION SYSTEMS (2019); Joh, *supra* note 39; Rich, *supra* note 39; Rashida Richardson et al., *Dirty Data, Bad Predictions: How Civil Rights Violations Impact Police Data, Predictive Policing Systems, and Justice*, 94 N.Y.U. L. REV. 15 (2019).

[101] For examples of this focus on accuracy, *see, e.g.*, Bamberger, *supra* note 33, at 676; Calo, *supra* note 76, at 415.

[102] *See* Sebastian Benthall & Bruce D. Haynes, *Racial Categories in Machine Learning*, *in* 2019 PROCEEDING OF THE CONFERENCE ON FAIRNESS, ACCOUNTABILITY, AND TRANSPARENCY 289, 294–96; Deirdre K. Mulligan & Kenneth A. Bamberger, *Saving Governance-by-Design*, 106 CAL. L. REV. 697, 718 (2018).

and biases that can widen the distance between the law on the books and how the law is applied.[103]

Different stages in the design and operation of machine learning algorithms can give rise to different distortions of a legal norm. Machine learning can be separated into two broad stages: the creation of the algorithm and the operation of the algorithm.[104] Although some machine learning technologies are designed to work "online," thereby retraining the algorithm as it operates,[105] the kind of work that commonly goes into preparing datasets often entails, at the very least, retraining that is performed on batches of new data.[106] This limitation is particularly true for legal algorithms as they commonly require a rigorous stage of preparation, often involving the extensive use of subject-matter expertise.[107]

The first steps in creating a legal algorithm involve translating the desired legal task into sufficiently exact specifications and the subsequent transformation of those specifications, either manually or through machine learning, into a programmable algorithm that can then be turned into user-friendly software.[108] This stepped transformation from law to software can give rise to various mistranslations, distorting the original meaning of the legal task, so that even when the algorithm works as specified, it fails to meet the requirements assumed by the relevant legal norm.[109] Oftentimes, the translation from norm to code and the specifications that control the translation are constrained by cost-effectiveness and the availability of relevant data.[110] As Deirdre Mulligan and Kenneth Bamberger demonstrate, these design decisions, as well as choices of method,

---

[103] *See* Bamberger, *supra* note 33, at 722–23; Andrew D. Selbst & Solon Barocas, *The Intuitive Appeal of Explainable Machines*, 87 FORDHAM L. REV. 1085, 1125 (2018).

[104] *See* Lehr & Ohm, *supra* note 30, at 655.

[105] *See* Kroll et al., *supra* note 99, at 660.

[106] For more information on online, batch, and mini batch training, *see* ETHEM ALPAYDIN, MACHINE LEARNING 90–91 (2016).

[107] *See* ENGSTROM ET AL., *supra* note 30, at 24–25.

[108] *See* KELLEHER, *supra* note 59, at 22–30.

[109] *See* Mulligan & Bamberger, *supra* note 102, at 718.

[110] *See* Lehr & Ohm, *supra* note 30, at 675.

media, and formulations, are rarely value-neutral and are further compounded by designers' cognitive biases.[111] These biases, Bamberger notes, are prone to skewing the ensuing algorithm's results, creating a mismatch with its original legal purpose.[112] A compelling example that Bamberger provides concerns risk prediction. As Bamberger suggests, translating legal risk determinations into risk-predicting algorithms involves the inherent danger of privileging measurable and quantifiable data and specifications with the result of downplaying the importance of information that is not easily quantified.[113] This bias, Bamberger warns, can, in turn, come to mean that implementation of the norm misrepresents the kinds of risks it is meant to address.[114]

Such distortions can also abound in the learning process itself, as training data that supposedly hold the key to the desired legal function are fed into a learning algorithm meant to extract this function. Training can fail to produce a representative model of the desired legal function, either due to problems with the training set's predictiveness or because of a failure to accurately extract the function from the data. When speaking of the first category of failures, many often remark that machine learning can only be as good as the data.[115] The adage "garbage in, garbage out" has come to represent the fact that machine learning is basically a way of modeling datasets, so any problem in the datasets is bound to be reflected in the ensuing algorithm.[116] At times, such problems concern datasets that do not include enough valuable information. The data in the training sets are meant to serve the learning algorithm as a gateway to the "ground truth" about the world, with every dataset adding more information accordingly.[117] Machine learning is generally contingent on the availability of "big data"—

---

[111] Mulligan & Bamberger, *supra* note 102, at 708–11.
[112] Bamberger, *supra* note 33, at 728.
[113] *See id.* at 712.
[114] *See id.* at 676, 708.
[115] *See* JOHN D. KELLEHER & BRENDAN TIERNEY, DATA SCIENCE 35–36 (2018).
[116] *See* NILSSON, *supra* note 31, at 10.
[117] *See* ENGSTROM ET AL., *supra* note 30, at 25; *see* Jane Bambauer & Tal Zarsky, *The Algorithm Game*, 94 NOTRE DAME L. REV. 1, 10 (2018).

often, smaller datasets mean a less accurate algorithm.[118] However, the datapoints also need to be pertinent to the desired task, meaning that the datapoints are capable of establishing a connection between relevant features and the desired outcome. Too little data that establishes this connection would usually result in an insufficiently precise algorithm that exhibits too much variance in its predictions to be of use, sending designers back to the drawing board.[119]

Too much variance, however, is not the worst problem that bad data can cause, as criticisms of the "bias in, bias out" type demonstrate.[120] As many scholars note, the law has been historically biased against minorities due to pervasive bigotry and more nuanced structural inequalities.[121] As a result, the historical databases used to create legal algorithms are steeped in discrimination and thus produce factual findings that result in the biased application of the legal norm.[122]

Beyond the imitation of inherent bias, distorted algorithms can result from inadequate training data. Using the typical example of credit scoring, an algorithm can be trained on a dataset that includes a large number of past loan applications but relatively few minority applicants, making ensuing predictions less accurate for future minority applicants.[123] More often, however, such problems are illustrative of a more profound failure of predictiveness. The dataset on which the algorithm trains is supposed to be a useful proxy for the environment with which the algorithm is meant to engage, capturing some "concepts" that express the underlying ground truth.[124] The concept of "default risk," for instance, can be ill-

---

[118] *See* KELLEHER, *supra* note 59, at 21.

[119] *See* Lehr & Ohm, *supra* note 30, at 675.

[120] *See* Sandra G. Mayson, *Bias in, Bias out*, 128 YALE L.J. 2218 (2019).

[121] *See, e.g.*, Simmons, *supra* note 36, at 1074–76.

[122] *See id.*, Ferguson, *supra* note 39, at 401; Kroll et al., *supra* note 99, at 681.

[123] *See* Lehr & Ohm, *supra* note 30, at 680.

[124] For more context on "concept attainment" and "concept drift," *see* Jeffrey C. Schlimmer & Richard H. Granger, Jr., *Incremental Learning from Noisy Data*, 1 MACH. LEARNING 317, 317–18 (1986); Gerhard Widmer & Miroslav Kubat, *Learning in the Presence of Concept Drift and Hidden Contexts*, 23 MACH. LEARNING 69, 69–71 (1996).

captured by a dataset that misrepresents its spread in the real world.[125] Thus, a dataset that includes only those applicants granted a loan can fail to adequately capture this concept—such as when it fails to include a representative number of minority applicants and therefore produces an algorithm that erroneously assigns minority applicants a high risk value.[126] Problems of concept can skirt the line between bug and feature. Training an algorithm on a dataset comprising of past decisions by bank officers will produce an algorithm that mimics bankers; this outcome may or may not be what the algorithm is intended to do.[127] If the algorithm is meant to rank insolvency risk as an abstract function, any biases that bank officers commonly display will taint the data, resulting in an algorithm that is similarly erroneous in performing this function.[128] If, in contrast, the purpose of the algorithm is to mimic human behavior, warts and all—for example, an algorithm used to predict judicial decisions—such distortions will be features of its operation rather than bugs.[129] Often the problem is that available data pushes the concept in the direction of mimicking behavior, making it difficult to weed out biases or even categorize them as errors.[130]

Conceptual problems can be particularly daunting since they can prove difficult to spot. The gold standard for evaluating an algorithm is to test it on "unseen data," meaning data that was not included in the datasets on which the algorithm was trained; an imprecise algorithm can be easily rooted out if it fails this test.[131] However, the data on which "unseen data" tests are run is commonly taken from the same source that produced the training set, so any conceptual problems endemic to the source will be undiscoverable by this form

---

[125] *See* ENGSTROM ET AL., *supra* note 30, at 25.

[126] *See* Lehr & Ohm, *supra* note 30, at 680–81.

[127] *See* Andrew D. Selbst, *Disparate Impact in Big Data Policing*, 52 GA. L. REV. 109, 140–42 (2017).

[128] *See* Kroll et al., *supra* note 99, at 680.

[129] *See* Selbst, *supra* note 127, at 141–42.

[130] *See id.*

[131] *See* KELLEHER, *supra* note 59, at 14–15.

of testing.[132] Even when the algorithm begins operating in the real world, it can be challenging to spot the ways in which bias-riddled concepts prevent the algorithm from producing accurate evaluations. Since the measurements for accuracy will often be intertwined with the source of the training data, as is the case with hiring decisions, algorithmic biases can blend into an already biased landscape.[133] This failure can be exacerbated when the algorithm operates "online," retraining on new results affected by its operation, hence creating "runaway feedback" problems in which the algorithm becomes progressively inaccurate as it relies on increasingly biased data.[134] As researchers have suggested, such problems can be characteristic of predictive policing algorithms: past over-policing can lead to progressively greater over-policing.[135]

Similar distortions can result from a faulty learning process. Even when the training set is sufficient and potentially representative, the learning algorithm can fail to model the training set in a useful manner. A familiar problem happens when the learner models the training set *too well*, "overfitting" the model to the training data, so that the model is not general enough to be of real-world use.[136] Overfitting, endemic to learning methods that create incredibly intricate models, produces faulty algorithms, not because the data is inherently under-representative, but instead because the training process assigns too much weight to irrelevant features in the training set.[137] To reuse the loan example, the mistake of assigning too much weight to the relative scarcity of minority applicants can also be described as a matter of overfitting the model to this

---

[132] *Id.* at 15. Recently, it has been suggested that this problem is further compounded by "underspecification." *See* Alexander D'Amour et al., *Underspecification Presents Challenges for Credibility in Modern Machine Learning* 2–3 (arXiv Working Paper No. 2011.03395), https://arXiv:2011.03395.

[133] *See, e.g.*, Tammy Wang, *How Machine Learning Will Shape the Future of Hiring*, LINKEDIN (Mar. 8, 2017), https://www.linkedin.com/pulse/how-machine-learning-shape-future-hiring-tammy-wang/ [https://perma.cc/2MNZ-9EWL].

[134] *See* Ensign et al., *supra* note 76.

[135] *See id.* at 10; ENGSTROM ET AL., *supra* note 30, at 25.

[136] *See* KELLEHER, *supra* note 59, at 21.

[137] *See id.*

incidental feature of the data, factoring in the applicants' membership in a minority group despite this detail's irrelevance to the function the learning process is meant to attain.[138] Overfitting problems can prove to be especially persistent, particularly when it comes to deep learning methods that operate by locating hidden patterns embedded in the data.[139] As a result, even when an algorithm is explicitly designed to ignore particular variables, such as participation in a protected class, deep learning algorithms are prone to fixate on the "noise" such features leave behind even in their absence and indirectly weigh them into the model.[140]

B. Algorithms and Interpretation

Distortions introduced through the use of auxiliary systems also affect how human decision-makers interpret the meaning of a legal norm. Several scholars have noted that reliance on algorithms can affect how decision-makers understand their determinations and the legal norms that guide those determinations. Ben Green and Yiling Chen, for instance, note that using risk assessment algorithms can make decision-makers overemphasize these algorithms' meaning.[141] Even when the algorithm's predictions are factually accurate, or especially when they are, reliance on these predictions can increase the salience of risk in decision-makers' overall decisions, affecting how the decision-makers interpret the very meaning of the legal norm that guides their decisions.[142]

Cary Coglianese and David Lehr discuss a similar algorithm-induced shift toward reliance on quantitative judgments, suggesting that this transformation can come to represent a substantial change in the law unaccompanied by political authorization.[143] As Coglianese and Lehr suggest, "the use of algorithms will often

---

[138] *See* Kroll et al., *supra* note 99, at 681; Lehr & Ohm, *supra* note 30, at 704.

[139] *See* Lehr & Ohm, *supra* note 30, at 704.

[140] *See* Lilian Edwards & Michael Veale, *Slave to the Algorithm: Why a Right to an Explanation Is Probably Not the Remedy You Are Looking for*, 16 DUKE L. & TECH. REV. 18, 28 (2017); Kroll et al., *supra* note 99, at 681.

[141] *See* Green & Chen, *supra* note 99, at 1.

[142] *Id.*

[143] *See* Coglianese & Lehr, *supra* note 47, at 1218.

compel agency decision makers to engage in quantitative coding of value judgments that have typically been made qualitatively."[144] Making this interpretive shift in the meaning of legal norms, as Coglianese and Lehr note, requires careful deliberation in light of its profound political effects.[145]

Richard Re and Alicia Solow-Niederman likewise argue that the efficiency of algorithmic adjudication can inspire a turn toward "codified justice," meaning an interpretation of legal norms that favors standardization over judicial discretion.[146] Machine learning, Re and Solow-Niederman suggest, introduces a new, correlation-based form of adjudication.[147] As this approach takes hold over lay and professional views of the law, the codified justice that accompanies this form of adjudication would gradually replace equitable notions of legal justice, creating a self-reinforcing cycle that continually pushes toward further codification.

In a similar fashion, Andrew Ferguson notes that reliance on the products of algorithmic data analysis can lead individuals to trust in their worst instincts with respect to the facts shaping the norm's meaning.[148] As discussed above, algorithmic biases can suggest that minorities are, for instance, more prone to being involved in crime and less likely to repay loans.[149] Even if these erroneous inferences are taken with a grain of salt, they can nevertheless reaffirm the human biases that led to the faulty algorithmic reasoning.[150]

Finally, Frank Pasquale and Glyn Cashwell demonstrate how machine learning algorithms push toward the emergence of an impoverished "jurisprudence of behaviorism."[151] As Pasquale and Cashwell suggest, given how machine learning operates, reliance on

---

[144] *Id.*

[145] *Id.*

[146] Richard M. Re & Alicia Solow-Niederman, *Developing Artificially Intelligent Justice*, 22 STAN. TECH. L. REV. 242, 246 (2019).

[147] *Id.*

[148] *See* Ferguson, *supra* note 39, at 402.

[149] *See* Lehr & Ohm, *supra* note 30, at 680.

[150] *See id.*

[151] *See* Pasquale & Cashwell, *supra* note 38, at 64–65.

machine learning systems, even in merely auxiliary functions, is prone to lead decision-makers to overemphasize the place of measurable data in their interpretation of the applicable norm.[152]

V. ASSISTIVE SYSTEMS AND LEGAL CHANGE

As suggested above, the effect that auxiliary systems can have on legal decisions and the meaning of legal norms can be sizable. Still, Solum and Volokh seem to suggest that this effect does not amount to the creation of legal meaning, as it does not generate new norms; these systems only affect how decision-makers implement and understand existing norms. In the thought experiments Solum and Volokh offer, the momentous step toward artificially intelligent legal meaning, with all the scrutiny this step demands, only occurs when algorithmic systems take over the task of adjudication—something that Solum and Volokh do not foresee happening in the near future.[153] In contrast, this Part suggests that meaningful algorithmic takeover can occur as early as the introduction of assistive machine learning systems—a development already underway. As will be shown, despite their limitations, these assistive systems' ability to emulate rudimentary legal analysis can have a negative effect on the path of legal development, as use of these systems can prevent the law's natural evolution.

A. Assistive Systems

Recent years have seen the rise of systems that advance from fact-finding and other auxiliary functions to limited legal analysis. Algorithmic systems have successfully made this transition, especially when assisting routine and repetitive decisions that take place within a narrow setting and follow a rule-bound structure.[154] These systems' ability to perform limited legal analysis has

---

[152] *Id.*

[153] *See* Solum, *supra* note 2, at 85; Solum, *supra* note 3, at 55; Volokh, *supra* note 3, at 1137.

[154] *See* Pasquale, *supra* note 32, at 29; Surden, *supra* note 35, at 1309.

propelled legal algorithms to the assistive stage, directly participating in the decision-making process.[155]

Legal algorithms are thus used for advanced legal research, e-discovery, and the evaluation of the strength of legal strategies.[156] Other functions increasingly delegated to algorithms include various triage responsibilities, such as processing, structuring, classifying, and generally filtering the information provided to decision-makers.[157] Thus, various notice and takedown procedures that demand responding to large numbers of complaints and targeting immense amounts of user content, routinely involve widespread automatic algorithmic decision-making, at least as an initial, often appealable, step.[158]

This change is most evident in social media content moderation. Particularly following the COVID-19 pandemic, platforms have begun implementing initial algorithmic screening systems to determine which potentially prohibited content is brought to the attention of human content moderators.[159] In March 2020, YouTube announced that it was implementing new measures in which "automated systems will start removing some content without human review," detecting "potentially harmful content and then send[ing] it to human reviewers for assessment."[160] In April of the

---

[155] *See generally* ENGSTROM ET AL., *supra* note 30 (discussing the use of artificial intelligence by federal administrative agencies).

[156] *See* Rich, *supra* note 39, at 872–73; Surden, *supra* note 35, at 1331; Surden, *supra* note 60, at 101.

[157] *See* ENGSTROM ET AL., *supra* note 30, at 22–23.

[158] *See* Perel & Elkin-Koren, *supra* note 89, at 183.

[159] *See* Elizabeth Dwoskin & Nitasha Tiku, *Facebook Sent Home Thousands of Human Moderators Due to the Coronavirus. Now the Algorithms Are in Charge*, WASH. POST (Mar. 24, 2020), https://www.washingtonpost.com/technology/2020/03/23/facebook-moderators-coronavirus/ [https://perma.cc/NJ7L-GM4B]; James Vincent, *Facebook is Now Using AI to Sort Content for Quicker Moderation*, THE VERGE (Nov. 13, 2020), https://www.theverge.com/2020/11/13/21562596/facebook-ai-moderation [https://perma.cc/4RHH-9M2R].

[160] *Protecting Our Extended Workforce and the Community*, YOUTUBE OFF. BLOG (Mar. 16, 2020), https://blog.youtube/news-and-events/protecting-our-extended-workforce-and [https://perma.cc/5AYS-AGQL].

same year, Twitter announced its implementation of algorithms trained on past moderation decisions to "surfac[e] content that's most likely to cause harm and should be reviewed first" and "proactively identify rule-breaking content before it's reported."[161] Likewise, Facebook has steadily increased its use of proactive filtering to identify materials that violate its community standards before these prohibited materials are reported.[162] Today, 97.6% of all hate speech violations on Facebook are proactively detected, with algorithms independently determining what speech falls under this classification.[163] Similarly, in the first quarter of 2021, YouTube reported using automated flagging to remove 9,091,315 videos, with only 478,326 removals originating from human sources.[164]

Outside of social media content moderation, another area in which algorithms have switched to legalistic measures of relevance has been in the prevention of child pornography. In 2018, Google announced that the company had developed an algorithm capable of autonomously identifying materials falling under the category of child pornography, which Google made freely available in the form of an Application Program Interface ("API") titled "Content Safety."[165] Google presents the Content Safety API as a screening

---

[161] Vijaya Gadde & Matt Derella, *An Update on Our Continuity Strategy During COVID-19*, TWITTER BLOG (Apr. 1, 2020), https://blog.twitter.com/en_us/topics/company/2020/An-update-on-our-continuity-strategy-during-COVID-19 [https://perma.cc/5J4A-LYPJ].

[162] *How Facebook Uses Super-Efficient AI Models to Detect Hate Speech*, FACEBOOK AI (Nov. 19, 2020), https://ai.facebook.com/blog/how-facebook-uses-super-efficient-ai-models-to-detect-hate-speech/ [https://perma.cc/PU7B-DUBU].

[163] *Hate Speech*, FACEBOOK TRANSPARENCY CTR., https://transparency.fb.com/data/community-standards-enforcement/hate-speech/facebook (last visited Oct. 3, 2021) [https://perma.cc/B7NG-2UHN].

[164] *YouTube Community Guidelines Enforcement*, GOOGLE TRANSPARENCY REP., https://transparencyreport.google.com/youtube-policy/removals?hl=en&total_removed_videos=period:2021Q1;exclude_automated:human_only&lu=total_removed_videos (last visited Oct. 3, 2021) [https://perma.cc/765H-B4XB].

[165] Nikola Todorovic & Abhi Chaudhuri, *Using AI to Help Organizations Detect and Report Child Sexual Abuse Material Online*, GOOGLE: THE KEYWORD (Sept. 3, 2018), https://www.blog.google/around-the-globe/google-europe/using-ai-help-organizations-detect-and-report-child-sexual-abuse-material-online/ [https://perma.cc/HTG4-3QHL].

tool to be used before any human evaluation of the material takes place, thus minimizing human contact with disturbing materials and scaling up human adjudication.[166] Although Google has not disclosed information on how its algorithm detects child sexual abuse, the company has suggested that the algorithm does so through the use of machine learning classifiers.[167] In 2021, Pornhub, responding to mounting public pressure sparked by a 2020 New York Times piece exposing its facilitation of illegal and exploitative materials,[168] announced its adoption of "Industry-Leading Measures for Verification, Moderation, and Detection," to be implemented across the properties of its parent company, MindGeek, which controls a significant portion of the online pornography production market.[169] These measures, the pornography colossus announced, will include proactive screening that involves manual human review and "a variety of automated detection technologies," including Google's Content Safety API.[170]

Although this degree of reliance on legalistic filtering is not yet dominant directly within legal proceedings, there is reason to believe that it is only a matter of time before algorithmic filtering expands from the virtual domain to legal decision-making, especially in routine adjudications where unassisted decision-making can result in intolerable backlogs.[171] Similarly, a

---

[166] GOOGLE, MEET THE CONTENT SAFETY API 1–2, https://static.googleusercontent.com/media/protectingchildren.google/en//static/pdf/content-safety-api.pdf (last visited Oct. 3, 2021) [https://perma.cc/4P2E-D5BV].

[167] Kristie Canegallo, *Our Efforts to Fight Child Sexual Abuse Online*, GOOGLE: THE KEYWORD (Feb. 24, 2021), https://blog.google/technology/safety-security/our-efforts-fight-child-sexual-abuse-online/ [https://perma.cc/BQ9Z-3WW8].

[168] Nicholas Kristof, *An Uplifting Update, on the Terrible World of Pornhub*, N.Y. TIMES (Dec. 9, 2020), https://www.nytimes.com/2020/12/09/opinion/pornhub-news-child-abuse.html [https://perma.cc/B69K-8QEU].

[169] *Pornhub Sets Standard for Safety and Security Policies Across Tech and Social Media; Announces Industry-Leading Measures for Verification, Moderation and Detection*, PORNHUB (Feb. 2, 2021), https://www.pornhub.com/press/show?id=2172 [https://perma.cc/5QFF-63CY].

[170] *Id.*

[171] *See* ENGSTROM ET AL., *supra* note 30, at 9–10.

collaboration between Stanford's Regulation, Evaluation, and Governance Lab and Carnegie Mellon's Language Technologies Institute is currently developing an algorithmic decision support system meant to assist the Board of Veterans Appeals in its mass adjudication of disability and veterans' benefits determinations.[172] In the most striking example thus far, the Brazilian judiciary is in the process of implementing machine learning triaging systems to assist in addressing the country's immense judicial backlog.[173]

Legal scholars have responded to these developments by pointing out that, despite their promise for improving the legal systems the algorithms aid, these algorithms are also prone to exacerbating existing problems and introducing new ones.[174] Special attention has been given to the potential opacity of algorithms,[175] the accountability deficit that algorithms can create,[176] their contribution to power inequalities inherent in the legal system,[177] the risk of implicit biases and discriminatory effects,[178] their stimulation of governmental overreach and endangerment of procedural safeguards,[179] and finally, the dehumanizing effect on those subjected to algorithmic decision-making.[180] In contrast, Solum and Volokh suggest that the legitimacy of such systems mainly hangs on

---

[172] Daniel E. Ho & Matthias Grabmair, *Toward a Decision Support System for Veterans Adjudication*, STAN. L. SCH., (Dec. 2, 2020, 12:45 PM), https://law.stanford.edu/event/codex-speaker-series-dan-ho/ [https://perma.cc/WLA7-GGXE].

[173] *See* KATIE BREHM ET AL., THE FUTURE OF AI IN THE BRAZILIAN JUDICIAL SYSTEM, https://www.sipa.columbia.edu/academics/capstone-projects/ai-driven-innovations-brazilian-judiciary [https://perma.cc/V3LN-YQYQ].

[174] *See generally* FRANK PASQUALE, THE BLACK BOX SOCIETY (Harv. Uni. Press 2015) (offering a thorough discussion of the challenges raised by decision-making algorithms).

[175] *See id.* at 6–7.

[176] *See* Kroll et al., *supra* note 99, at 638.

[177] *See* PASQUALE, *supra* note 174, at 3–4.

[178] *See* Solon Barocas & Andrew D. Selbst, *Big Data's Disparate Impact*, 104 CALIF. L. REV. 671, 673–74 (2016).

[179] *See* Danielle Keats Citron & Frank Pasquale, *The Scored Society: Due Process for Automated Predictions*, 89 WASH. L. REV. 1 (2014).

[180] *See* Kiel Brennan-Marquez, *"Plausible Cause": Explanatory Standards in the Age of Powerful Machines*, 70 VAND. L. REV. 1249, 1251–52 (2017).

comparing their capabilities to comparable human agents and, more importantly, that the legitimacy of such systems only comes into serious question as algorithms take over the decision-making process.[181]

B. Preventing Legal Change

Legal development, however, can be affected not just by the creation of new norms but also by the prevention of legal change. This prevention is precisely what is prone to occur as a result of relying on assistive systems that filter cases brought before legal decision-makers, limiting them to those cases that accord with current legal norms.[182]

As machine learning systems learn to emulate legal analysis, the systems do so in a very particular manner, eerily similar to Oliver Wendell Holmes's description of legal reasoning: the prediction of how decision-makers would decide on a given case—based on experience rather than logic.[183] Holmes couples this idea with his famous "bad man's" view of the law, which suggests that legal analysis should only be interested in the tangible legal consequences a rule would have in a specific case.[184] However, in bringing up the bad man's point of view, Holmes certainly did not mean that law is an immoral or draconian project—quite the contrary. Holmes meant that legal decision-makers will be best served by recognizing the inherent amorality of law's internal perspective, so that in making legal decisions, particularly those that involve policy considerations, legal decision-makers need to be well aware of the need to supplant the legal perspective with a view focused on the social

---

[181] *See* Solum, *supra* note 3, at 61–62; Volokh, *supra* note 3, at 1138.

[182] *See* Tarleton Gillespie, *Content Moderation, AI, and the Question of Scale*, 7 BIG DATA & SOC'Y 1, 1–5 (2020).

[183] *See* Oliver Wendell Holmes, *The Path of the Law*, 10 HARV. L. REV. 457, 476–77 (1897); *see also* Surden, *supra* note 35, at 1331.

[184] *See* Bryan Casey, *Amoral Machines, or: How Roboticists Can Learn to Stop Worrying and Love the Law*, 111 NW. U. L. REV. ONLINE 231 (2016); Davis, *supra* note 13, at 187–88.

circumstances in which the legal norm operates and the social purposes the law serves.[185]

However, given supervised machine learning's reliance on past training data to create the normative model that animates the decision-making algorithm, machine learning algorithms are inherently incapable of weighing considerations that lie outside of existing legal paradigms, tethering their future determinations to past ones. As Tarleton Gillespie puts it, "An effective tool may learn to make the same kinds of distinctions as before. But while consistency might sound like a good thing, these policies should actually adapt over time."[186] For Holmes, in order to ensure that existing legal paradigms are congruent with the social advantages the law aims to produce, adjudicators must always keep in mind the need to adapt law to changing social realities.[187] This determination of the path of the law is not only something adjudicators do while consciously altering legal precedents or actively creating norms, but is also an inevitable part of legal adjudication, whether made explicitly and positively, or negatively through omission; "the result of the often proclaimed judicial aversion to deal with such considerations is simply to leave the very ground and foundation of judgments inarticulate, and often unconscious."[188]

Likewise, reliance on path-dependent algorithmic systems to determine which cases are brought before human adjudicators means that such systems have a negative, hidden effect on legal development by hindering legal evolution. Although these systems' reproduction of past legal decisions may appear to leave law unaffected, their stability in fact has the hidden effect of determining law's progress by preventing its development. As Holmes keenly noted, the choice not to update the norm is as consequential as the choice to actively alter it, only less forthright.[189] Likewise, Robert Cover famously described positive law as inherently "jurispathic,"

---

[185] *See* E. Donald Elliott, *Holmes and Evolution: Legal Process as Artificial Intelligence*, 13 J. LEGAL STUDS. 113, 115 (1984).

[186] Gillespie, *supra* note 182, at 3–4 (citations omitted).

[187] *See* Elliott, *supra* note 185, at 115.

[188] *See* Holmes, *supra* note 183, at 467.

[189] *Id.*

as the choice to keep the law unaltered is the implicit choice to suppress an alternative legal narrative that could rejuvenate the law.[190] To prevent the law from becoming moribund, healthy legal development requires the continuous inclusion of "jurisgenerative" narratives external to the law's current stance.[191] By cementing dominant legal norms and preventing the inclusion of extralegal consideration, assistive systems therefore have a considerable, albeit negative, effect on legal development, as these assistive systems make it harder for future decisions to substantially divert from past ones.[192]

Legal progress is often contingent on realizing that existing legal categories are incapable of adequately responding to the behaviors currently outside the law's purview. The use of supervised machine learning classifiers generally entails taking the meaning of its classifications as granted, seeing them as fixed end results with which to match new cases.[193] This appearance of legal immutability can undercut decision-makers' ability to appreciate their role in shaping and updating the meaning of legal norms.[194] As social sentiments change through time, the algorithm's continued reliance on past decisions is likely to prevent legal definitions from keeping up with the times.[195]

This limitation is reminiscent of, but distinct from, the familiar problem of "concept drift" or the infamous runaway-feedback

---

[190] Robert M. Cover, *Foreword: Nomos and Narrative*, 97 HARV. L. REV 4, 40–42 (1983).

[191] *See id.*

[192] *See* Niva Elkin-Koren, *Contesting Algorithms: Restoring the Public Interest in Content Filtering by Artificial Intelligence*, 7 BIG DATA & SOC'Y 1, 7 (2020); Davis, *supra* note 13, at 189.

[193] *See* SIMONE BROWNE, DARK MATTERS 114 (2015).

[194] *See* Eileen Oak, *A Minority Report for Social Work? The Predictive Risk Model (PRM) and the Tuituia Assessment Framework in Addressing the Needs of New Zealand's Vulnerable Children*, 46 BRIT. J. SOC'Y WORK 1208, 1215 (2016).

[195] *See* Rich, *supra* note 39, at 897.

problem.[196] In these two types of malfunctions, a growing mismatch between the concepts that underlie the training set and the ground truth the training set purports to represent leads to increasing inaccuracy in the algorithm's operation.[197] Such distortions, however, speak to the algorithm's failure to accurately capture the meaning of the legal norm. In contrast, the problem here is that the algorithm *is* congruent with the relevant legal classification but does not prevent the classification's natural development by insulating the algorithm from societal changes.

VI. PREDICTION AND SUSPICION: THE CASE OF MANDATORY REPORTING

To demonstrate how reliance on assistive supervised machine learning systems can produce "artificial legal meaning" even with contemporary technology, this Article suggests a thought experiment much like Solum's and Volokh's, only much closer to present day, involving the use of auxiliary and assistive algorithmic systems to assist legal actors in reporting child maltreatment and responding to such reports.

The initial part of this thought experiment, namely the creation of norm-influencing auxiliary systems, is already in place today in Allegheny County, Pennsylvania.[198] The thought experiment then hypothesizes a move to the next stage: using supervised machine learning systems to assist healthcare professionals who are legally mandated to report suspected child maltreatment by flagging cases that involve reportable suspicion. Rather than a hypothetical set in some future time, sufficiently remote to allow society to calmly

---

[196] "Concept drift" and the runaway-feedback problem are the two frameworks through which the problem of legal evolution is commonly addressed. *See* Mulligan & Bamberger, *supra* note 102, at 739; Pasquale, *supra* note 32, at 52; Pasquale & Cashwell, *supra* note 38, at 64; Rich, *supra* note 39, at 884.

[197] *See* Schlimmer & Granger, *supra* note 124; Widmer & Kubat, *supra* note 124.

[198] *See* Stephanie Cuccaro-Alamin et al., *Risk Assessment and Decision Making in Child Protective Services: Predictive Risk Modeling in Context*, 79 CHILD. & YOUTH SERVS. REV. 291, 294 (2017); Dan Hurley, *Can an Algorithm Tell When Kids Are in Danger?*, N.Y. TIMES (Jan. 2, 2018), https://www.nytimes.com/2018/01/02/magazine/can-an-algorithm-tell-when-kids-are-in-danger.html [https://perma.cc/S35L-7DU9].

ponder its theoretical meaning, this example illustrates the immediacy of this discussion. As detailed below, mandatory reporting is already subject to extensive data collection and analysis efforts in collaboration between state and federal agencies. Admittedly, this effort is still far short of amassing the comprehensive data required for training supervised machine learning algorithms that could animate assistive systems of the kind discussed here. Nevertheless, the existing record-keeping efforts, the willingness and ability to make data accessible for analysis, and the availability of relatively straightforward evaluation criteria make it likely that this development is not far off.[199]

A. Mandatory Reporting

In 2019, the most recent year for which national aggregate data is publicly available, about 4.4 million suspected child maltreatment cases were reported to Child Protective Services ("CPS").[200] Of those reports, about 2.4 million cases met CPS screening criteria, and the cases of some 3.5 million children were serviced.[201] Approximately 656,000 of the children were classified by CPS as victims of abuse or neglect;[202] an estimated 1,840 children died as a result of reported maltreatment—more than the number of children who fell to cancer that year.[203]

---

[199] Indeed, in 2017 the U.S. Department of Health and Human Services stated its interest in exploring the use of predictive analytics. U.S. DEP'T OF HEALTH & HUM. SERVS., REPORT TO THE CONGRESS PRESENTING HHS'S RESPONSE TO THE RECOMMENDATIONS OF THE COMMISSION TO ELIMINATE CHILD ABUSE AND NEGLECT FATALITIES 10–11 (Sept. 2016), https://aspe.hhs.gov/system/files/pdf/208766/ResponseReport.pdf [https://perma.cc/3T3U-HDAE] [hereinafter 2016 REPORT TO CONGRESS].

[200] *See* U.S. DEP'T OF HEALTH & HUM. SERVS., ADMIN. FOR CHILD. & FAM., CHILD. BUREAU, CHILD MALTREATMENT 2019 7 (last visited June 3, 2021) https://www.acf.hhs.gov/sites/default/files/documents/cb/cm2019.pdf [https://perma.cc/J2HS-7TFZ] [hereinafter CHILD MALTREATMENT 2019].

[201] *See id.* at 6, 18.

[202] *See id.* at 20.

[203] *See* CHILD MALTREATMENT 2019, *supra* note 200, at 53; Hurley, *supra* note 198.

This horrible tally reflects not only a tragic reality but also a concerted effort to fight the child abuse plague through data collection and analysis.[204] Prodded by the Child Abuse Prevention and Treatment Act ("CAPTA"), first passed in 1974[205] and amended by the CAPTA Reauthorization Act of 2010,[206] all jurisdictions today mandate the reporting of suspected child maltreatment to CPS agencies.[207] Referrals to CPS are either "screened in"—where "reports" are created and the agency responds in some way—or "screened out" because the referrals either fail to meet CPS reporting criteria, are lacking in information, or are outside the jurisdiction of CPS.[208] In 1988, CAPTA was amended to direct the Secretary of Health and Human Services to create a national data collection and analysis program; this directive led to the establishment of the National Child Abuse and Neglect Data System ("NCANDS").[209] States voluntarily submit records of reported cases to the NCANDS, including completed reports and their findings.[210] CPS reports are supplemented by agency files containing aggregate data from agencies outside of CPS.[211] The collected data are analyzed and put into annual reports by the Children's Bureau in the Administration on Children, Youth and Families.[212] In addition, the National Data Archive on Child Abuse and Neglect was established in 1988 to house this data along with data from individual

---

[204] *See* B. L. FORTSON ET AL., PREVENTING CHILD ABUSE AND NEGLECT: A TECHNICAL PACKAGE FOR POLICY, NORM, AND PROGRAMMATIC ACTIVITIES 35 (2016), https://www.cdc.gov/violenceprevention/pdf/can-prevention-technical-package.pdf [https://perma.cc/UX9W-UZQL]; 2016 REPORT TO CONGRESS, *supra* note 199, at 9.

[205] Child Abuse Prevention and Treatment Act, Pub. L. No. 100-294, 102 Stat. 103 (codified as amended at 42 U.S.C. § 5102).

[206] CAPTA Reauthorization Act of 2010, Pub. L. No. 111-320, 124 Stat. 3459 (2010).

[207] *See* 42 U.S.C. § 5106a(b)(2)(B)(i).

[208] *See* CHILD MALTREATMENT 2019, *supra* note 200, at 6.

[209] 1988 Act to amend the Child Abuse and Treatment Act, Pub. L. No. 100-294, 102 Stat. 102 (1988) (codified as amended at 42 U.S.C. §§ 5101–5119(c)).

[210] *See* CHILD MALTREATMENT 2019, *supra* note 200, at 2.

[211] *See id.* at 3.

[212] *See id.* at viii.

researchers, prepare the data for research, and disseminate the prepared data to qualified researchers.[213]

The duty to report child maltreatment is today a familiar fixture of the legal landscape. First applicable to physicians, and later expanded to all healthcare professionals, the duty to report attaches to as many as forty professions that routinely come into contact with children or are otherwise likely to encounter information indicative of maltreatment, at times explicitly tied to the reporter's professional role.[214] Who exactly is obligated to report varies from state to state.[215] Physicians remain prominent in the reporting process, specified as mandatory reporters in forty-seven jurisdictions,[216] and often assess the medical significance of findings reported by others.[217] Other professions notably include social workers, educators, therapists, childcare providers, law enforcement officers, film and photograph processors, and computer technicians.[218] In approximately eighteen jurisdictions, *any* person who suspects child maltreatment is required to submit a report, in most cases supplementing professional obligations.[219] In 2019, various professionals submitted 68.6% of screened-in reports, of which, healthcare and mental health professionals were responsible for 17%.[220]

---

[213] *See id.* at 4; *All Datasets*, NAT'L DATA ARCHIVE ON CHILD ABUSE & NEGLECT (last visited Aug. 23, 2021), https://www.ndacan.acf.hhs.gov [https://perma.cc/8CLW-MMK5].

[214] *See* Alan Sussman, *Reporting Child Abuse: A Review of the Literature*, 8 FAM. L.Q. 245, 272 (1974) (discussing the original Act); Art Hinshaw, *Mediators as Mandatory Reporters of Child Abuse: Preserving Mediation's Core Value*, 34 FLA. ST. U. L. REV. 271, 289–291 (2007) (discussing the expansion of reporting duties).

[215] *Cf.* U.S. DEP'T OF HEALTH & HUM. SERVS., CHILD.'S BUREAU, CHILD WELFARE INFO. GATEWAY, MANDATORY REPS. OF CHILD ABUSE AND NEGLECT (2019) [hereinafter, MANDATORY REPORTERS] (discussing the variance in reporting requirements).

[216] *See id.* at 2.

[217] *See* Ellen Wright Clayton, *To Protect Children from Abuse and Neglect, Protect Physician Reporters*, 1 HOUS. J. HEALTH L. & POL'Y 133, 136–37 (2001).

[218] *See* MANDATORY REPORTERS, *supra* note 215, at 2.

[219] *See id.*

[220] *See* CHILD MALTREATMENT 2019, *supra* note 200, at 9.

The definition of maltreatment has expanded over time, and states differ in regard to what must be reported.[221] CAPTA defines "child abuse and neglect" as "at a minimum, any recent act or failure to act on the part of a parent or caretaker, which results in death, serious physical or emotional harm, sexual abuse or exploitation, or an act or failure to act which presents an imminent risk of serious harm."[222] While some state statutes explicitly enunciate what falls under these categories, statutes and regulations often seek to define maltreatment in broad terms in an effort to encourage reporting—although not all jurisdictions agree that this broad definition is a constructive approach, fearing that doing so could overburden CPS and needlessly create animosity between families and reporters.[223]

B. Auxiliary Reporting Systems

To aid CPS in responding to the massive number of complaints they receive, several jurisdictions have begun implementing algorithmic systems to assist screeners in determining the urgency of the complaint.[224] Although CPS decisions are contingent on referrals from reporters, it is often up to CPS—or other law enforcement agencies—to determine whether a case justifies state

---

[221] *See e.g.*, U.S. DEP'T OF HEALTH & HUM. SERVS., CHILD.'S BUREAU, CHILD WELFARE INFO. GATEWAY, WHAT IS CHILD ABUSE AND NEGLECT? RECOGNIZING THE SIGNS AND SYMPTOMS 2 (2019), https://www.childwelfare.gov/pubpdfs/whatiscan.pdf [https://perma.cc/5QS8-3Q4L] [hereinafter, WHAT IS CHILD ABUSE AND NEGLECT?]; Thomas L. Hafemeister, *Castles Made of Sand? Rediscovering Child Abuse and Society's Response*, 36 OHIO N. U. L. REV. 819, 850 (2010); Margaret H. Meriwether, *Child Abuse Reporting Laws: Time for a Change*, 20 FAM. L.Q. 141, 143 (1986).

[222] 42 U.S.C. § 5101.

[223] *See, e.g.*, Hafemeister, *supra* note 221, at 845 (discussing the disagreement on the scope of the duty); Hinshaw, *supra* note 214, at 286–87 (discussing the definition of maltreatment); Sarah H. Ramsey & Douglas E. Abrams, *A Primer on Child Abuse and Neglect Law*, 61 JUV. & FAM. CT. J. 1, 9 (2010) (suggesting that "[s]tatutes and regulations often define abuse and neglect broadly in an effort to effectuate their child protective purposes").

[224] *See* Sarah Valentine, *Impoverished Algorithms: Misguided Governments, Flawed Technologies, and Social Control*, 46 FORDHAM URB. L.J. 364, 380 (2019).

involvement.[225] In an attempt to balance the terrible risks of under-involvement and the costs of over-involvement, risk-prediction algorithms are increasingly employed to assist CPS in making these decisions.[226] Such systems use as input data various indicators that are presumably predictive of maltreatment to track an outcome variable translated to a risk predictor.[227] Of particular note is the Allegheny Family Screening Tool ("AFST"), implemented by the Department of Human Services in Allegheny County, Pennsylvania, in 2016.[228] The system uses information provided in the referral combined with additional information found in child welfare information systems to assign a risk score to the referral as a supplement to the human screener's evaluation.[229]

Although these systems are "merely" assessment tools meant to assist human screeners in evaluating the (factual) risk involved in complaints they receive, Erin Dalton, the leader of Allegheny County's data-analysis department, was not shy about the system's more ambitious goal, telling the New York Times that the system also aims to "change the mind-set of the screeners . . . . It's a very strong, dug-in culture. They want to focus on the immediate allegation, not the child's future risk a year or two down the line. They call it clinical decision-making. I call it someone's opinion."[230]

Indeed, as discussion on similar uses of auxiliary fact-finding systems illustrates, the system's design can significantly influence how human decision-makers implement and understand the law. As Virginia Eubanks illustrates, "the model is already subtly changing

---

[225] *See* CHILD MALTREATMENT 2019, *supra* note 200, at 6.

[226] *See* Valentine, *supra* note 224.

[227] *See* VIRGINIA EUBANKS, AUTOMATING INEQUALITY 137–38 (Picador 2019) (2018).

[228] *See* Hurley, *supra* note 198.

[229] *See* RHEMA VAITHIANATHAN ET AL., CTR. FOR SOC. DYNAMICS, DEVELOPING PREDICTIVE MODELS TO SUPPORT CHILD MALTREATMENT HOTLINE SCREENING DECISIONS: ALLEGHENY COUNTY METHODOLOGY AND IMPLEMENTATION, (2017), https://www.alleghenycountyanalytics.us/wp-content/uploads/2017/04/Developing-Predictive-Risk-Models-package-with-cover-1-to-post-1.pdf [https://perma.cc/5XSE-4WEB].

[230] Hurley, *supra* note 198.

how some intake screeners do their jobs."[231] As the system enjoys an aura of scientific objectivity, human screeners increasingly defer to the algorithm's judgment when it conflicts with the screener's judgment.[232] The algorithm, however, is reliant on past data, making it susceptible to familiar biases.[233] One such bias, Eubanks suggests, results from the increased exposure of lower-income families to the kind of data collected by state authorities and analyzed by the algorithm.[234] As a result, screeners' exercise of their legal roles, and their subsequent understanding of these roles, is shaped by the design choices that produced the algorithmic predictions, creating a likewise biased legal notion of risk.[235]

C. The Need for Algorithmic Assistance

Moving from auxiliary algorithms aiding screeners to assistive systems aiding mandated reports involves a change from factual, forward-facing assessment of risk to backward-looking legal analysis of the meaning of the suspected behavior.[236] Admittedly, the normative space between reporter and screener is sometimes minimal or nonexistent. Severe maltreatment cases clearly need to be reported and screened in by CPS. In borderline cases, though, the medical, legal, and protective questions can diverge. There are instances in which an injury is likely nonaccidental yet does not mandate reporting from a legal point of view, either because the harm is not "serious" or because the facts do not meet the threshold of reasonable suspicion.[237] Likewise, given the forward-facing nature of CPS services, there are cases in which reportable

---

[231] EUBANKS, *supra* note 227, at 141.

[232] *See id.* at 142.

[233] *See supra* Part IV(A).

[234] EUBANKS, *supra* note 227, at 157–62.

[235] *See id.* at 167–73.

[236] Furthermore, failing to recognize these distinct tasks can lead to resentment between reporters and CPS. *See* Benjamin H. Levi & Sharon G. Portwood, *Reasonable Suspicion of Child Abuse: Finding a Common Language*, 39 J.L. MED. & ETHICS 62, 62 (2011); Melton Strozier et al., *Experiences of Mandated Reporting Among Family Therapists,* 27 CONTEMP. FAM. THERAPY 177,186 (2005).

[237] *See* 42 U.S.C. § 5101 (defining "child abuse and neglect").

maltreatment does not justify protective intervention—the extreme case being the intentional killing of a child with no siblings.[238]

Assisting mandated reporters will therefore require a tool different from the one in use today—one that will help reporters make the medical determination concerning the injury's diagnosis and etiology *as well as* determine whether the circumstances mandate reporting under the law. The latter, as with any reasonable suspicion determination, requires a totality-of-the-circumstances approach that takes into account both medical considerations and the other considerations that go into CPS decisions.[239] Such systems will require machine learning algorithms capable of developing a model of "reportable suspicion" to accordingly classify new cases.

The ability to create such algorithms is inseparable from the urgent need that demands their creation. Today, the work of mandated reporters increasingly takes place in a world of big data, with every decision potentially informed by vast amounts of pertinent information. Even when suspicion results from a "small data" setting—e.g., a single visit to the doctor's office—the services rendered will often be assisted by machine learning algorithms that are the products of big data.[240] In this reality, the introduction of algorithmic systems to assist reporters seems to be merely a matter of time. Without algorithmic assistance, the breadth of data that can give rise to reportable suspicion can become so massive as to be unmanageable.[241] Medical histories can be used by trained professionals to routinely locate suspected abuse; however, going through a decade's worth of case histories in a large hospital in

---

[238] *See* CHILD MALTREATMENT 2019, *supra* note 200, at 53.

[239] *See, e.g.*, Rich, *supra* note 39, at 887.

[240] Indeed, what makes data "big" is not necessarily sheer size, but rather how the data lend themselves to big data usages. *See* MAYER-SCHÖNBERGER & CUKIER, *supra* note 37, at 5–7; Danah Boyd & Kate Crawford, *Critical Questions for Big Data*, 15 INFO. COMM. & SOC. 662, 663 (2012).

[241] *See, e.g.*, Desai & Kroll, *supra* note 65, at 50–51; R. GREGG DWYER ET AL., PROTECTING CHILDREN ONLINE: USING RESEARCH BASED ALGORITHMS TO PRIORITIZE LAW ENFORCEMENT INTERNET INVESTIGATIONS, TECHNICAL REPORT 5 (2016).

search of suspicious patterns would, even if it were humanly possible, undoubtedly be so time-consuming as to be impractical.[242]

Big data also opens the door to including previously untapped data sources that may shed additional light on a single decision, from statewide Child Welfare Information Systems to other public and commercial databases.[243] Again, consulting this precious information will require algorithmic assistance. Finally, big data can be used to train algorithms that can take over routine administrative and even clinical functions. As algorithms become better at emulating human healthcare professionals' work, algorithms offer increasingly appealing applications for handling stored information by analyzing and structuring patient data, so that the analyzed data are presented to physicians in a way that best meets their professional needs.[244] With such algorithms in use, information needed to form suspicion of maltreatment could be hidden away from mandated reporters, unless the assistive systems are explicitly designed to flag that information.

As medical diagnosis increasingly operates in a big data environment, the scattered information on which medical reporters often rely in determining whether a case is reportable gets placed behind a veil that can only be pierced by enlisting the help of machine learning systems. No doubt, using machine learning in this context can be a perilous quantum leap. The case may be, as Frank Pasquale and Glyn Cashwell write, that the efficiency machine learning offers cannot, in and of itself, justify the significant jurisprudential risks machine learning creates.[245] However, in the categories discussed above, a real risk exists that, without using machine learning, the duty to report will fall into desuetude—reporting will become delayed to the point that the purpose of reporting is effectively denied.[246] This consequence is already becoming a reality in other legal domains, and it seems safe to

---

[242] *See, e.g.*, Surden, *supra* note 35, at 1326.
[243] *See, e.g.*, Hurley, *supra* note 198.
[244] *See, e.g.*, Pasquale & Cashwell, *supra* note 38, at 65.
[245] *See id.* at 12.
[246] *See, e.g.*, Bamberger, *supra* note 33, at 673; Calo, *supra* note 76, at 415.

assume that given the grave implications of unreported suspicions, mandatory reporting will soon follow suit.[247]

Furthermore, medical and other "algorithmic professionals" are rapidly becoming a reality, providing users with services that have little to no human involvement.[248] In a remarkable development, Google, for instance, has recently announced the anticipated launch of an AI-powered dermatology tool meant to provide medical diagnosis of common skin conditions.[249] Naturally, such services would be particularly appealing to those who wish to keep suspicious information away from mandated reporters. Although such information may be divulged with the intention of concealment from human view, failing to expand the duty to report to include such cases—meaning, requiring that such algorithms report their suspicions to their human counterparts—would significantly diminish the scope of the duty.[250]

D. Assistive Reporting Systems

Recognizing this need, in the proposed thought experiment, a supervised machine learning system is introduced to assist medical professionals in exercising their reporting duties by flagging cases that may involve reportable suspicion. The discussion is limited to the handling of information *voluntarily* provided, albeit for medical reasons, and ignores the potential proactive use of algorithms to detect suspected abuse in other sources. The discussion is further limited to algorithms used *at most* to flag and triage suspected abuse cases so that suspected abuse can be brought to the attention of human mandated reporters. This limitation should not be inferred as

---

[247] *See, e.g.*, Jane R. Bambauer, *Dr. Robot*, 51 U.C. DAVIS L. REV. 383, 383 (2017); Clayton, *supra* note 217, at 146; ENGSTROM ET AL., *supra* note 30, at 11, 16–17, 19; Perel & Elkin-Koren, *supra* note 89, at 183.

[248] *See, e.g.*, Bambauer, *supra* note 247, at 386–87.

[249] Peggy Bui & Yuan Liu, *Using AI to Help Find Answers to Common Skin Conditions*, GOOGLE (May 18, 2021), https://blog.google/technology/health/ai-dermatology-preview-io-2021/ [https://perma.cc/F8FM-62GF].

[250] *See, e.g.*, Barbara Daly, *Willful Child Abuse and State Reporting Statutes,* 23 U. MIAMI L. REV. 283, 342 (1969) (discussing the need to reach cases that do not reach healthcare professionals).

suggesting that algorithms cannot or should not be used as independent reporters, or even as adjudicators, enforcers, or legislators—questions that have already invoked some scholarly debate.[251] Rather, this Article seeks to demonstrate that these algorithmic systems can effectively create legal meaning, even when the algorithm is used in a purely assistive function. For this reason, this thought experiment also assumes that the human decision is made *de novo*, disregarding the fact that the algorithm found reasonable suspicion.[252]

In the thought experiment, the supervised machine learning algorithm used to make these determinations would be trained on datasets comprising past case histories labeled according to whether the cases were reported or not. For the sake of argument, the experiment assumes that these past determinations have left copious information about the circumstances in which the decisions were made, including both pertinent and irrelevant information, information gathered from medical history files, legal proceedings, and any other available sources.

For the purpose of the discussion, the thought experiment assumes that the output variable is tethered to the classification of the facts of the case as mandating reporting, mapped on previous reporting decisions. This decision is not obvious but is the decision that would most reasonably be made. Already as has become evident in the case of the AFST, no clear indicator of actual maltreatment is readily available, making AFST rely instead on measuring re-referral of screened-out cases and placement of children in foster care as proxies for actual maltreatment.[253] Likewise, one could suggest that the right decision to model—meaning to use as the outcome variable—should be the CPS screening decision, with the intention of creating an algorithm that helps close the gap between reporting and screening in. However, as discussed above, doing so will unnecessarily eliminate the vital distinction between mandated

---

[251] *See* Bambauer, *supra* note 247, at 395; Calo, *supra* note 76, at 423; ENGSTROM ET AL., *supra* note 30, at 22. *Contra* Volokh, *supra* note 3, at 1159.

[252] *Cf.* Simmons, *supra* note 36, at 1086.

[253] *See* EUBANKS, *supra* note 227, at 143–44.

reporters' independent function and that of CPS.[254] Closing the gap between reporting and screening in would necessarily leave out cases of reportable maltreatment that do not fall within CPS's reporting criteria;[255] even though such cases would not lead to legal ramifications, reporting can nonetheless have critical social purposes. An appropriate model for reporting purposes will have to make explicit this distinction, as an expression of the distinct normative function of mandated reporters.[256]

However, choosing past reporting as the training process' output variable would entail that the system's decision of whether to bring any new case to the reporter's attention is tethered to the preexisting legal meaning of "reportable suspicion." No matter how multifaceted the input data that the algorithm weighs to make this determination is, any datapoint would only be appraised in light of its ability to affect the legal meaning of reporting, as this legal category existed at the time the training sets were created. Even though final reporting decisions would remain in human hands, what human decision-makers decide upon and see in making these decisions would be determined according to its relevance to past legal meaning.[257]

Healthcare professionals would likely be highly susceptible to this winnowing effect. The algorithm's success in correctly implementing the strict meaning of "reportable suspicion" is likely to cause these mandated reporters to at most question the algorithm's accuracy but not the basic premises of its operational model and especially not the legalistic output classification that drives it—meaning the determination of whether a case falls under

---

[254] *See* Robert Deisz et al., *Reasonable Cause: A Qualitative Study of Mandated Reporting*, 20 CHILD ABUSE & NEGLECT 275, 284 (1996).

[255] *See* Brett Drake, *Unraveling "Unsubstantiated"*, 1 CHILD MALTREATMENT 261 (1996).

[256] For more discussion on the importance of this choice, *see* Emily Keddell, *Decision-Making and Risk Prediction in Child Protection Systems*, 12 POL'Y Q. 46, 48 (2016).

[257] *See* Bamberger, *supra* note 33, at 676; Coglianese & Lehr, *supra* note 47, at 1218; Pasquale, *supra* note 32, at 11–12.

the preexisting category of "reportable suspicion."[258] Healthcare professionals, and mandated reporters more generally, are typically not trained lawyers; although different mandated reporters often participate in dedicated training to help those professionals determine what constitutes reportable suspicion, adequately trained legal algorithms theoretically could offer expert-level advice—far surpassing the minimal training reporters currently receive.[259] Reporting duties ensnare reporters in a complicated web of conflicting and ill-defined legal obligations.[260] As lay legal classifiers, mandated reporters are prone to relying on moral intuition and nonlegal considerations to make the legal portion of their decision,[261] but the algorithm's reasoning would more closely track liability rules and considerations shaped by legal findings on file and assisted by the legal expertise injected into the training process.[262] Furthermore, the holistic approach that is the hallmark of machine learning could more closely follow the totality-of-the-circumstances standard and the need to consider unintuitive exculpating evidence.[263]

For these reasons, the use of machine learning to assist in the legal classification of reportable suspicion will likely become highly influential in shaping best practices, thereby helping to shield reporters from liability.[264] Even beyond liability considerations, the duty to report frequently introduces an unwelcome conflict between healthcare professionals and their patients or clients; there is good reason to believe that reliance on algorithmic decision-making could

---

[258] *Cf.* EUBANKS, *supra* note 227, at 167–68.

[259] For a discussion on the need for training, *see* Levi & Portwood, *supra* note 236, at 64–65; Victor I. Vieth, *Unto the Third Generation: A Call to End Child Abuse in the United States Within 120 Years (Revised and Expanded)*, 28 HAMLINE J. PUB. L. & POL'Y 1, 21 (2006).

[260] *See* SETH C. KALICHMAN, MANDATED REPORTING OF SUSPECTED CHILD ABUSE 26–30 (2d ed. 1999); Levi & Portwood, *supra* note 236, at 65.

[261] *See* KALICHMAN, *supra* note 260, at 64; Levi & Brown, *infra* note 282; Levi & Portwood, *supra* note 236, at 64.

[262] *See* Casey, *supra* note 184.

[263] *See* Calo, *supra* note 76, at 421; Simmons, *supra* note 36, at 1076.

[264] *See* Ryan Abbott, *The Reasonable Computer: Disrupting the Paradigm of Tort Liability*, 86 GEO. WASH. L. REV. 1, 5 (2018).

be used to alleviate some of this tension by sharing the burden of reporting with the algorithm, even if the ultimate decision remains in human hands.[265]

The system's constriction of reporting decisions seems even more likely given the dualistic nature of mandatory reporting. Mandated reporters are commonly not law enforcement professionals tasked with locating and investigating suspected abuse;[266] rather, most mandated reporters are professionals whose lines of work make them likely to come into contact with incidental information that could form the basis for suspicion. While at times reporters are explicitly notified of the child's maltreatment, reportable suspicion is often based on indicators found in information not directly related to the reason for which the information was provided: the behavior of a patient or a family member,[267] a pattern of otherwise unrelated traumatic episodes, or a host of other indicators that trained medical professionals come to identify as suspicious.[268] Even when the information's legal significance is of little doubt, locating the information will often require inferences and probabilistic assessment, leaving much room for the algorithm's discretion—which the algorithm will exercise based mainly on the legalistic definition the algorithm drew from its experience with the training set.[269] As the human decision-maker's sole source of information, these legalistic definitions will serve as the basis for the human decision, likewise limiting decision-making to established legal categories.[270] As algorithms become increasingly responsible for determining what information is brought before reporters, fewer such incidental details, irrelevant to the reporter's

---

[265] *See* Strozier et al., *supra* note 235, at 186.

[266] Reporting by law enforcement officers represented only 19.1% of all reporting in 2019. *See* CHILD MALTREATMENT 2019, *supra* note 200, at 53.

[267] *See* WHAT IS CHILD ABUSE AND NEGLECT?, *supra* note 221, at 5–6.

[268] *See* Nouman & Alfandari, *supra* note 77, at 6; Emalee G. Flaherty et al., *Clinical Report—The Pediatrician's Role in Child Maltreatment Prevention*, 126 PEDIATRICS 833, 834 (2010).

[269] *See* Rich, *supra* note 39, at 897.

[270] *See* Bamberger, *supra* note 33, at 711–13; Pasquale, *supra* note 32, at 11–12.

main professional function, will come out into the open—unless an algorithm is designed to locate them.

E. Creating the Legal Meaning of Maltreatment

In the above thought experiment, the result of using an assistive system to make medical professionals aware of cases that fall under the legal category of reportable suspicion is that medical professionals only encounter "incidental" information—of the kind that can produce reportable suspicion—when these details are deemed relevant to the preexisting meaning of this legal category. The result of such constriction is the tethering of any future legal meaning reporters produce to the meaning that informed past decisions. The more that healthcare professionals' encounters with potential maltreatment are contingent on previous algorithmic classifications, the less these professionals become aware of forms of maltreatment that do not fall under this formal category.[271]

"Reportable maltreatment," however, is hardly a static notion; preventing its congruence with changing norms is no less consequential in its creation of legal meaning than positive adaptation. The definition of reportable maltreatment is an unsettled amalgamation of fact and law, leading some to view the definition as "inherently problematic and variable."[272] Physical abuse is most commonly understood to mean serious nonaccidental harm to a child by a person responsible for the child; the definition does not include physical disciplining, as long as the disciplining is reasonable and causes no bodily injury.[273] As Virginia Eubanks notes, this definition, even with the requirement that harm be "serious," still leaves much room for subjectivity on what constitutes maltreatment.[274] For example, "Is spanking abusive? Or

---

[271] *See* EUBANKS, *supra* note 227, at 141.

[272] Drake, *supra* note 255, at 265; *see also* Nouman & Alfandari, *supra* note 77, at 2 ("[T]he concept of child maltreatment has no single, accepted and detailed definition, rather, it is variable and depends on social and cultural circumstances.").

[273] *E.g.*, WHAT IS CHILD ABUSE AND NEGLECT?, *supra* note 221, at 3; *see, e.g.*, Meriwether, *supra* note 221, at 144.

[274] EUBANKS, *supra* note 227, at 130.

is the line drawn at striking a child with a closed hand? Is letting your children walk to a park down the block alone neglectful? Even if you can see them from the window?"[275] As Brett Drake further illustrates, the legal facet of the definition necessarily assumes that some intentional and non-disciplinary harm will not be regarded as maltreatment, either because the harm lacks probative value or because the harm is insufficiently injurious: "A parent may state openly that he or she has caused a given injury, but if that harm was sufficiently minor, the [mandated reporter] may be unable to substantiate physical abuse."[276]

Similar indeterminacy surrounds the question of when a report must be made—a matter that is of particular import to healthcare professionals.[277] Although state statutes differ in their wording, most statutes establish that the duty to report does not require the reporter to have knowledge of maltreatment; instead, it is enough that the reporter possess *reasonable suspicion*.[278] The reasonable suspicion standard indicates that reporting requires less than a firm belief but more than the mere possibility that the observed injuries resulted from maltreatment.[279] With little more than this vague definition to rely upon, reasonable suspicion has often been condemned as hopelessly indeterminate to the point of unconstitutionality.[280] As a result, the scope of the duty to report is extremely context-sensitive,[281] with studies showing that reporters radically differ in their understanding of reportable suspicion.[282] These studies found that such decisions are as informed by tacit professional intuition as

---

[275] *Id.*

[276] Drake, *supra* note 255, at 265; *see also* Levi & Portwood, *supra* note 236, at 63.

[277] *See, e.g.*, Meriwether, *supra* note 221, at 146.

[278] *See, e.g.*, Hinshaw, *supra* note 214, at 287–89.

[279] *See, e.g.*, Levi & Portwood, *supra* note 236, at 64.

[280] *See, e.g.*, *id.* at 63–64; Gail L. Zellman & Stephen Antler, *Mandated Reporters and CPS: A Study in Frustration*, PUB. WELFARE 30, 37 (1990).

[281] *See, e.g.*, KALICHMAN, *supra* note 260, at 65–92; Nouman & Alfandari, *supra* note 77, at 2.

[282] *See* Deisz et al., *supra* note 254, at 279; Benjamin H. Levi & Georgia Brown, *Reasonable Suspicion: A Study of Pennsylvania Pediatricians Regarding Child Abuse*, 116 PEDIATRICS 5, 5 (2005); Levi & Portwood, *supra* note 236, at 63.

such decisions are attributable to general principles.[283] This result is commonly seen as an unsatisfactory situation. Aligning with this viewpoint, Benjamin Levi and Sharon Portwood have argued that, "If reasonable suspicion is to entail more than the 'mere possibility' that a child was abused[,] . . . potential reporters need guidance on how likely abuse must be before reporting is required"—guidance that Levi and Portwood believe is found neither in available legal advice nor in CPS guidance.[284]

Levi and Portwood's arguments suggest that the meaning of reportable suspicion is continually evolving—shaped not just by its factual implementation and interpretation in specific cases, but also by the changing meaning of "seriousness" and the acceptability of physical disciplining, both concepts introduced into the legal meaning of reportable suspicion through reporters' decisions. As these decisions become reliant on the assistance of algorithmic systems and are constricted by their regressive legal analysis, these systems effectively create their own static version of maltreatment, superimposing their definition over the legal norm.

Even though the systems assisting mandated reporters in this hypothetical would not be creating original norms, the systems would nonetheless be determining the legal meaning of reportable suspicion by deciding whether reportable suspicion applies in given cases—shaping how reporters understand reportable suspicion—and, ultimately, by preventing the natural evolution of legal elements, such as acceptable physical disciplining. In preventing the law's development, algorithms will essentially be functioning as Holmes's inert judges, charting the path of the law by omission. Although this development seems less dramatic than the emergence of active norm-generation by algorithms, the jurisprudential concerns this development raises as to algorithms' legitimacy are no less troubling.

---

[283] *See, e.g.*, Nouman & Alfandari, *supra* note 77, at 6.
[284] Levi & Portwood, *supra* note 236, at 65 (emphasis omitted).

VII. CONCLUSION

By examining the hypothetical cases of an algorithmic system meant to assist mandated reporters in determining which cases require reporting of child maltreatment, this Article seeks to demonstrate the dangers of delaying normative deliberation over the legitimacy of using law-making algorithms to the time in which these systems are capable of *replacing* human adjudicators. As this hypothetical case demonstrates, focus on the moment of replacement hides the negative effect that reliance on assistive legal analyses is prone to have on the law's natural development. This effect, as this Article suggests, can amount to the "jurispathic" creation of legal meaning by constricting legal decision-makers' worldviews in a way that insulates these decision-makers from social changes and prevents the adoption of alternative legal narratives. Although less visible than the positive creation of new norms by autonomous decision-making systems, this effect is just as consequential, and unlike positive algorithmic law-making, this effect is just around the corner. Addressing this challenge will require not only understanding the inherent limits of machine learning but also seeing how legal decisions must be free—to some extent—from the constrictions of past decisions as decision-makers chart the future path of the law.