VOL 5 NO 2 | SPRING 2025

International Journal of Law, Ethics, and Technology

La Nouvelle Jeunesse

The International Journal of Law, Ethics, and Technology

cite as Int'l J. L. Ethics Tech.

ISSN: 2769-7150 (Online) | 2769-7142 (Print)

Publisher: La Nouvelle Jeunesse

Address: 655 15th Street NW, Washington, DC 20005

Date: June 14, 2025

THE INTERNATIONAL JOURNAL OF LAW, ETHICS, AND TECHNOLOGY assumes a paramount role as a dynamic and intellectually stimulating platform dedicated to the meticulous exploration of the intricate interplay between technology, ethics, and the law. As a distinguished peer reviewed publication, we aim to highlight emerging legal issues by prioritizing exceptional original scholarship that traverses diverse academic disciplines. In our unwavering pursuit of academic excellence, we actively foster contributions that exhibit profound depth and insightful analyses within doctrinal and critical frameworks. Moreover, we enthusiastically embrace interdisciplinary research endeavors that aim to unveil the multifaceted dimensions of the law, drawing upon the diverse perspectives offered by the social sciences and humanities. Rather than regarding law, ethics, and technology as distinct and isolated realms, our journal proudly stands as a nurturing ecosystem that fosters a dynamic and inclusive dialogue. Through a holistic amalgamation of these traditionally delineated fields, we strive relentlessly to engender a comprehensive understanding of the ever-evolving contemporary society we find ourselves in. With open arms and genuine enthusiasm, we sincerely invite scholars from every corner of the globe, urging them to contribute their invaluable knowledge and expertise to this vibrant and intellectually stimulating forum of global knowledge exchange.

SUBSCRIPTIONS: The print version of the *International Journal of Law, Ethics and Technology* is available to individuals and institutions as pre the approval by the editors. To request a place on the list, please email us at info@ijelt.org.

SUBMISSIONS: Please send articles, responses, letters to the editors, and anything else we ought to consider for publication to the *International Journal of Law, Ethics, and Technology* at submissions@ijlet.org.

CORRESPONDENCE: Please write to the *International Journal of Law, Ethics, and Technology* at info@ijlet.org.

Table of Contents

This page intentionally left blank

THE OTHER BRADY RIGHTS: HOW SOFTWARE CAN SOLVE THE BRADY RIGHTS CLASH BETWEEN DEFENDANTS AND POLICE OFFICERS

Joshua L. Herzberg[*]

Abstract: This Comment explores a critical, yet unexamined feature of Brady doctrine: police officers' rights in Brady disclosures. Prosecutors disclosing from officers' personnel files must consider and respect officers' procedural and substantive due process rights. Examining those rights also requires understanding the tools the government uses to carry out disclosure. So, this article examines the software applications that departments use to identify Giglio material and provide it to defendants, the first account of such applications. Many city police departments rely on IA Pro, an internal affairs software program. Others are just starting to shift to digital data collection, opening new avenues for compliance, tracking, and disclosure.

Keywords: Brady; Giglio

[*] United States District Court for the Southern District of New York, US.

Table of Contents

INTRODUCTION

In 2011, Philadelphia police officer Christopher Hulmes arrested Gilbert Narvaez for possession of narcotics with intent to distribute.[1] On Hulmes's word alone, Narvaez was convicted and sentenced to prison. But unbeknownst to Narvaez, the judge, and the jury, Hulmes had admitted to perjuring himself earlier that year in an unrelated case.[2] That information could have changed the entire outcome of the case. And by law, the prosecuting attorney should have disclosed that admission to Narvaez thanks to a seminal Supreme Court case from 1963.

Brady v. Maryland held that due process requires prosecutors to turn over certain information about witnesses to defendants, including information that impeaches a witness by casting doubt on his truthfulness.[3] When a police officer takes the stand to testify, *Brady* demands that the prosecutor disclose any information known to the officer or the prosecutor that could cause the jury to discredit the testimony. The defendant and his attorney can use that information—termed "*Giglio* material"—to cross-examine the officer about any time the officer lied, was charged or arrested, used excessive force, or exhibited bias or prejudice.[4] That sort of cross-examination can be devastating to the prosecutor's case.

Defendants need *Brady*: it safeguards their innocence by constraining prosecutors' worst adversarial instincts.[5] But flaws abound in its ability to protect their due process rights.[6] When it comes to police personnel files, webs of statutory provisions, protective orders, and misaligned incentives deny defendants "critical impeachment evidence to which they are entitled under *Brady* [and] harm society by undermining due process and by allowing dishonest officers to stay on the job."[7] In large part, fervent resistance from the police and their allies has kept prosecutors from complete *Brady* compliance.[8] That resistance has sometimes taken the form of legal action to protect officers. Defendants need *Brady*, but overcorrecting in favor of their interests risks violating the rights of police officers.

[1] Commonwealth v. Narvaez, J-S62013-13 1, 2 (Pa. Super. Ct. 2013).

[2] *See* Maura Ewing & Daniel Denvir *Cascade of Overturned Cases May Emerge in Wake of Philly DA's 'Bad Cop' List*, THE APPEAL (Apr. 9, 2018) https://theappeal.org/cascade-of-overturned-cases-may-emerge-in-wake-of-philly-das-bad-cop-list-27f8bcd4fc9a/.

[3] 373 U.S. 83 (1963); Kyles v. Whitley, 514 U.S. 419, 433 (1995) (holding that due process, as interpreted by Brady, requires disclosure of exculpatory material in the possession of the police).

[4] *See* Giglio v. United States, 405 U.S. 150, 154 (1972) (extending *Brady* disclosures to cover evidence that impeaches the testimony of a government witness); Rachel Moran, Brady *Lists*, 107 MINN. L. REV. 657, 701-04 (2022).

[5] U.S. v. Olsen, 737 F.3d 625, 630 (9th Cir. 2013) (Kozinski, J., dissenting) ("A robust and rigorously enforced *Brady* rule is imperative because all the incentives prosecutors confront encourage them not to discover or disclose exculpatory evidence.").

[6] *See* STEPHANOS BIBAS, Brady v. Maryland *in* CRIMINAL PROCEDURE STORIES 130 (Carol S. Steiker ed., Foundation Press 2006) ("Brady's ringing rhetoric of innocence, then, is in some ways a hollow promise. Far from transforming the adversarial system into a question for truth, it has merely tinkered at its margins."); Brady*'s Bunch of Flaws*, 67 WASH. & LEE L. REV. 1533, 1535 ("the doctrine as presently constituted may provide a disservice to the very concept of justice.").

[7] *See* Jonathan Abel, Brady*'s Blind Spot: Impeachment Evidence in Police Personnel Files and the Battle Splitting the Prosecution Team*, 67 STAN. L. REV. 743, 807 (2015).

[8] *Id.* at 779-89.

In 2017, to comply with *Brady*, then-Philadelphia District Attorney Seth Williams put together a list of officers into a "Do Not Call" list or "*Brady*" list.[9] That list comprised officers whose personnel files held reports and evidence that significantly diminished their credibility as witnesses. Williams classified sixty-six officers into three categories: "do not call" as a witness, "may use" but disclose misconduct to defendants, and "use without restriction" but be aware of misconduct.[10] None of the officers knew that he had listed them. Years later, Williams' successor and self-proclaimed "progressive prosecutor," Larry Krasner, found the list.

As DA, Krasner sought to develop more robust "procedures and regulations" for *Brady* compliance.[11] He created a Police Misconduct Database to digitally track reports and complaints of misconduct. He also filed contempt motions against the police department for failing to turn over evidence of misconduct in specific cases.[12] Then, under court order, Krasner released the list of sixty-six officers, and the Philadelphia Inquirer published it for the whole city to see.[13]

Brady tracking, disclosures, and lists, publicized or not, cause harm to officers. The Williams list, once published, damaged the officers' reputations in their community and limited, if not ended, their employment prospects—all before any of the officers had a chance to dispute the propriety of including them on the list or to challenge the veracity of the information that landed them there.[14] To protect the officers' rights and interests, Philadelphia's dominant police union, Lodge 5 of the Fraternal Order of Police (FOP), sued.[15] They alleged that Williams' list and Krasner's database infringed on the officers' constitutionally protected interests in their employment, reputation, and privacy without providing them due process.[16]

This Comment explores a critical, yet unexamined feature of *Brady* doctrine: police officers' rights in *Brady* disclosures. Prosecutors disclosing from officers' personnel files must consider and respect officers' procedural and substantive due

[9] Fraternal Ord. of Police Lodge No. 5 by McNesby v. City of Philadelphia, 267 A.3d 531, 535 (Pa. Commw. Ct. 2021).

[10] *See id.* at 535, 556, n.6.

[11] *See* Giglio v. United States, *supra* note 4, at 154; PHILADELPHIA DISTRICT ATTORNEY'S OFFICE, OVERTURNING CONVICTIONS— AND AN ERA: CONVICTION INTEGRITY UNIT REPORT JANUARY 2018—JUNE 2021 (2021) [hereinafter DAO REPORT]. https://github.com/phillydao/phillydao-public-data/blob/25293f7340f98f6007e944a4a3943bac71faa99a/docs/reports/Philadelphia%20CIU%20Report%202018%20-%202021.pdf.

[12] *See* William Bender, *DA Krasner hits Police Department with contempt motions over not sharing misconduct data*, THE PHILADELPHIA INQUIRER (Aug. 11, 2021) https://www.inquirer.com/news/larry-krasner-philadelphia-police-misconduct-contempt-database-20210811.html.

[13] *See* Maura Ewing and Daniel Denvir *Cascade of Overturned Cases May Emerge in Wake of Philly DA's 'Bad Cop' List*, THE APPEAL (Apr. 9, 2018) https://theappeal.org/cascade-of-overturned-cases-may-emerge-in-wake-of-philly-das-bad-cop-list-27f8bcd4fc9a/; Mark Fazlollah, et al., *The full list of Philadelphia's 66 problem cops*, (Mar. 16, 2018) https://www.inquirer.com/crime/inq/full-list-philadelphias-66-problem-cops-20180316.html (publishing the officers identified by former DA Williams, released by DA Krasner).

[14] The Commonwealth Court treated eleven of the sixty-six officers as "exonerated," making their claim to injustice all that much stronger. Judge Ceisler, concurring in part and dissenting in part, argued that each of the eleven situations was more complicated. *See* Fraternal Ord. of Police Lodge No. 5 by McNesby, *supra* note 9, at 558, n.8.

[15] *Id.* at 535.

[16] Second Amended Complaint, Fraternal Order of Police, Lodge No. 5 v. City of Philadelphia, No. 01465, 2019 WL 7816597 (Pa.Com.Pl. Nov. 18, 2019).

process rights. Examining those rights also requires understanding the tools the government uses to carry out disclosure. So, this article examines the software applications that departments use to identify *Giglio* material and provide it to defendants, the first account of such applications. Many city police departments rely on IA Pro, an internal affairs software program.[17] Others are just starting to shift to digital data collection, opening new avenues for compliance, tracking, and disclosure.

Police unions tend to resist technological changes like these, foreseeing threats to their livelihoods. This is unfortunate—good software development and practice could protect officers' interests just much as bad development and practice can hurt them. If prosecutors and police departments build and adopt software with the right features, they can better comply with *Brady*'s demands for defendants while protecting officers' rights in the process.

This Comment discusses how to do that and proceeds in four parts. Part I explains the *Brady* rule with respect to police personnel files. Part II reviews how police departments use software to track misconduct and comply (or fail to comply) with *Brady*. Part III examines the FOP's lawsuit to understand police officers' due process rights in the face of *Brady* lists and disclosures. Finally, Part IV argues for two categories of improvements to software development and practice—those that facilitate *Brady* compliance and those that protect officers' rights.

I. *BRADY* AND POLICE PERSONNEL FILES

A. The *Brady* Rule

In 1963, the Supreme Court decided *Brady v. Maryland*, announcing that prosecutors violate a criminal defendant's right to due process when they suppress information favorable to his defense.[18] This landmark decision requires prosecutors in local and federal courts to disclose "favorable" information, "material either to guilt or to punishment."[19] *Brady* initially required only that prosecutors turn over information "upon request," not that they turn over information unprompted.[20]

Eight years later, in *Giglio v. United States*, the Court extended prosecutors' disclosure duty to cover information affecting the credibility of government witnesses, also known as impeaching information.[21] In *Giglio*, the Court emphasized that "the good or bad faith of the prosecution" is irrelevant to whether the prosecutor violated the defendant's due process rights.[22] In other words, the government need not hide the information to cause a *Brady* violation; they need only fail to disclose it. In *Bagley*, the

[17] *See* Part II.C, *infra*.
[18] *See* Brady v. Maryland, *supra* note 3; United States v. Bagley, 473 U.S. 667, 676 (1985) ("The jury's estimate of the truthfulness and reliability of a given witness may well be determinative of guilt or innocence, and it is upon such subtle factors as the possible interest of the witness in testifying falsely that a defendant's life or liberty may depend" (citing Napue v. Illinois, 360 U.S. 264, 269 (1959)).
[19] *Id.* at 87.
[20] *Id.* at 85.
[21] *See* Giglio v. United States, *supra* note 4, at 154 ("When the 'reliability of a given witness may well be determinative of guilt or innocence,' nondisclosure of evidence affecting credibility falls within this general rule.") (quoting Napue v. People of State of Ill., 79 S. Ct. 1173, 1177 (1959)).
[22] *Id.* at 153-54.

Court held that prosecutors must turn evidence over even if defendants do not ask for it.[23]

Brady and *Giglio* do not require police and prosecutors to search for or collect evidence on behalf of the defendant.[24] But the defense cannot know what information the prosecution has not disclosed, so the "prosecutor has a duty to learn of any favorable evidence known to others acting on the government's behalf in the case, including the police."[25] In *Kyles v. Whitley*, the Court specifically held that a *Brady* violation occurs when the prosecutor fails to disclose evidence known to a police officer.[26] On top of that, the prosecutor's office must establish "procedures and regulations" to carry out their disclosure duties.[27] The Court repeated the holding in *Strickler v. Greene* and *Youngblood v. West Virginia*: prosecutors must disclose to defendants even evidence "known only to police investigators."[28]

To warrant vacatur of a conviction, a *Brady* violation must satisfy three elements.[29] (1) The evidence must be favorable to the defendant as exculpatory or impeaching. (2) The state must have suppressed the evidence, "either willfully or inadvertently." And (3) the evidence must be material: it must have prejudiced the result.[30] When truly exculpatory evidence appears, prosecutors may decline to pursue the case, knowing they cannot convince a jury beyond a reasonable doubt. In this way, *Brady* encourages fairer trials because prosecutors choose not to prosecute rather than to suppress exculpatory information. Yet, between prosecutors' adversarial trial mindset and the difficulties of evaluating materiality as a case develops, *Brady* makes a tall ask of prosecutors.[31]

Impeaching evidence reduces the jury's estimate of the credibility of a witness's testimony.[32] That can take the form of evidence that the witness could not properly observe the events, prior acts of untruthfulness of the witness, prior non-lying acts that

[23] United States v. Bagley, *supra* note 18.

[24] *See* Kyles v. Whitley, *supra* note 3, at 437 (1995).

[25] *Id.*

[26] *Id.* at 438 (The state of Louisiana "suggested below that it should not be held accountable . . . for evidence known only to police investigators and not to the prosecutor. [This] would amount to a serious change of course from the *Brady* line of cases.").

[27] Giglio v. United States, *supra* note 4, at 154 (laying responsibility on the prosecutor for nondisclosure as "spokesman for the Government."); *See also* Angela J. Davis, *The Legal Profession's Failure to Discipline Unethical Prosecutors*, 36 HOFSTRA L. REV. 275, 287 n.59 (2007) (distinguishing the prosecutor's duty to disclose not their entire file, but any evidence material and favorable to the defendant).

[28] Strickler v. Greene, 527 U.S. 263, 280-81 (1999); Youngblood v. W. Virginia, 547 U.S. 867, 869 (2006).

[29] *Id.* at 281-82.

[30] *Id.*

[31] United States v. Bagley, *supra* note 18, at 3379-80 (*Brady*'s "purpose is not to displace the adversary system . . . , but to ensure that a miscarriage of justice does not occur."); Edward P. Stringham, *Prosecutors Are Rewarded for Convictions, Not Justice*, INDEP. INST. (May 22, 2007) ("our highly politicized legal system, which rewards law enforcement officials for high conviction rates, rather than meting out justice") https://www.independent.org/news/article.asp?id=2024; R. Michael Cassidy, *Plea Bargaining, Discovery, and the Intractable Problem of Impeachment Disclosures*, 64 VAND. L. REV. 1429, 1436-40 (2011) ("appellate courts assessing post-conviction claims of undisclosed impeachment evidence struggle with the materiality issue and often produce split opinions.").

[32] *See, e.g.*, FED. R. EVID. 607.

suggest untruthfulness, or evidence of bias.[33] For example, in *Kyles*, the State relied on eyewitness testimony, but suppressed impeaching evidence that the eyewitnesses' testimony evolved over time.[34] Had the prosecutor disclosed that fact, the defense counsel could have introduced it, making "a different result reasonably probable."[35]

B. Police Personnel Files

As the state's main investigators, police officers "testify again and again as part of their jobs."[36] They also often act as accusers in criminal trials, such as when a citizen assaults an officer or when the officer identifies contraband in someone's possession.[37] Both when officers act as investigators and when they accuse, defense counsel cross-examines them and introduces impeaching evidence to cast doubt on their truthful character and the veracity of their testimony. Evidence of that character is often found within a police officer's personnel file where only the police officer and supervisors know about it.

Although personnel files contain information unrelated to the criminal prosecution at hand, their contents still fit squarely within the *Brady* rule as "favorable," "material," and known to someone acting on the government's behalf because they can decrease the jury's credence in the officer's testimony.[38] That is, police personnel files are full of potential *Giglio* material. Such information is especially powerful in he-said, she-said situations where the only witnesses to events are the officer and the defendant. The impeaching value of evidence often depends on context, and busy, rushed prosecutors don't review every officers' file for every case they try, so they rely on officers to alert them to impeaching information within their own files.

That includes evidence of dishonesty like "instances in which the officer lied under oath, filed a false report, covered up misconduct by another officer, cheated, or stole,"[39] but also less direct evidence of untruthful character like "internal affairs reports, disciplinary write-ups, and performance evaluations."[40] Criminal convictions of the officer provide valuable impeaching evidence because a court establishes the truth of the events. Unfortunately, many police departments continue to employ officers with convictions for planting drugs on suspects, bribing witnesses or informants, and perjury. All these can be used to impeach the witness officers.[41] Rules of evidence even allow pending charges to impeach witnesses.[42]

[33] *Impeachment Evidence*, BLACK'S LAW DICTIONARY (11th ed. 2019); *See, e.g.*, FED. R. EVID. 608, 609; RACHEL HARMON, THE LAW OF THE POLICE 200 (2021) (listing examples of impeachment evidence).

[34] *See* Kyles v. Whitley, *supra* note 3, at 443.

[35] *Id.* at 441.

[36] HARMON, *supra* note 33, at 226; JONATHAN SUAREZ, POLICE OFFICER EXAM 50 (2d ed. 2003) (police officers often testify in court as part of their job).

[37] Rachel Moran, *Contesting Police Credibility*, 93 WASH. L. REV. 1339, 1341-42 (2018).

[38] *See* Abel, *supra* note 7, at 748.

[39] *See* Harmon, *supra* note 33, at 224 (examining the doctrine of *Brady* disclosure of police personnel files).

[40] *See* Abel, *supra* note 7, at 745.

[41] *See* Moran, Brady *Lists supra* note 4, at 668; FED. R. EVID. 609.

[42] *Id.*; FED. R. EVID. 609(e) ("conviction that satisfies this rule is admissible even if an appeal is pending.").

Judicial findings of dishonesty can impeach officers by demonstrating their untruthful character.[43] Rachel Moran highlights an Arizona police officer who lied under oath at least four times and two Colorado officers who presented testimony inconsistent with body camera footage.[44] In fact, "[t]he problem of police officers lying in police reports or during testimony is so common that it has its own well-known euphemism, 'testilying.'"[45]

Internal affairs departments also conduct investigations and make findings that stay in officers' files.[46] Internal findings are particularly valuable "because they are the police department's own assessment of the officer's credibility."[47] In one egregious case, a police officer questioned a mother about the murder of her child.[48] A man had already confessed that he and another man had murdered the boy.[49] Other than the officer's testimony, "[t]here were no other witnesses or direct evidence linking [the mother] to the crime,"[50] yet she sat on death row for twenty two years.[51] But when she was tried, the judge and jury did not know about the officer's "long history of lying under oath and other misconduct."[52] He had made sexually inappropriate advances toward a woman during a traffic stop and lied about it under oath. The internal report wrote "[Y]our image of honesty, competency, and overall reliability must be questioned."[53] He had even "lied under oath in order to secure a conviction." After twenty years, Judge Kozinski granted the mother habeas relief, releasing her.[54]

Beyond direct evidence of dishonesty, internal affairs departments track officers' uses of force, stops, firearm discharges, pursuits, and collisions.[55] Individual incidents are part of officers' normal policing duties and are unlikely to amount to *Brady* material. But in the aggregate, patterns may emerge that reveal problematic behavior or racial biases. Internal affairs departments also collect and review citizen complaints. Unfortunately, departments rarely conduct thorough investigations or hold officers accountable for misconduct.[56]

[43] FED. R. EVID. 608.
[44] *See* Moran, Brady *Lists supra* note 4, at 669; *See, e.g.*, Oliver v. Flahive, No. 20-CV-1129-JPS-JPS, 2022 WL 1576686 (E.D. Wis. May 2, 2022), *appeal dismissed,* No. 22-1922, 2022 WL 17176378 (7th Cir. July 27, 2022).
[45] Moran, Brady *Lists*, *supra* note 4, at 669, citing Christopher Slobogin, *Testilying: Police Perjury and What to Do About It*, 67 COLO. L. REV. 1037, 1040 (1996).
[46] *See* Moran, Brady *Lists supra* note 4, at 670-71 (collecting examples).
[47] *See* Abel, *supra* note 7, at 746.
[48] *See* Milke v. Ryan, 711 F.3d 998, 1000-01 (9th Cir. 2013).
[49] *Id.*
[50] *Id.* Neither of the two men testified against her. Both insisted for years that the boy's mother had nothing to do with the murder. *Id.* at 1022 (Kozinski, J., concurring).
[51] *Id.*
[52] *Id.*
[53] *Id.* at 1012.
[54] *Id.* at 1019.
[55] *See, e.g.*, CI TECHNOLOGIES, INC., IAPRO INSTRUCTIONAL MANUAL 68, 141-42 (Mar. 2019) (available at https://wokewindows-data.s3.amazonaws.com/iapro/somerville/IAProManual.pdf).
[56] *See* Rachel Moran, *Ending the Internal Affairs Farce*, 64 BUFF. L. REV. 837, 844 (2016) ("Saying internal affairs units are the best means of protecting citizens from police misconduct is like saying foxes are the best guards for the henhouse.").

II. DISCLOSURE TOOLS

A. *Brady* Lists

The simplest of tools to carry out *Brady* disclosure is the *Brady* List. In pursuit of *Kyles* and *Giglio*'s required "procedures and regulations," many district attorneys' offices create a list of officers to aid their *Brady* compliance. *Brady* lists are "lists some prosecutors maintain of law enforcement officers with histories of misconduct that could impact the officers' credibility in criminal cases."[57] Jonathan Abel describes the lists, also known as "Do Not Call" lists, as "the mechanism by which prosecutors within an office alert each other to an officer's credibility problems."[58] Once alerted, prosecutors preparing for trial can choose to disclose the officer's *Giglio* evidence or avoid calling the officer as a witness at all. *Brady* lists are uncommon but impactful. They can facilitate disclosure for prosecutors, save defendants from testimony by dishonest officers, and impede or even end the careers of officers who find themselves on the list.

Although they are "highly controversial," very little law governs Brady lists in practice.[59] State laws requiring prosecutor offices to maintain them are rare, there is no federal law mandating their creation, and there are "inconsistent prosecutorial practices around maintaining and using such lists."[60] As one police association leader noted, "there appears to be no set standard for placing an officer on the list, removing an officer from the list, or . . . defining [who] makes those decisions."[61] In a few localities, the prosecutor must overcome a burden of proof before adding an officer to the list. For example, some offices require "substantial information" before including an officer on a list.[62] However, most prosecutors create *Brady* lists without articulating the decision rules they apply to establish which officers they include on the list. Nor do they communicate policies for removing officers from lists. Insofar as *Brady* lists help prosecutors avoid relying on officers with damning personnel files as witnesses, clear decision rules make it easier to weigh the costs and benefits of calling a specific officer. And lists without removal processes can grow stale to the point that prosecutors cannot trust their accuracy or usefulness.

Offices that keep active *Brady* lists often assemble committees that review officers' files to determine whom to include on their *Brady* list.[63] As Rachel Moran notes, many of these committees trust law enforcement agencies to share the relevant information about their officers. Not only does this mean that *Brady* list committees rely on parties to share information who have incentives to conceal that information, but police departments also don't have the legal expertise to discern what evidence

[57] *See* Moran, Brady *Lists supra* note 4, at 658. These lists have many other names including *Giglio* lists. *See id.* at 658 n. 2.

[58] *See* Abel, *supra* note 7, at 780, n. 198 (citing CAL. GOV'T CODE § 3305.5 (West 2014)).

[59] *See* Moran, Brady *Lists supra* note 4, at 660-63.

[60] *Id.*

[61] *See* Abel, *supra* note 7, at 780, (quoting Jim Parks, former "president of Arizona's largest police association.").

[62] *See* Moran, Brady *Lists supra* note 4, at 707.

[63] *See id* at 696-701 (reviewing the mechanics of how officers get on *Brady* lists, but leaving out any descriptions of decision rules that *Brady* list committees apply).

merits *Brady* disclosure. That has resulted in major failures of prosecutors to comply with *Brady* mandates in Texas, Colorado, Oregon, and elsewhere.[64]

In most localities that maintain *Brady* lists, the prosecutor's office keeps the list internally and limits outside access.[65] Not everywhere, though. Middlesex County, Massachusetts, maintains a publicly available *Brady* list but has faced significant criticism for doing so,[66] and in Philadelphia, the DA's Office faced the lawsuit demanding disclosure of Williams' *Brady* list. Likewise, in the Bronx, a Freedom of Information Act (FOIA) request exposed the DA Office's *Brady* list.[67] The DA then released the list with "heavy redactions" to protect officer privacy.

Citizens sometimes create and maintain misconduct databases analogous to *Brady* lists. These don't help prosecutors avoid certain officers like traditional lists do. Instead, they provide accountability to the community and assist defense attorneys protecting their clients against accusations from certain police officers.[68]

B. Software Systems in Use

The academic literature on *Brady* disclosure has taken no critical review of the software systems that police departments use to collect, analyze, and disclose officer personnel files, creating a major gap in our understanding of the practice. Rachel Moran offers a helpful jumping off point by reviewing the ways that the law allows data collection systems and rules of evidence to prevent defendants from accessing information about accusing police officers.[69] She notes that no national database exists to collect officer misconduct information except for the National Decertification Index (NDI).[70]

Funded by the justice department, the NDI is "a national registry of certificate or license revocation actions relating to officer misconduct," but it relies on agencies to self-report officers.[71] The International Association of Directors of Law Enforcement Standards and Training maintains the NDI, and limits use to law enforcement agencies and Peace Officer Standards and Training (POST) organizations.[72] They restrict access to individuals to cases of legitimate need, which excludes defense attorneys fishing for impeaching information about an officer witness.[73] In other words, criminal defendants cannot rely on the NDI for *Brady* disclosures. Moreover, the NDI lists officers with

[64] *Id.* at 698-99.
[65] *Id.* at 709 (listing the lengths to which prosecutors go to keep their lists secret in New Hampshire, California, Utah, and D.C.) Some offices release the lists to comply with open access laws. *Id.* at 710.
[66] *Id.*
[67] George Joseph, *Bronx Prosecutors Release Secret Records on Dishonest Cops*, Gothamist (Oct. 7, 2019), https://gothamist.com/news/bronx-prosecutors-release-secret-records-dishonest-cops.
[68] *See, e.g.,* The Legal Aid Society, *The Cop Accountability Project*, (last visited Apr. 15, 2023) https://legalaidnyc.org/programs-projects-units/the-cop-accountability-project/; Nathan Story and Jacob Lurye, *The Woke Windows Project* (last visited Apr. 15, 2023) https://www.wokewindows.org/.
[69] Moran, *Contesting Police Credibility supra* note 37, at 1360-64.
[70] *Id.*
[71] *National Decertification Index*, INT'L ASS'N OF DIRS. OF LAW ENF'T STANDARDS & TRAINING, https://www.iadlest.org/about.
[72] *See* Moran, *Contesting Police Credibility supra* note 37, at 1362 n.132 and accompanying text.
[73] *Id.* Only one case on Westlaw mentions the NDI. Cmty. Coll. of Rhode Island v. CCRI Educ. Support Pro. Ass'n/NEARI, 184 A.3d 220, 223 (R.I. 2018).

revoked certifications or licenses. Police departments don't employ those officers, and they are unlikely to stand witness at trial.

Beyond the NDI, individual police departments use software to track active officers' files. Many use a software program called "IA Pro" to track and manage misconduct complaints and personnel files.[74] Almost one thousand American police departments in forty-eight states use IA Pro, including departments and agencies in New York City, Chicago, Los Angeles, Miami, Philadelphia, and the federal government.[75] Despite its prevalence, almost no caselaw and even fewer secondary sources discuss the software.[76] In one case that does, the City of Philadelphia explained how it uses the software:

> [T]he City asserts that as part of its investigatory and disciplinary process, the [Philadelphia Police Department] maintains an Internal Affairs Case Management System ("IAPro") that was initiated in 2002. IAPro can display officers by name and complaint summaries, and identify documents associated with an investigation. According to the City, IAPro provides officer "alerts" with regard to complaints against a particular officer for use of force, IAD intake information, discharge of firearm investigations, statistical reports and graphs, and Police Board of Inquiry case processing (including notification to chains of command and the tracking of disciplinary case outcomes).[77]

In short, IA Pro collects and organizes reports concerning officers. It also provides a platform for civilians to submit complaints, although some departments restrict access to internal use only. IA Pro is capable of generating and tracking disclosures, as well as efficiently purging officer records.[78]

Some departments rely on IA Pro's "Personnel Early Warning System,"[79] referred to in the user manual as "Early Intervention." IA Pro offers early warning

[74] *See* IA PRO https://www.iapro.com/pages/united-states-of-america (last visited Mar. 22, 2023).

[75] *See Clients*, IA PRO https://www.iapro.com/clients (last visited Mar. 22, 2023).

[76] The author ran a Westlaw search on Mar. 23, 2023, for "adv: ("IA Pro" or "IAPro") % "IA, Pro"" which returned nine cases and one secondary source. IA Pro also alerts internal affairs when numerous complaints accumulate against an officer through its "Personnel Early Warning System", a related function to *Brady* disclosure. A Westlaw search for "Personnel Early Warning System" returned six other cases and the same secondary source. Besides the quoted E.D.P.A. and N.D. Ga. discussions, each mention was brief and not central to the case.

[77] Lyons v. City of Philadelphia, No. CIV.A.06-5195, 2007 WL 3018945 at *3 (E.D. Pa. Oct. 12, 2007) (internal citations removed). The City of Lakewood, Washington uses IA Pro to digitally collect files that they keep in hardcopy in a locked closet down a secure hallway. Martin v. City of Lakewood, 21 Wash. App. 2d 1067 at *2 (2022).

[78] *See* CI TECHNOLOGIES, INC., IAPRO INSTRUCTIONAL MANUAL 68, 141-42 (Mar. 2019) (available at https://wokewindows-data.s3.amazonaws.com/iapro/somerville/IAProManual.pdf) ("We recommend that you normally purge the employee from the incident instead of purging the entire incident. This allows you to keep important statistical information such as incident counts, allegation counts and actions taken, just to name a few.").

[79] *See, e.g.*, Thompson v. City of Lebanon, No. 3:11-CV-00392, 2014 WL 12677063, at *2 (M.D. Tenn. June 10, 2014) ("LPD became aware through alerts from its "Personnel Early Warning System" ("PEWS") that [Thompson] had an unusually high number of uses of force and car chases"); Florida courts have twice spoken of the software. *See* Johnson v. Dixon, No. 3:14-CV-579-J-39PDB, 2015 WL 12851563, at *17 (M.D. Fla. Nov. 20, 2015) ("the JSO utilizes [PEWS], which is a computerized system that automatically monitors and flags officers "whose behavior indicates potential problems."); Ramsey v. Fields, No. 3:10-CV-238-J-32MCR, 2012 WL 6803518, at *6 (M.D. Fla. Dec. 4, 2012)

analytics in a few different ways. First, departments can input thresholds for certain categories of reports, and IA Pro will alert them when an officer exceeds the threshold.[80] For example, a department sets the twelve-month threshold for firearm discharges at two. Then, when an officer fires his weapon a third time (and the department inputs the event into IA Pro), IA Pro places a warning icon next to the officer's name. IA Pro also allows a user to run a report that ranks employees by certain categories like Use of Force, or to identify "top percentile" employees in a category. IA Pro has a scoring mechanism to aggregate incidents, allegations, and types of force, weighing each so that administrators can identify whether to intervene with an officer.[81] Police administrators input the scoring weights that best suit their departments.

These features don't always work as intended. When Officer Matthew Johns of the Atlanta Police Department physically assaulted Antraveious Payne following a car chase, Payne's mother sued the department and the city.[82] In the past, Officer Johns had suffered from psychological distress stemming from his military service, and the plaintiff argued that Johns' personnel record and psychological evaluations should have alerted supervisors that he might endanger others.[83] The city countered that it uses IA Pro to alert the Office of Professional Standards for issues and even conceded that the software should have alerted someone, but that it didn't.[84] The federal court for the Northern District of Georgia held that the facts could not support a showing of deliberate indifference on the part of the city.[85] Thus, IA Pro might immunize a police department and city from liability simply because they track officers' files with its software and rely on its alerts.

Personnel file management software isn't the only software relevant for *Brady* disclosure. New York City police officers famously used to carry leather memo books to log their activity. In 2020, the NYPD retired the memo books and replaced them with an iOS app, installed on the officers' department issued iPhone.[86] Although reports tout the app's crime fighting capabilities, digitally collected data will also allow the NYPD to produce analytics on officers themselves and to identify patterns in their behavior, potentially relevant for *Brady* disclosure. Furthermore, the police department will control access to the data, instead of officers themselves, who used to keep the books in their lockers or homes.[87]

C. *Brady* Compliance

Ideally, internal affairs software facilitates *Brady* compliance, streamlining disclosures and misconduct review. And IA Pro offers several features to this end. Users can link officers to the files for charges and hearings, allowing quick access to relevant

(describing PEWS). An audit in Easton, Pennsylvania recommended that the city adopt a PEWS. *See* Hogan v. City of Easton, No. CIV A 04-759, 2006 WL 3702637, at *5 (E.D. Pa. Dec. 12, 2006).

[80] *See* IAPRO INSTRUCTIONAL MANUAL, *supra* note 78, at 176-79.

[81] *Id.* at 184.

[82] Brown v. City of Atlanta, No. 1:17-CV-04850, 2020 WL 5633399 at *12 (N.D. Ga. Sept. 21, 2020).

[83] *Id.* at *13.

[84] *Id.* at *6.

[85] *Id.* at *13.

[86] *See* Corey Kilgannon *Why the N.Y.P.D. Dropped One of Its Oldest Crime-Fighting Tools*, N.Y. TIMES (Feb. 5, 2020) https://www.nytimes.com/2020/02/05/nyregion/nypd-memo-book.html.

[87] *Id.*

personnel files.[88] Statistical alerts and threshold warnings provide innovative ways for internal affairs departments to identify problematic officers.[89]

But other features reflect the resistance of police departments and unions to *Brady*. IA Pro's limited disclosure functionality only allows users to print reports to PDF or send electronic ones internally, rather than enabling external disclosure to defense attorneys and prosecutors from within the application.[90] Additionally, when a department can escape liability for *Brady* or other civil rights violations simply by using IA Pro—even when it is used ineffectively—departments have no incentive to ensure its effective use.[91] Nor do they have reason to request or purchase new features to help them better comply with constitutional mandates.

Brady lists also imperfectly protect defendants' rights. Rather than encouraging disclosure, lists aim to ensure compliance by preventing officers with *Giglio* material in their files from testifying in the first place. When this system functions effectively, tainted officers' testimony never reaches a courtroom. Prosecutors refrain from charging defendants when the evidence relies solely on the officer's word. However, these lists can be both over- and underinclusive from the defendants' perspective.

Lists are underinclusive insofar as they omit officers with non-misconduct *Giglio* information, such as personal motives to testify against a specific defendant. They may also exclude an officer whose file contains less misconduct than the prosecutor's minimum threshold for inclusion on the list even though a defense attorney finds the information valuable for impeachment purposes. A results-oriented defense attorney might also identify statistical patterns in an officer's file that a prosecutor's cursory list review misses.

Brady lists are also overinclusive. Prosecutors include officers with false misconduct reports, minimal reports, or reports immaterial to the case at hand. Recall that *Brady*'s materiality standard means a violation occurs only when undisclosed evidence would have prejudiced the outcome of the trial.[92] Some officers engage in misconduct severe enough to discredit their testimony in any case, while others appear on *Brady* lists due to conduct that could influence the result in specific cases, but not all situations. Indeed, most disclosed evidence is consistent with the defendant's guilt.[93] Likewise, an officer's personnel file that convinces the DA to place them on their *Brady* list may not be persuasive enough in the hands of the defense attorney during cross-examination to convince a jury of the defendant's innocence.

Both *Brady* lists and internal affairs software fall short in protecting a defendant's *Brady* rights when officers can escape the file by changing departments. Police officers, even those dismissed for misconduct, often secure employment with new police departments.[94] And they tend to join departments with fewer resources and

[88] *See* IAPRO INSTRUCTIONAL MANUAL, *supra* note 78, at 27-39.
[89] *See id.* at 176-79.
[90] *See id.* at 68-69.
[91] *See* Brown v. City of Atlanta, supra note 82, at *13.
[92] *See* United States v. Bagley, supra note 18, at 681-82 (Blackmun, J. & O'Connor, J., concurring).
[93] *See* BIBAS, Brady v. Maryland, *supra* note 6, at 146.
[94] *See* Grunwald and Rappaport, *The Wandering Officer*, 129 YALE L.J. 1676, 1758–59 (2020) (highlighting that the Cleveland Police Department failed to review the personnel file of the officer who killed Tamir Rice).

serving communities of color.[95] As long as the officer doesn't lose his license, the new department might never share the previous department's personnel file with local prosecutors. Consequently, those prosecutors cannot determine if the contents warrant disclosure or placement on a list. While the file retains its persuasive weight to a jury, it evades local prosecutors' knowledge and access, complicating their constitutional burden to learn of evidence impeaching or exculpatory. In this situation, defendants suffer most of all, unable to effectively cross-examine the officer.

Beyond lists and internal affairs software, certain aspects of the criminal justice system that hobble *Brady*'s ability to protect defendants doubly inhibit the doctrine's ability to protect defendants through disclosure of officers' personnel files.[96] Police and prosecutors take an adversarial mindset to investigations and trials which can cause them to overlook evidence that fails to change their minds about the defendant's guilt. They can presume an officer's file fails *Brady*'s materiality requirement because the information within is consistent with their theory of the defendant's guilt, even if the file is riddled with disclosable *Giglio* information. On top of that, *Brady*'s weak enforcement mechanism means the worst consequence officers suffer is a retrial for the defendant, hardly a deterrent strong enough to encourage compliance.

Additionally, *Brady* applies only to trials, not to pre-trial proceedings like plea bargaining, where the vast majority of prosecutions are resolved.[97] Consequently, because they cannot evaluate the trustworthiness or effectiveness of witnesses against them, nearly all defendants face an information asymmetry at moment the criminal justice system resolves their case.

III. OFFICERS' DUE PROCESS RIGHTS

As crudely as *Brady* protects defendants, it offers even less to the officers whose files prosecutors disclose. When DA Larry Krasner implemented a Police Misconduct Disclosure Database in Philadelphia and published DA Williams' "Do Not Call" list,[98] the city's Fraternal Order of Police (FOP) responded by filing a lawsuit claiming that Krasner violated the officers' due process rights by publishing the list and placing them in the database. The FOP alleged "unconstitutional and illegal treatment of Police Officers with respect to the improper and illegal disclosure of their confidential personnel records. . . . without regard to . . . the due process rights of Police Officers."[99] The Court of Common Pleas of Pennsylvania dismissed the complaint, but the Commonwealth Court—the Pennsylvania appellate court—held that the FOP and officers plead sufficient allegations to state a claim that Krasner and the City violated their procedural due process rights and caused them actual harm.[100]

[95] *Id.* at 1687.

[96] *See* BIBAS, Brady v. Maryland, *supra* note 6, at 139-40.

[97] *See* Stephanos Bibas, *Regulating the Plea-Bargaining Market: From Caveat Emptor to Consumer Protection*, 99 CALIF. L. REV. 1117, 1118 (2011).

[98] In the lawsuit, "Do Not Call List" refers to both the tracking mechanisms and the published list. *See* Fraternal Ord. of Police Lodge No. 5 by McNesby, *supra* note 9, at 536 n.8.

[99] *See* Second Amended Complaint, *Id.*

[100] *See id.* at 550.

When an officer in Washington state was placed on a *Brady* list in 2007, he successfully sued for damages under § 1983 for violation of his due process rights.[101] But in Louisiana, an officer was unsuccessful on the same theory. In *Adams v. City of Harahan*, a police department chief filed internal disciplinary action against a captain.[102] But before the captain could appeal to the civil service board, the chief emailed the disciplinary records to the DA's office, landing the captain on the office's *Giglio* list. The Fifth Circuit held that due process did not protect the officer's liberty interest in his employment.[103] Likewise, the Oregon Court of Appeals found "no evident procedural requirements imposed on prosecutors in Oregon in deciding whether a particular witness should be removed from a *Brady* list."[104] On the legislative front, lawmakers in Arizona and California enacted statutes to protect officers from abuse by *Brady* lists.[105] California's Public Safety Officers Procedural Bill of Rights (POBR) enacts comprehensive protections for police officers in internal affairs investigations and when adding documents to officers' personnel files.[106]

These events urge more critical review of the claim that police officers have due process rights against placement on *Brady* Lists or in *Brady* databases.[107] This Part investigates these lawsuits, focusing on Pennsylvania, to answer what due process rights the officers claim and whether those claims are valid so that Part IV can examine how to protect them.

A. Procedural

Procedural due process refers to the requirement that the state not deny a person life, liberty, or property without notice, hearing, and a decision by an unbiased authority.[108] Under *Cleveland Board of Education v. Loudermill*, public employees facing termination are entitled to "oral or written notice of the charges against him, an

[101] *See* Moran, Brady *Lists supra* note 4, at 719 (citing Wender v. Snohomish Cnty., No. C07-197Z, 2007 WL 3165481 (W.D. Wash. Oct. 24, 2007)).

[102] 2023 WL 2945725, *1 (5th Cir. 2023).

[103] *Id.*

[104] *See* Lane v. Marion Cnty. Dist. Att'ys. Off., 310 Or. App. 296, 307 (2021).

[105] *See* Moran, *Brady Lists*, *supra* note 4, at 678-679 (referring to H.B. 2295, 55th Leg., 1st Reg. Sess. (Ariz. 2021), subdiv. (E)(1), subdiv. (H); ARIZ. REV. STAT. ANN. § 38-1117 (2021) (effective Sept. 24, 2022) and CAL. GOV'T CODE § 3301-3313 (West 2022)).

[106] *See* CAL. GOV'T CODE § 3303, 3305-06 (West 2022). The California Supreme Court balanced statutory protections over personnel files against defendants' *Brady* disclosure rights in Ass'n for L.A. Deputy Sheriffs v. Superior Court, 8 Cal.5th 28 (Cal. 2019), concluding that the statute did not bar disclosure. *See Whose Rights Matter More—Police Privacy or a Defendant's Right To A Fair Trial?* 54 LOY. L.A. L. REV. 495, 502. ("Ultimately, the court concluded that the confidentiality created by the *Pitchess* statutes does not forbid disclosure to prosecutors of *Brady* alerts.")

[107] Mary Ellen Reimund, *Are* Brady *Lists (aka Liar's Lists) the Scarlet Letter for Law Enforcement Officers? A Need for Expansion and Uniformity*, 3 INT'L J. HUMANS. & SOC. SCI. 1, 4 (2013), https://www.ijhssnet.com/journals/Vol_3_No_17_September_2013/1.pdf (reviewing how officers in King County, Washington end up on Brady lists but noting that prosecutors' offices leave "due process considerations" to the police agencies, and the *Brady* committee does not reconsider the agency's internal processes). Professor Abel notes the "grave employment consequences" and "potential for police management to misuse" *Brady* lists. Abel, *supra* note 7, at 780-81 (collecting statements from both officers and prosecutors concerned about the potential for abuse of *Brady* lists).

[108] Henry J. Friendly, *"Some Kind of Hearing"*, 123 U. PA. L. REV. 1267, 1279-82 (1975).

explanation of the employer's evidence, and an opportunity to present his side of the story."[109]

Loudermill represents a broader right to procedural due process protections where state action infringes on a public employee's interest in continued employment.[110] That due process should include "oral or written notice," especially when the employee's employment is in danger.[111]

Pennsylvania courts have applied the federal Supreme Court's balancing test for determining what process is due in pretermination hearings: "a court must balance the private interest involved, the risk of an erroneous deprivation of that interest through the procedures used and the probable value of any additional procedural safeguards, and the government's interest."[112] *Mathews v. Eldridge* thus offers guideposts for balancing the officers' and state's interests to ensure that the officers are afforded due process.[113]

1. Fair Notice

Notice refers to the opportunity for a someone to react or respond. Notice is "fair" or "due" when it occurs sufficiently in advance of particular consequences. In trial contexts, for example, a plaintiff must afford the defendant fair notice of his claim and reasoning.[114] California's broad POBR requires internal affairs investigations to provide written notice of the subject matter of the investigation and names of investigators.[115] It also mandates notice before documents are added to an officer's personnel file.[116] In Pennsylvania public employment contexts, written notice should also explain the reasons for the potential dismissal.[117]

a. Advanced Notice

The most basic element of fair notice requires that sufficient time elapse between notice and consequences. [118] Prosecutors must provide police officers enough time to object before they suffer substantive consequences.

[109] Bethel Park Sch. Dist. v. Bethel Park Fed'n of Tchrs., Loc. 1607, 55 A.3d 154 (Pa. Commw. Ct. 2012) (citing Cleveland Board of Education v. Loudermill*,* 470 U.S. 532, 546 (1985)).

[110] *See* Gniotek v. City of Philadelphia, 808 F.2d 241 (3d Cir. 1983) (holding that suspensions demand notice and hearing); City of Philadelphia v. Fraternal Ord. of Police*,* Lodge No. 5, 140 Pa. Cmwlth. 235, 592 A.2d 779 (1991) (finding that suspension proceedings satisfied due process); *contra* Vander Zee v. Reno, 73 F.3d 1365, 1371 (5th Cir. 1996) (finding that a future law enforcement career instead of termination did not violate a constitutionally protected property interest).

[111] *See* Veit v. N. Wales Borough, 800 A.2d 391 (Pa. Commw. Ct. 2002) (citing Loudermill, *supra* note 109, at 542-46 (explicating the right to notice and hearing)).

[112] *Id.* (citing Mathews v. Eldridge*,* 424 U.S. 319, 332-50 (1976)).

[113] *See* Dee v. Borough of Dunmore, 549 F.3d 225 (3d Cir. 2008) (applying the Mathews v. Eldridge framework to a suspension).

[114] *See* Bell Atlantic Corporation. v. Twombly, 550 U.S. 544, 545 (2007).

[115] *See* CAL. GOV'T CODE § 3303(b) (West 2022).

[116] *See id.*, § 3305.

[117] *See* Veit v. N. Wales Borough, *supra* note 111, at 398.

[118] *See* Fraternal Ord. of Police Lodge No. 5 by McNesby v. City of Philadelphia, *supra* note 9, at 638-39 ("Emphasizing that advanced notice is the most basic of all requirements, we held that publication of the report without giving the petitioners notice . . . was in violation of article I, sections 1 and 11 of the Pennsylvania Constitution.").

Unfortunately, they rarely do. As Jonathan Abel notes, "prosecutors can make *Brady*-cop designations . . . without giving officers an opportunity to contest the allegations beforehand or to appeal the decisions afterwards."[119] Abel's observation manifests in practice. A Police Chief in Michigan received no notice before he found himself on the prosecutor office's *Brady* list. [120] An Assistant Prosecuting Attorney reviewed the situation, noting that not one officer had received notice that the prosecutor's office placed them on the list.[121] Likewise, the *Harahan* officer found himself on the DA's *Brady* list before he had time to challenge the disciplinary actions.[122]

In Philadelphia, former DA Williams never told officers that he created a *Brady* list, and the officers only found out later when a court mandated that DA Krasner release the list.[123] Since the officers suffered immediate reputational harm when the Inquirer published the list, the DA's office failed to provide notice sufficiently advanced of the reputational consequences.[124]

b. Form and Description of Consequences

While the form of fair notice only needs to comply with other due process strictures, it should identify the consequences that the employee may face. In *Brady* contexts, fair notice must alert police officers to the prosecutor's disclosure obligations and the danger to their employment, privacy, and reputation posed by that disclosure.[125]

The Philadelphia officers received notice in two ways. They first found their names on the Do Not Call list when the Philadelphia Inquirer published it. The DA Office also alerted officers directly by sending a letter to each officer, informing them that the office had placed their name on the Do Not Call list. Those letters "informed them of the District Attorney's disclosure obligations 'ahead of any hypothetical disclosure' and invited them to communicate in writing if they believed the information in the Letter was 'incorrect.'"[126]

Notice by city-wide publication immediately and permanently harmed the officers' reputations and failed to explain either how the officers ended up on the list or what further consequences they could face. Had that been the only notice, it would hardly have passed muster under federal or Pennsylvania due process strictures.

The letters partially solved the problem. They were sent privately to the officers so that their reputations were kept safe for the time being. The letters also explained that the law requires the DA to disclose certain contents of their files. In keeping with

[119] *See* Abel, *supra* note 7, at 781.

[120] *See* Val Van Brocklin, *Do Brady and Giglio trump officers' due process rights?*, POLICE1 BY LEXIPOL (Jan. 25, 2022) https://www.police1.com/patrol-issues/articles/do-brady-and-giglio-trump-officers-due-process-rights-g585QOS4UeSOSF5u/. Compare Latty v. Polk Cnty. Sheriff's Off., No. 21-35794, 2022 WL 5241297 (9th Cir. Oct. 6, 2022) ("It is undisputed that Latty received an opportunity to be heard prior to his disciplinary suspension and an opportunity to be heard by DA Felton before he was placed on the *Brady* list.")

[121] *Id.*

[122] *See* Adams v. City of Harahan, *supra* note 102, at 269-70.

[123] *See* Fraternal Ord. of Police Lodge No. 5 by McNesby, *supra* note 9, at 546-47.

[124] *Id.*

[125] *See id.*; Matthews v. Eldridge, *supra* note 112.

[126] *See id.* at 549.

Mathews v. Eldridge, the letters included information within the letter justifying their placement on the *Brady* list.[127]

2. Hearing

To comply with due process, a hearing must meet several requirements. It must occur before an impartial or unbiased tribunal. It must offer the employee opportunity for redress.[128] And the employee must be able to present and defend their version of events, assisted by counsel.[129]

a. Access to Counsel

In most contexts, fair hearing requires that the accused have access to counsel.[130] In fact, Krasner's letters to the officers said, "if you believe our information is incorrect, feel free to communicate to us in writing through *counsel*."[131] In Philadelphia, the police union provides counsel for officers in lawsuits like this one. The union in this suit—Lodge No. 5 of the Fraternal Order of Police—has represented police officers in Philadelphia for many years.[132] Had any of the "Do Not Call" list officers sought the hearing provided for in the letters, their union could have represented them.

b. Impartial tribunal

"[A]n impartial decision maker is essential,"[133] especially "in the public employment context, where the reason for the challenged dismissal may well be related to some personal antagonism between the employee and his superior."[134] The further the decider is removed from the agency and disputed events, the better. The police officers in Philadelphia complained that DA Krasner would have sole discretion over whether the officers would remain on the list or not.[135] Since he placed them on the list, the Commonwealth Court found it inadequate to conduct hearings before him. In other words, the Court found that the DA could not be an impartial decider.

In termination situations, both federal and Pennsylvania courts hold that an employee does not have a due process right to an impartial decider at the *pre*-termination hearing as long as an impartial decider has authority over the *post*-termination hearing.[136] Officers are not terminated per se when they are placed on *Brady* lists or when internal affairs includes *Giglio* material in their personnel file. Lists

[127] *Id.*; Matthews v. Eldridge, *supra* note 112.
[128] *See* Friendly, *supra* note 108, at 1279-81.
[129] *See* Veit v. N. Wales Borough, 800 A.2d 391 (Pa. Commw. Ct. 2002) (citing Loudermill, *supra* note 109, at 542-46).
[130] *See e.g.*, CAL. GOV'T CODE § 3303(i) (West 2022).
[131] *See* Fraternal Ord. of Police Lodge No. 5 by McNesby, *supra* note 9, at 537 (emphasis altered).
[132] *See* FRATERNAL ORDER OF POLICE, PHILADELPHIA LODGE #5, https://www.fop5.org/ ("We are the labor union for Police Officers and Sheriffs in the city of Philadelphia…. We vigilantly and vigorously protect, promote and improve the working conditions, legal rights, salary compensation, pensions & benefits of Philadelphia Police Officers and Deputy Sheriffs."); *See, e.g.*, City of Philadelphia v. Fraternal Ord. of Police, Lodge No. 5, *supra* note 110.
[133] Goldberg v. Kelly, 397 U.S. 254, 271 (1970).
[134] Arnett v. Kennedy, 416 U.S. 134, 216 (1974).
[135] *See* Fraternal Ord. of Police Lodge No. 5 by McNesby, *supra* note 9, at 549.
[136] *See* McDaniels v. Flick, 59 F.3d 446, 460 (3d Cir. 1995), Edwards v. Beaver Cnty. Career & Tech. Ctr*., No.*, 2020 WL 1130853, at *6 (Pa. Commw. Ct. Mar. 9, 2020).

and disclosures may threaten officers' employment, but they do not directly cause dismissal. This weighs on the private interest factor of the *Mathews v. Eldridge* balancing test for determining the situationally necessary due process.[137] The private interest—the officer's interest in his employment—is lesser because the list or disclosure does not directly threaten his employment. This is a fact dependent inquiry. In each situation, reviewing courts must grapple with whether that means that hearings over *Brady* list determinations and personnel file disclosures require a more impartial and neutral arbiter than oversaw the hearing in the first place.

Who ideally should preside over the hearings? In Louisiana, the police captain alleged that the police chief intended to "clean house," using disciplinary action and consequent *Giglio* list placement to end the captain's and colleagues' careers.[138] The department thus afforded the captain insufficient due process because the hearing was conducted by a biased decision maker.[139]

District Attorneys compose *Brady* lists and make disclosure determinations. As the Commonwealth Court observes, they are thus too involved to decide impartially. On the other hand, internal affairs departments work closely with officers and are likely to decide in their favor unduly often. *Brady* committees are often staffed by prosecuting attorneys or attorneys from the internal affairs department.[140] Courts could oversee hearings in exceptional cases, but they would quickly get overwhelmed if they had to review every dispute over *Brady* list designations and information included in personnel files. Identifying a truly impartial party is a challenge, and District Attorneys, sitting as they do between the courts and police departments, might be the *least* biased competent party.

c. Redress

Lastly, a hearing must offer opportunity for redress. *Brady* list hearings need the power to expunge officers from the list. Review of personnel file reports must be able to purge the report from the officer's file (or, as IA Pro recommends, purge the officer from the report).[141] In Philadelphia, the DA's letters included information about why the officer was placed on the Do Not Call list. The union complained and the court agreed that this did not allow the officers to seek removal from the list. The court held that "there must be some post-placement mechanism available for an officer to seek removal from the Do Not Call List." So too in Louisiana, the police captain did not receive proper procedural due process because the Civil Service Board, which he could appeal to, did not have the authority to remove him from the DA's *Giglio* list.[142]

The private interest in redress is strong. When officers seek to be removed from a *Brady* list or to purge a report from her file and a hearing finds that the misconduct report is accurate, then the state's interest outweighs the private interest, but redress is not warranted anyway. When the hearing tribunal determines that the relevant report is

[137] Mathews v. Eldridge, supra note 112, at 334-35.
[138] Adams v. Walker, 2021 WL 5833965, *4 (E.D. La., 2021) (overturned on the grounds that the captain did not have a liberty interest in his law enforcement career on Supreme Court or Fifth Circuit precedent).
[139] *See* Matthews v. Eldridge, *supra* note 112, at 334-35.
[140] *See* Moran, Brady *Lists supra* note 4, at 696-700.
[141] *See* note 78 *supra* and accompanying text.
[142] *See* Adams v. Walker, *supra* note 138, at *4.

false or mistaken, the private interest predominates and the hearing must have the power to remove the officer from a *Brady* list or purge the report. Otherwise the process given is not the process due.

B. Substantive

In Pennsylvania, a due process claim requires that the state deprive individuals of "life, liberty, or property interest" attended by constitutionally insufficient procedure.[143] The FOP asserted that the officers maintained interests in their reputations and their employment,[144] and the court found both rights in the Pennsylvania constitution.[145] The Commonwealth Court firmly rejected the trial court's ruling that harms to reputation and employment were speculative or hypothetical, holding that the union pled sufficient facts to demonstrate actual harm. One officer "was removed from his new assignment . . . resulting in a loss of overtime income and the potential of other career promotions."[146]

The Do Not Call list and Police Misconduct Disclosure database each had a different and distinct impact on police officers' employment prospects, reputation, and privacy. The Commonwealth Court lumped them together, but future courts should keep separate the harms from *Brady* disclosures and the harms from *Brady* lists.[147]

The state has a distinct interest in each too. It has an outsize interest in disclosures: the prosecutor's duty is "inescapable."[148] Officers' interests can counterbalance the state's up to the point that notice and hearing interfere with the prosecutor's ability to carry out constitutionally mandated disclosure. After that point, the state's interest in *Brady* must mean due process for the officer comes second. For example, if evidence of dishonesty arises late at night before trial, disclosure to the defendant trumps the officer's interest in employment, reputation, or privacy.

The state's interest in *Brady* lists is less concrete. Prosecutors use *Brady* lists to avoid the surprise at trial of calling a witness that the jury won't believe.[149] The lists also encourage police departments to prevent listed officers from conducting investigations, making arrests, or carrying out any other police functions that could land them on the witness stand.[150] *Brady* lists are thus useful but not critical to the state. They are one means of conveying information amongst prosecutors and to police command staff, but by no means the only way.

[143] Commonwealth. v. Turner, 622 Pa. 318, 335 (2013).

[144] *See* Second Amended Complaint, Fraternal Order of Police, Lodge No. 5 v. City of Philadelphia, *supra* note 16.

[145] *See* Fraternal Ord. of Police Lodge No. 5 by McNesby, *supra* note 9, at 546 (citing Pa. Const. art. I, §§ 1, 11).

[146] *See id.* at 552 (citing Second Amended Complaint, Fraternal Order of Police, Lodge No. 5 v. City of Philadelphia, *supra* note 16 at ¶ 180).

[147] In the lawsuit, "Do Not Call List" refers to both the tracking mechanisms and the published list. *See* Fraternal Ord. of Police Lodge No. 5 by McNesby, *supra* note 9, at 536 n.8.

[148] *See* Kyles v. Whitley, *supra* note 3 at 438 (1995).

[149] *See* HARMON, *supra* note 33, at 228; section II.B, *supra*.

[150] *Id.*

1. Employment

The federal Constitution does not protect public employees' property interests in their employment on its own, but the Fourteenth Amendment can do so through state law.[151] The Pennsylvania Constitution does just that, protecting individuals against deprivation of property without due process.[152] Citing a case from New Hampshire, the Commonwealth Court held that when police officers contend that the allegations against them are false, they have a constitutionally protected interest in their continued employment that the state cannot deprive them of without due process.[153]

Each of the actions a prosecutor can take with respect to *Giglio* evidence in an officer's file—collection, disclosure, *Brady* list placement—affects the officer's employment interest differently. Simply collecting information and reporting it in a Police Misconduct Database or internal affairs software causes the least damage to the officer's employment prospects (and reputation) because it does not affect his ability to testify until a prosecutor reviews it. Notably, it does make it easier for a prosecutor to find the evidence, leading to disclosures later.

Disclosure of *Giglio* evidence can indirectly cause an officer's de facto termination. When impeaching information piles up in an officer's file or a single devasting piece of evidence like a perjury conviction arises, disclosure publicizes that information. Then prosecutors cannot call him as witness, and an officer who cannot testify cannot effectively make arrests or investigate crimes.[154] At best, the department will shift a *Brady* list officer to desk duty.[155] At worst, he will get "fast-tracked for termination and hard-pressed to find future work."[156] In sum, *Brady* disclosures affect officers' employment interest when they are many or when they are great.

Brady lists work more directly to jeopardize officers' careers, labeling them as tainted or liars.[157] It only takes one appearance on a list to scare prosecutors off pursuing the officer's cases, regardless of the substantiality of the evidence that landed him there. As the Commonwealth Court said "[t]here can be little question that placement on a formal list of officers who are deemed untrustworthy or unworthy [of testifying] could very well be detrimental to their reputations . . . in the employment context if released."[158] In that way, *Brady* lists compound the effect of prior misconduct into significant employment consequences.

[151] *See* Adams v. City of Harahan, 2023 WL 2945725, *3, *5 (5th Cir. 2023) (reviewing two cases relied on by the district court—Kerry v. Din, 576 U.S. 86 (2015) and Meyer v. Nebraska, 262 U.S. 390 (1923)—and finding neither supports a liberty interest in one's profession); Dee v. Borough of Dunmore, 549 F.3d 225, 229-30 (3d Cir. 2008) (citing Board of Regents v. Roth, 408 U.S. 564, 577 (1972), Kelly v. Borough of Sayreville, 107 F.3d 1073, 1077 (3d Cir. 1997), and Brown v. Trench, 787 F.2d 167, 170 (3d Cir. 1986)).

[152] *See* Fraternal Order of Police, Lodge No. 5 v. City of Philadelphia, *supra* note 16, at 546, citing Pa. Const. art. I, §§ 1, 11.

[153] *Id.* at 549 (citing Duchesne v. Hillsborough Cnty. Att'y, 167 N.H. 774, 783-84 (2015))

[154] *See* Abel, *supra* note 7, at 780-81.

[155] *See* HARMON, *supra* note 33, at 229.

[156] *See* Abel, *supra* note 7, at 781.

[157] *See* HARMON, *supra* note 33, at 228.

[158] Fraternal Ord. of Police Lodge No. 5 by McNesby, *supra* note 9, at 547-48.

2. Reputation

The Commonwealth Court also identified in the Pennsylvania Constitution a protected interest in officers' reputations which the government may not infringe upon without due process.[159] By published the Do Not Call list, the Inquirer exposed for the whole community of Philadelphia the "blacklist of sorts."[160] That kind of public shaming and exposure exacerbated the harm to the officers, but a more limited audience does not cure the harm to the officers' reputation.[161]

An officer's reputation amongst his colleagues and the prosecutors he works with matters. *Brady* lists, published in local newspapers or not, inform prosecutors and police command staff that the listed officer cannot effectively carry out his duties. That sort of reputational harm merits due process before it is deprived.

Disclosure to defendants functions differently. Where a *Brady* list announces that prosecutors cannot rely on an officer, disclosure of impeaching evidence might or might not affect whether prosecutors can confidently call the witness to the stand. In most cases, disclosed material won't change whether the officer can make arrests, investigate cases, and provide testimony. When disclosure doesn't change the officer's ability to do his job, he suffers only minimal damage to his reputation. Further, protective orders often cabin the audience of the disclosure, limiting the impact on the officer's reputation to a judge, defendant, and the defendant's attorney.[162]

3. Privacy

Lawmakers often justify police protections against *Brady* disclosures as concern for officers' privacy.[163] Yet while the FOP argued in their complaint that the Pennsylvania Constitution "recognizes a right to privacy" which protects officers' reputations,[164] the Commonwealth Court largely declined to discuss any independent right to privacy. The concurrence added in a footnote that a constitutional right to privacy could not overcome the prosecutor's Fourteenth Amendment *Brady* obligations. In part, the Commonwealth Court may have been reluctant to stand on an officer's right to privacy because its opinion focused on the published Do Not Call list rather than the Police Misconduct Database or *Brady* disclosures. The Do Not Call list contained only sparse facts and notes, a weak intrusion into the officers' privacy in comparison to the reputational and employment harm it could cause.

[159] *See* Fraternal Order of Police, Lodge No. 5 v. City of Philadelphia, *supra* note 16, at 546, citing Pa. Const. art. I, §§ 1, 11.

[160] *Id.*; *see* Fazlollah, *supra* note 13, (publishing the officers identified by former DA Williams, released by DA Krasner).

[161] *Id.* at 547 (citing Pennsylvania Bar Ass'n v. Commonwealth Insurance Department, 147 Pennsylvania. Cmwlth. 607 A.2d 850, 855 (1991) ("we conclude that reputation is a fundamental right under the Pennsylvania Constitution, and it is entitled to the protection of procedural due process even in the absence of a more "tangible" right.")).

[162] *See* Abel, *supra* note 7, at 804.

[163] *See* Moran, *Contesting Police Credibility supra* note 37, at 1369 (2018) ("As a doctrinal matter, lawmakers frequently justify their unwillingness to allow defense counsel access to police records by theorizing that such access would violate officers' rights to privacy.").

[164] *See* Second Amended Complaint, Fraternal Order of Police, Lodge No. 5 v. City of Philadelphia, *supra* note 16.

Courts in Pennsylvania have identified a right to privacy covering personnel files in other contexts: "a constitutional right to privacy protects public school teachers and their personnel files," but they always weigh that right against competing restrictions.[165] Other state constitutions protect government workers' privacy rights too.[166] However, in *Brady* contexts, constitutional demands outweigh the limited intrusion into officers' privacy, especially where policies such as protective orders and redacted disclosures can limit the extent of the intrusion. Accordingly, the federal Supreme Court's jurisprudence suggests that police misconduct reports fall outside the zone of protected informational privacy.[167]

IV. RECOMMENDATIONS

As police unions fight against *Brady* lists and push back against progressive reforms that rely on greater disclosure, police departments have adopted software systems to manage their personnel files that provide analytics capabilities and generate reports for easy disclosure. Those software systems are an opportunity to improve *Brady* compliance. But easier disclosure means police and their unions will resist the change. Still, a number of improvements to software applications like IA Pro and other file managers can facilitate *Brady* compliance while also protecting officers' interests.

Make disclosure harder, and departments will slouch away from constitutionally sufficient disclosure. Make disclosure easier, and departments will drift towards it. Better software could alleviate police unwillingness to disclose misconduct, especially when it simultaneously mitigates negative employment consequences.[168] Progressive, reform-minded prosecutors in particular should welcome ways to effect disclosures along this path of least resistance, while maintaining positive relationships with the officers whose cooperation and good will they need to retain in order to remain effective. This final Part maps out the software functions that can attend that goal. Section IV.A outlines features to promote *Brady* compliance. Then Section IV.B describes software practices and functionality to best protect officer's due process rights.

A. *Brady* Compliance

1. Defendant Access

The most important *Brady* reform focuses on guaranteeing that defendants receive *Brady* and *Giglio* evidence without which they cannot experience a fair trial. For starters, digitized documents are easier to distribute. IA Pro allows users to generate reports for printing and can track those disclosures, but they don't need to rely on physical copies or PDFs at all. Instead of producing photocopies and couriering manilla envelopes to defense attorneys, software programs should build in email functions to send reports. Better yet, software that offers an internal portal for defense attorneys

[165] Bangor Area Educ. Ass'n v. Angle, 720 A.2d 198, 201 (1998).

[166] *See, e.g.*, CAL. CONST. art. I, § 1 ("All people . . . have inalienable rights. Among these are enjoying and defending life and liberty, . . . and privacy."). State courts have also recognized the same right where constitutions omit explicit protection. *See, e.g.*, Jegley v. Picado, 349 Ark. 600, 632 (2002) ("the fundamental right to privacy [is] guaranteed to the citizens of Arkansas.").

[167] *See* Rachel Moran, *Police Privacy*, 10 U.C. IRVINE L. REV. 153, 177, 198 (2019) (concluding, after examining the interests at stake when concealing police misconduct records, that "privacy is overused as a justification for denying public access to misconduct records.")

[168] *See* Abel, *supra* note 7, Section III at 779.

would allow them access while keeping control with the government over what information they see and how they can redistribute it.

Digital applications also facilitate data collection and organization. Once filled, the NYPD's old leather memo books sat unexamined in officers' lockers. Any police department relying on physical files and analog organization has to keep the files in cabinets or boxes. By storing reports and documents in databases, departments can access them faster and organize them more efficiently for disclosure. For example, whenever prosecutors begin a case, they should review internal files for any officers they plan to call to the witness stand. With software, they can do that from their own laptops, instead of visiting the backrooms of the precinct.

2. Analytics

IA Pro offers rudimentary statistical review of officers' reports and files. As discussed *supra*, it can warn users when an officer exceeds certain thresholds such as three firearm discharges in twelve months. More sophisticated software could alert supervisors and prosecutors when an officer conducts traffic stops disproportionately on people of color. It could compare written warrant justifications and descriptions of the reliability of informants to estimate their veracity. Consider *State v. Duarte* where an officer compiled an affidavit in support of a warrant but copied over half of the language from an affidavit written by a different officer in a different case.[169] That behavior is all too common: police often use substantially identical language when composing warrant affidavits.[170] Software should catch that dishonesty.[171]

Artificial Intelligence might soon be capable of comparing the facts that an officer reports with, for example, the facts in a citizen complaint. At the time of this writing, cutting-edge AI like ChatGPT cannot reliably compare two versions of events for consistency, but that may soon change. An AI plugin could compare the officer's and citizen's versions of events, alerting a human reviewer when they diverge in an important way.

3. Aggregating Likely Brady Material

Some types of information that sit in officers' personnel files are likely to qualify as favorable and material to any criminal defendant. That includes reports of dishonesty, arrests or convictions, reports of prejudice, negative performance reviews, and mishandling of evidence. All this should be easily accessible to certain software users. For example, when a prosecutor intends to try a case with certain officers as witnesses, he should be able to see all those reports in a single view.

Access to that information should be restricted. Police department leaders should have permanent access, but prosecutors should not.[172] They should only be able

[169] *See* State v. Duarte, 389 S.W.3d 349, 351 n.3 (Tex. Crim. App. 2012) and accompanying text.

[170] *See* L. Paul Sutton, *Getting Around the Fourth Amendment*, in THINKING ABOUT POLICE 433, 440-441 (Carl B. Klockars & Stephen D. Mastrofski eds., 2nd ed. 1991).

[171] Exactly when it should identify and alert that dishonesty is an open question. Alert the officer as soon as he writes the affidavit and he might adjust the language until it escapes algorithmic warning without altering that the warrant remains unjustified. Alert the officer too late, and he might complain of unfairness.

[172] *See* Abel, *supra* note 7, at 770-73.

to see the files at a point close to trial proceedings. Prior to that, they should only see something like IA Pro's EI score, so that they can make informed decisions about trying cases without invading the officer's privacy.

4. Forced Checklists

Software can build in manual checklists or design workflows that require users to complete specific steps.[173] An application like IA Pro should require prosecutors to review every police officer-witness's personnel file, to check for common impeaching information like falsified reports, and to review analytics on the officers' files. The checklist or workflow should then require the prosecutor to assert that the information is not *Brady* material in the case, or to send it to the defendant.

Of course, this would be harder in jurisdictions like California that restrict prosecutors' access to personnel files.[174] Checklists can still help. A single software checklist can require tasks from different users. A prosecutor could complete their checklist section, while the internal affairs department completes the *Brady* review section.

5. Case Linking

Officers testify repeatedly in various cases. If a prosecutor finds *Giglio* evidence in one of those cases that discredits the officer's testimony by casting doubt on his capacity for truthfulness, then that evidence probably casts the same doubt in every other case where the officer gave testimony. Software should track every time an officer testifies. When *Brady* material arises, prosecutors can validate that former defendants received the information, or disclose it to them.

Presently, defendants only rarely find out about *Brady* material via other prosecutions.[175] Officers use "litigation, legislation, and informal political pressure to blunt *Brady*'s application to their files." Judges impose protective orders against sharing *Brady* material with attorneys in past or future cases.[176] As Jonathan Abel argues, reducing the use of protective orders could go a long way towards resolving *Brady* violations. Case linking would be tough to sell to police without concomitant concessions to protect officers' interests. Still, software developers and *Brady* stakeholders can and should build the functionality so that it can be turned on if policymakers and special interests agree to do so.

B. Due Process Protection

1. Don't Create Brady Lists

Brady lists have (roughly) two benefits. First, they warn prosecutors against calling certain officers to the witness stand. If an officer's file contains too much

[173] *See* ATUL GAWANDE, THE CHECKLIST MANIFESTO: HOW TO GET THINGS RIGHT (Metro. Bks. 2009).

[174] *See* Abel, *supra* note 7, at 762-64 (describing regimes where prosecutors get "no access" to personnel files).

[175] This does happen though. *See, e.g.,* Fraser v. City of New York, No. 20CIV4926CMOTW, 2023 WL 144448, *1-*2 (S.D.N.Y. Jan. 10, 2023).

[176] *See* Abel, *supra* note 7, at 796 (listing procedural problems of systems that attempt *Brady* compliance through balancing systems).

impeaching material, he cannot testify effectively in the face of cross-examination about his truthfulness. Second, they expose officers' misconduct and dishonesty.[177] Both of these functions can be achieved without *Brady* lists qua lists, better protecting officers' reputations and employment prospects. Software can alert prosecutors who intend to call an officer as witness that a file contains impeaching evidence. With greater access to information, prosecutors can make individualized determinations about whether to pursue a trial and whether to call certain officers to the stand. One simple way to achieve this would be with IA Pro's Early Intervention (EI) tool: prosecutors should take caution before calling to the stand an officer identified by EI.

The core insight is that good *Brady* technology can obviate the need to "designate" officers anything at all. Providing prosecutors with access to personnel files—anonymized if needed—allows them to carry out their constitutional duties. No designation as a "*Brady* cop" or placement on a list, no reputational harm, no need for review before an impartial decisionmaker.[178]

As an added benefit to choosing not to create a *Brady* list, the DA office cannot face a FOIA request for it, and local newspapers cannot publish it. Information kept inside internal databases is more likely to stay private. That would protect officers against harms to their reputation, privacy, and employment prospects. Even if departments want to create a Brady list, there are opportunities to make lists that better serve prosecutors while simultaneously protecting officers.

2. Flexible *Brady* List Decision Rules

Although many offices that maintain *Brady* lists rely on committees to create *Brady* lists, many other possibilities exist. Legislatures could enact decision rules for their entire constituency, or district attorneys could define rules for their whole office. Then voters could hold policymakers accountable for the *Brady* decision rules in their community.

More focused policies could more appositely align practice with constitutional requirements. Because "case-specific knowledge is required to determine what is and is not Brady material," *Brady* committees have to make decisions in the abstract, unable to evaluate whether evidence is "favorable" or "material," two of the three elements of a *Brady* violation.[179] More flexible policies could allow individual prosecutors to specify their own *Brady* list decision rules. And internal affairs software programs could apply them to the needs and facts of a case.

3. Open Policy Discussions

Police officers complain they don't know why they are placed on *Brady* lists nor what standards and rules govern those decisions. When departments implement software to make those decisions algorithmically, the decision rules codified into the code have to be made in the abstract, away from considerations about particular officers. That would allow stakeholders and interested parties to negotiate over what rules are

[177] *Id.*; *See also*, Fraternal Ord. of Police Lodge No. 5 by McNesby, *supra* note 9, at 547 ("the [*Brady* list] is a blacklist of sorts.").
[178] *Contra*, Moran, Brady *Lists supra* note 4, at 728-32 (arguing that *Brady* lists should be "a minimal requirement for all prosecutor officers.").
[179] *See* Abel, *supra* note 7, at 796-98.

most appropriate for the department and community. A neutral party—an algorithm—would then apply those rules. Abstract discussions and neutral application of rules could depoliticize some of *Brady* practice, calming the "battle" between officers and prosecutors to the benefit of both parties and defendants.[180]

4. Immediate Notice and Hearing

To give officers their due process before *Brady* disclosure deprives them of any protected interest, software should provide them immediate notice whenever someone enters common types of *Brady* information into their file. They should receive an email or notification alerting them that they can challenge the information with instructions how to do so. The email should clearly explain what consequences may follow from the information including harm to their reputation, to their ability to testify, and to their employment. The email or notification should require the officer to acknowledge receipt and that they understand that it constitutes fair notice.

Internal affairs offices should likewise conduct periodic reviews of officers' files and remind them that the information within is subject to disclosure. *Brady* sometimes requires disclosure of evidence because it casts doubt on an officer's testimony in a specific case, rather than his truthfulness generally. So ensuring officers remain aware of *Brady*'s demands and obligations guarantees they won't be surprised when the prosecutor discloses. IA Pro could send periodic reminders to internal affairs to conduct these reviews.

Delayed notice harms defendants and police, while early notice gives officers time to correct their behavior. When officers find themselves on *Brady* lists unexpectedly, they cannot always identify what they did to merit the disfavored designation. Delayed notice also creates animosity between prosecutors and police, especially in localities run by progressive prosecutors.[181] Case in point—the Philadelphia FOP's lawsuit.[182] When officers don't know that their files contain *Giglio* material, they risk surprise while testifying at trial. Early warning also gives them a chance to perform mitigating actions to limit the impeaching effect of cross examination focused on their files. Officers more aware of their exposure to disclosure might also behave better, wishing to remain effective witnesses and crime fighters.

Delayed notice makes it harder for prosecutors to do their jobs too. Lawsuits distract from important prosecutorial functions. And in criminal trials, failing to timely notify officers delays prosecutors from disclosing files to defendants. That decreases how much information reaches criminal defense attorneys, reducing the effectiveness of their cross-examinations and limiting their ability to successfully defend their clients.

[180] *See id.* at 779-87 (analyzing the "battle" over *Brady* disclosure splitting officers and prosecutors).

[181] Mary Ellen Reimund, *Are* Brady *Lists (aka Liar's Lists) the Scarlet Letter for Law Enforcement Officers? A Need for Expansion and Uniformity*, 3 INT'L J. HUMANS. & SOC. SCI. 1, 4 (2013), https://www.ijhssnet.com/journals/Vol_3_No_17_September_2013/1.pdf (reviewing how officers in King County, Washington end up on Brady lists but noting that prosecutors' offices leave "due process considerations" to the police agencies, and the *Brady* committee does not reconsider the agency's internal processes). Professor Abel notes the "grave employment consequences" and "potential for police management to misuse" *Brady* lists. Abel, *supra* note 7, at 780-81 (collecting statements from both officers and prosecutors concerned about the potential for abuse of *Brady* lists).

[182] *See* DAO REPORT, *supra* note 11, at 33.

In most contexts, it would not take much to provide officers sufficiently advanced notice for opportunity to appeal. Officers can suffer from two consequences of *Brady* material: placement on a public *Brady* list and a prosecutor refusing to call them as witness. Prosecutors control *Brady* lists and trials move slowly. Neither proceeds with such rapidity to preclude sending formal notice to officers well in advance. Police departments already track officers' activities and log incidents contemporaneously, especially with new software like the NYPD's.[183] Prosecutor's offices can and should immediately alert officers that certain activity—individual incidents or aggregated statistics—may require *Brady* disclosure or cause the department to place the officer on a *Brady* list.

FOIA requests should also trigger immediate notice to any involved officers so that they can seek review of the veracity of the sought information.

5. Anonymized Review

Especially in jurisdictions that restrict access to personnel files outside the police department, software can reveal to prosecutors and defense attorneys only the minimum amount of information necessary. It can redact names, dates, and details from reports and writeups so that the attorneys can determine whether they have to disclose the whole file without intruding at all on officer privacy.

On top of that, digital locks should prevent users from disseminating files without authority. User-tracking and digital watermarks can easily track leaks when information does escape. Once internal affairs identifies which user abused their access to the software and took data, they can immediately shut down that user's access. For a defense attorney, that would hamper their ability to provide effective counsel to their clients, providing a strong incentive against violating the department's terms of use.

CONCLUSION

Criminal defendants need *Brady*. Flawed as it is, *Brady* safeguards their right to a fair trial, ensuring exculpatory evidence is not concealed from the innocent, and preventing untruthful or biased witnesses from presenting unchallenged testimony. Yet, when prosecutors' *Brady* practice uncritically prioritizes the rights of defendants, it risks violating police officers' due process rights. Officers have rights to their employment, reputations, and privacy, all protected by procedural rights to notice and hearing before a neutral arbiter. Constitutional *Brady* practice must contemplate these rights just as it must protect defendants' rights. Where these rights are in tension, police and their unions will resist disclosures and technological changes.

This Comment highlights how thoughtful and considered software can resolve that tension, protecting both parties' rights. These recommendations will benefit prosecutors working to carry out their constitutional duties, defendants putting on a vigorous defense, and importantly, the police officers trying to do their jobs protecting their cities. Used carelessly, software can harm those parties or operate neutrally to their interests. IA Pro's current disclosure function is hardly better for defendants and prosecutors than analog practice. A software application without careful access restrictions and anti-sharing features would allow more disclosure than intended, at

[183] *See supra* text accompanying note 86.

great cost to officers. But informed software development and practice can bring about *Brady* practice fairer and more efficient for all: prosecutors, defendants, and police.

COURTS SHOULD TEST PROMOTION OF GENERAL WELFARE UNDER THE U.S. CONSTITUTION – ON CUTTING MEDICAID PROGRAM TO PAY FOR TAX CUTS BENEFITING PREDOMINANTLY THE WEALTHY AND A NOTE ON TARIFFS; BACK TO THE BASICS, BACK TO THE FUTURE

Rafal Pruchniak[*]

Abstract: Courts indolently and unconstitutionally ignore economics of legislation even if from the beginning it is known that the legislation creates poverty and damage to health contrary to the constitution overarching purpose of promoting general welfare. Courts base its inaction on precedents from times before big data modeling and advancements in behavioral economics made economics an empirically certain science. Given current enormous deficit economic duds cause risks to welfare of generations of Americans which the constitution aims to protect. A note on application to tariffs follows.

Keywords: General Welfare; Promotion; Deficit; Constitution; Poverty; Economic; Legislation; Tariffs

[*] New York University, US.

Table of Contents

I. PROMOTION OF GENERAL WELFARE

The constitution empowers congress to spend and tax for welfare that is general in nature only. Budgeting and cost benefit analysis requirements should be implied.

Article I, Section 8, Clause 1: “The Congress shall have Power To lay and collect Taxes, Duties, Imposts and Excises, to pay the Debts and provide for the common Defence and general Welfare of the United States.”

In its preamble, promotion of general welfare is the essential purpose of the supreme law.

“We the People of the United States, in Order to form a more perfect Union, establish Justice, insure domestic Tranquility, provide for the common defence, promote the general Welfare, and secure the Blessings of Liberty to ourselves and our Posterity, do ordain and establish this Constitution for the United States of America.”

Although, the preamble is non-binding its language has been used interpretatively for understanding Constitution’s broader objectives. In that way preamble expresses purpose but does not provide power.[1]

Nevertheless, spending or taxation that knowingly does not create overall general welfare should be unconstitutional.

This should be especially true if congress arbitrarily cut spending on the poor in order to arbitrarily cut taxes predominantly benefiting the rich; thus both of the cuts in spending and taxes knowingly do not create nor lead to overall general welfare. In other words, these cuts will not promote general welfare.

Logically, it seems courts must test whether there is overall welfare i.e. promotion of welfare which must be general in nature. Saying it differently “wider well-being" result and not only an attempt need to be present or forecasted for spending and taxes to be constitutional. This in and of itself would require objective testing of economic evidence and forecasts, and not mere deference to subjective and wishful political narrative.

II. RATIONAL BASIS TEST DEFERENCE

Under the rational basis test the spending and tax legislation law needs to be rationally related to a legitimate government interest, there must be a rational connection between the legislation and the legitimate purpose, and the purpose be reasonably achievable by the legislation, under this deference standard. The Rational Basis Test stems from the decision in U.S. vs Carolene Products Co., 304 U.S.144 (1938) where the court signaled judicial deference to economic regulations.

However, a tax law cut for the rich only made affordable by a spending cut on the poor that knowingly does not result or will not result in general welfare or even it’s promotion is not constitutional and/or even rationally legitimate. We are dealing here

[1] Jacobson vs. Massachusetts 197 U.S. 11 (1905).

with two steps Article I and rational basis test neither of which is resulting in general welfare or its promotion.

The fact that this legislation is knowingly not resulting in overall general welfare obviously does not reasonably achieve that general welfare nor creates a rational connection with that welfare. To put it plainly a spending cut to pay for tax cut that causes poverty does not create any welfare as poverty is not welfare.

III. OBJECTIVE ECONOMIC REVIEW & BALLOONING DEFICIT

Constitution's general welfare result requirement trumps courts rational basis test.[2] However also, creation of or resulting poverty through spending or taxation does not pass either of the tests: general welfare test or rational basis test, provided there are working alternatives for spending or taxation and there is no necessity for cuts due to unaffordability. Again, rational interest test would only be passed as long as it would result in general welfare, not poverty.

Often overlooked the constitution's preamble also focuses on future generations' welfare. Hence, future general welfare and its forecasts should also matter. This is especially important when reviewing the deficit and making long-term decisions impacting it or connected to it.

Would that mean that any tax cut promotes poverty? No. Just as any spending would not result or promote general welfare today or in the future. Hence, objective economic data analysis and forecasts should used in judicial review. First, to answer whether general welfare will result and secondly if rational basis i.e. legitimate connection between the general welfare result and the legislation test is met.

An important question arises how to test a tax cut for the rich in times of ballooning deficit? For example, economists predict that our current deficit could cause as much as 10% reduction in wage income over next 30 years.[3] Does this mean that the courts should hear economists on this long-term angle too? The author is certain the courts should do exactly that.

Judicial review without deferring to the facts, actual mechanics or objective estimates and forecasts would be no more than symbolic rubber-stamping, of often populist political agenda, that these days can be simply amplified by media and social media.

Judicial branch ruled itself to be an equal branch of the government under the plain language of the Constitution. Namely, under Article III judiciary is established as an equal branch of the government and it would seem courts' deferment to the legislature on policy decisions by not objectively reviewing economics of the tested legislation, resembles more dereliction of court's duties. This is especially true that neither the legislature members nor judges are economists.

[2] Supremacy Clause in Article VI Clause 2 U.S. Constitution.

[3] "Ballooning U.S. budget deficit is killing the American dream"., by Medora Lee www.usatoday.com 06.27.2024 https://www.usatoday.com/story/money/personalfinance/2024/06/27/us-budget-deficit-cuts-incomes/74211526007/

The power of judicial review was affirmed in Marbury vs Madison 5 U.S. (1 Cranch) 137 (1803). However, this was done in times where economics was a theoretical science and not an empirically certain science as it is becoming today, equipped with precision of mathematical modeling on big data and use of advancements in behavioral economics.

Consequently, the courts and legislature should rely on objective economic science and economic experts to promote general welfare. Furthermore, objective economic review would certainly be more consistent with courts currently practiced responsibility of considering animus or bias in legislation in determining its unconstitutionality.

IV. TAX CUTS PAID BY MEDICAID CUTS

Today, Trump administration is considering extension of 2017 tax cuts which majority of benefits went to the wealthy and ultra wealthy.[4] To cover the shortfall in revenue caused by these tax cuts, the administration is considering spending cuts which naturally will impact negatively most the middle class and the poor. The administration is also considering cuts to Medicaid, a healthcare program for the poor impacting rural areas the hardest.[5]

Furthermore, the tax cuts themselves proved not to result in general welfare but rather welfare of the wealthy.[6] The courts should take into account this unfairly biased economic situation towards the middle class and the poor. To do that, it should review economic and social arguments against income inequality and rule the cuts unconstitutional as they simply do not result in general welfare but are rather biased in favor of the already rich. Of course, the legislation should be ruled as constitutional if tax cuts also benefit the middle class and the poor and other balancing measures are passed that promote general welfare such as taxing the ultra rich.

A negative balance on welfare of these tax revenue cuts and spending cuts should be taken into courts decision of such legislation unconstitutionality. Lesser spending cuts which ideally would less negatively impact the middle class and/or the poor, tying or preceding this legislation to additional spending that promotes general welfare such as spending on education or healthcare and offsetting reforms that promote same such as immigration reform would have a positive effect in determining constitutionality of this legislative package.

The courts should also hear expert economist testimony on deficit these tax cuts would generate and how it will particularly impact the middle-class and the poor by forecasting interest rates on credit cards and mortgages to accurately rule on constitutionality of this tax cut legislation which again must result in general welfare –

[4] "The 2017 Trump Tax Law Was Skewed to the Rich, Expensive, and Failed to Deliver on Its Promises"., By Chuck Marr, Samanta Jacoby and George Fenton, 06.13.2024, www.cbpp.org https://www.cbpp.org/research/federal-tax/the-2017-trump-tax-law-was-skewed-to-the-rich-expensive-and-failed-to-deliver

[5] "The most likely Medicaid cuts hit rural areas the hardest." By Scott S. Greenberger. www.medicalexpress.com 03.17.2025 https://medicalxpress.com/news/2025-03-medicaid-rural-areas-hardest.html

[6] "The economic consequences of major tax cuts for the rich"., Socio-Economic Review. 20(2): 539-559. Doi:10.1093/ser/mwab061. ISSN 1475-1461.

please see again preamble to the U.S. Constitution and Article I section 8 of the U.S. Constitution.

It should be noted that so far courts acted quite flexibly in ruling on economic policies constitutionality and were more strict when detecting animus or bias.

However, a policy that from its onset is not known to result general welfare such as tariffs, effectively a tax on the poor, coupled or close in time to spending cuts on things related directly to general welfare, obviously tax cuts for the rich should risk court's review and action.[7]

Therefore, the courts should hear economist panel testimony on broader economic implications of reviewed legislation. This would be a logical expansion of court economic review functions from currently mainly anti-trust cases.

V. AVAILABLE ALTERNATIVE

Spending cuts that limit benefits for the poor such as Medicaid cuts, even if are necessary to balance the budget would still be susceptible if that budget shortfall is caused by tax cuts predominantly benefiting the rich. In this timely example court should rule that the legislature has an option to raise the tax revenue from the rich to pay for the poor and not gut Medicaid. Thus, doing this in an equitable manner by overall creation of general welfare. Had the court allow for Medicaid cuts to balance the tax cuts for the rich it would make the poor poorer and the rich richer thus, not creating or even promoting general welfare.

Note that if the rich would pay their fair i.e. proportionate share of taxes to pay for the Medicaid program for the poor, the rich would have similar tax rate to the poor. This is not the current case as the rich effective tax rate is low in comparison to the general population.[8]

Taxing the rich would create more equity and would be just, as it makes poor less poor while the rich remain still rich. Of course, the courts would also need to take into account optimal tax rate as agreed by economists that would not restrain growth.[9] This court testing activity would clearly be in line with promotion of general welfare which both the legislature and courts are tasked with under the Constitution.

There are, however, several cases baffling the author where Supreme court ignores income inequality despite the Constitutional article requiring general welfare result.

[7] "The Trouble with Tariffs, By David Kelly, Notes on the Weak Ahead"., www.am.jpmorgan.com 03.03.2025 https://am.jpmorgan.com/us/en/asset-management/adv/insights/market-insights/market-updates/notes-on-the-week-ahead/the-trouble-with-tariffs/

[8] "The Forbes 400 Pay Lower Tax Rates Than Many Ordinary Americans", By Seth Hanlon and Nick Buffie, www.americanprogress.org 10.07.2021 https://www.americanprogress.org/article/forbes-400-pay-lower-tax-rates-many-ordinary-americans/

[9] "Krugman on optimal taxes", The Grumpy Economist, https://johncochrane.blogspot.com 01.06.2019 https://johnhcochrane.blogspot.com/2019/01/krugman-on-optimal-taxes.html In that blog economists Diamond and Saez are cited in estimating the optimal top marginal tax rate at 70% a long way to go from current rate of 37%.

On the other hand, there are a couple of notable cases where the courts got on the side of economic fairness against economic disparities. Particularly, in NFIB vs Sebelius court ruled that the government is permitted to use taxation as a policy tool and in that case tax citizens to expand healthcare coverage and thus reduce inequality.[10]

Currently, the government is planning on doing almost exactly the opposite; namely, mechanically reducing taxes for the rich and balancing this revenue shortfall with cuts to the Medicaid's benefits for the poor. Thus, increasing inequality and consequently reducing the general welfare.

NFIB vs Sebelius implies that the supreme court should be interested and can rule on this fact pattern, as well. Merely ignoring it because it renders an opposite result to NFIB vs Sebelius would be illogical and inconsistent to the overarching principle of promotion of general welfare. Furthermore, Medicaid cuts may also lead to similar in size revenue losses at the hospitals for uncompensated care, often shifting and enlarging overall costs to taxpayers through worst health outcomes, higher healthcare prices and healthcare job losses.[11] Medicaid's work requirement did not increase employment either[12]

VI. DEFERENCE OR DERELICTION

To rule on whether policy / legislation results and/or promotes general welfare logically, court has to test the economics of such policy / legislation. Relying on the congress opinion truly does not answer the question at all. Had the court not review the economics of the legislation it would not test at all whether the legislation is constitutional because it would simply not test whether the legislation economics will result or even promote general welfare.

Such answer can only be given by economic subject matter experts and specifically reliance and deference to political narrative is likely not answering the question with proper level of objective necessary expertise.

If the legislation economics are questioned, they can be litigated through sworn economic, subject matter expert testimony just like any other legal issue. The notion the judges are not equipped with economic expertise is thus flawed and deference to the congress on economic policy for the same reason must also be flawed when congress does not rely on subject matter expertise that is forecasted to result in general welfare.

[10] 597 U.S. 519 (2012)

[11] "Medicaid cuts would cost hospitals billions, spike uncompensated care costs: Report", By Alan Condon. www.beckershospitalreview.com 03.11.2025. https://www.beckershospitalreview.com/finance/medicaid-cuts-would-cost-hospitals-80b-in-2026-spike-uncompensated-care-costs-report/ See also "Commentary: Preserve Medicaid funding to safeguard healthcare for our neighbors." By Damond Boatwrite. www.myjournalcourier.com (Last visited 3.17.2025), https://www.msn.com/en-us/health/other/commentary-preserve-medicaid-funding-to-safeguard-health-care-for-our-neighbors-damond-boatwright/ar-AA1AJKX7

[12] "Congressional Republicans Can't Cut Medicaid by Hundreds of Billions Without Hurting People." By Allison Orris and Elisabeth Zhang. www.cbpp.org 03.17.2025 https://www.cbpp.org/research/health/congressional-republicans-cant-cut-medicaid-by-hundreds-of-billions-without-hurting

Moreover, deference on economics of legislation and economic policy to the legislature is not based on any provision in the constitution. This effective division of work was rather invented by the Supreme Court, generations ago when economics where not a highly empirical and certain science, as it is today.

VII. KNOWN TO BE WRONG

The issue whether rational basis test allowing only rational laws, allows laws that do not result or promote general welfare can get more complicated. In our particular example, already comparatively low corporate tax rate is planned to further be lowered and extended, and the consequent shortfall in tax revenue is to be covered by cuts to spending on healthcare for the poor clearly not resulting or promoting general welfare.

A natural question develops again whether an even more competitive tax rate is a legitimate interest when it falls behind the optimal tax rate beyond which it benefits diminish and/or done during ballooning deficit. If the current precedent rules that any legitimate rational interest whether better or not would be constitutional, it does not answer if a known to be completely wrong economically legislation is legitimate as it certainly would not produce general welfare. Given that we already discussed and agreed of court obligation of reviewing economics of legislation what policy sense would it make to approve a known economic dud post review. Thus, lowering tax rate below optimal level would not create overall welfare whether in times of deficit or not.

The administration to be treated seriously on further tax cuts for onshore manufacturing would also have to close tax loopholes incentivizing offshore manufacturing at the least.[13] Bringing manufacturing jobs will not be easy, given high labor costs in U.S., inexistent supply-chain, costly regulations - things that the administration is also not discussing nor has no plan for.[14] It also seems tariffs would increase costs for U.S manufacturing.[15] Last not least higher paid job openings are available right now in the U.S. in services.[16]

We know that from economists who teach us that trickledown economics do not work in converting corporate tax savings into greater employment or salaries.[17] From the beginning it is also well known that the tax rates are already competitive[18] and

[13] "The U.S. Tax System's Curious Embrace of Manufacturing Job Losses", By J. Clifton Fleming, Robert J. Peroni, Stephen E. Shay. www.taxnotes.com, https://www.taxnotes.com/special-reports/base-erosion-and-profit-shifting-beps/u.s-tax-systems-curious-embrace-manufacturing-job-losses/2024/09/30/7lscm 10.01.2024

[14] "Bringing Manufacturing Back To The U.S. Easier Said Than Done." By Guankai Zhai www.forbes.com, 08.28.2024. https://www.forbes.com/councils/forbesbusinesscouncil/2024/08/28/bringing-manufacturing-back-to-the-us-easier-said-than-done/

[15] *Id.*

[16] CNN.com, GPS, By Fareed Zakaria. aired on 03.23.2025. https://www.msn.com/en-us/news/politics/fareed-s-take-manufacturing-is-the-way-of-the-past/vi-AA1BuUHg?ocid=BingNewsSerp

[17] "Trickle-Down Tax Cuts Don't Create Jobs", By Seth Hanlon and Alexandra Thorton. www.americanprogress.org 08.24.2017 https://www.americanprogress.org/article/trickle-tax-cuts-dont-create-jobs/

[18] "International Tax Competitiveness Index 2024" By Alex Mengden, www.taxfoundation.org 10.21.2024 https://taxfoundation.org/research/all/global/2024-international-tax-competitiveness-index/ . Note also the fact that U.S. had a composite marginal effective tax rate of about 11.2% making

further extension of tax cuts covered by cuts to Medicaid will not result or promote general welfare. On the other hand, we also know from economists that higher taxes would not affect negatively economic growth and would make the cuts to Medicaid unnecessary, thus preserving general welfare. Since court would know this from the testimony of experts what policy sense would it make to allow general welfare destruction?

Therefore, even if these tax cuts would be viewed as rational, the fact that the consequent revenue shortfall is covered by spending cuts on the poor and would overall negatively impact general welfare the legislation should be viewed as unconstitutional.

To conclude, it must be emphasized, that promotion of welfare is the preeminent goal of spending and taxing powers and is clearly defined in the constitution. The rational basis test is only stemming from courts interpretation and thus cannot directly contradict defined in constitutional article overall general welfare requirement. Furthermore, how can legislation be rational in example where it from the onset destroys general welfare and other alternatives are available which promote general welfare such as tax raise.

VIII. COURT INACTION

Courts lack of interest in reviewing economic outcomes of legislation may also stem from some justices partisanship caused by how they are appointed. Justices reach old age, due to their lifetime tenures, lack of performance incentives, tight court docket and fewer justices number may also amplify the restraint to review economics of legislation.

However, by no means should this be an excuse especially in situations where legislation is apparently flawed and promotes poverty against one of the main purposes of the constitution.

Previously, nonexistent code of conduct, but currently still non-binding code should have been tied to incentives for objective performance such as hours spent and complexity of cases including complex economic cases.

Failure to critically assess known poverty creating duds resembles dereliction of duty permitting economic inefficiencies often on weakest and poorest members of our society contradicting the very principles constitution seeks to uphold.[19] Courts should be uniquely sensitive to economic populism spread on social media.

it more competitive than the statutory rate of 21% is suggesting. "U.S. Effective Corporate Tax Rate Is Right in Line With Its OECD Peers",
By Daniel Bunn, Garrett Watson. 04.02.2021 https://taxfoundation.org/blog/us-effective-corporate-tax-rate-oecd-peers/

[19] "Takings: Private Property and the Power of Eminent Domain", Richard A. Epstein (1985) (Last visited 03.25,2025) https://doi.org/10.2307/j.ctvjghwth

Another panel of supreme court justices would make the court twice as effective.[20] However, the weakened legitimacy of the court coupled with the battue on civil servants makes its expansion unlikely.

Public outrage should a government made recession be triggered could however, easily push the court itself to require empirical economic analysis as part of rational basis review[21] or even to push the legislature to reform the supreme court. The reform may however, depend on the level of the economic harm public would face which could also be triggered by squandered technological advantage against U.S. biggest competitor China.

The necessary bipartisan cooperation seems unlikely at this moment and will most likely happen after the fact, on heels of an economic crisis or embarrassing science cuts driven competition loss.

Media behavior control capabilities give hope but at the same time surprise with weak patriotism, timid by partisanship and further eroded by highly paid positions tied to ratings of simplified 24/7 content for a society which half is without tertiary education.[22]

IX. WHAT TO DO?

There needs to be a will by the media to report more on economics of legislation its short-term and forecasted long-term effects and court inaction. Economics of legislation should be more of an objective evidence-based topic instead of wishful partisan experiment.

A well-crafted, long and determined media campaign perhaps would generate enough self-reflection at the courts' levels to properly apply the law and review whether there is overall resulting general welfare or even its promotion in legislative agenda while using even the rational basis test.

If not, media should prepare the public for likely recession, loss of competitiveness, but also U.S. debt downgrade, higher interest rates, social security cut and housing unaffordability. But objective and knowledge building economic criticism of populistic policies would allow for fairer democratic process and legislators accountability.

Otherwise, current short-term partisan oversimplifications foretell a bleak economic future if not the end of the republic as we know it, one ignored court order at a time.

[20] "Modern Constitutional Reform- Rebalancing the 3 Branches of Government for Greater Governance Efficiency on U.S.A Example". By Rafal Pruchniak. The International Journal of Law, Ethics and Technology. https://www.ijlet.org/wp-content/uploads/2025/03/IJLET-5.1.pdf (Last visited 03.25.2025)

[21] "Economic Analysis of Law", Richard A. Posner (9th ed.2014)

[22] "Share of adults who have earned a tertiary education in OECD countries in 2022", www.statista.com (last visited on 03/16/2025) https://www.statista.com/statistics/1227287/share-of-people-with-tertiary-education-in-oecd-countries-by-country/

The positive resolution of the need for courts economic review could perhaps see a self-imposed minimum number of complex cases, including economic cases on the dockets of justices as a start of a more prosperous tomorrow.

X. STRICT SCRUTINY

Medicaid cuts can disproportionately affect protected classes, including racial minorities and individuals with disabilities.[23] Courts may apply strict scrutiny if Medicaid cuts disproportionately or intentionally harm protected classes violating the equal. protection clause.[24] It has to be emphasized that poverty is not a suspect classification.[25]

Under strict scrutiny government must show legislation serves compelling interest, is narrowly tailored and uses least restrictive means.

However, all this does not mean the government can further impoverish Medicaid recipients through these cuts. This would be against the welfare clause as discussed earlier.

Would enormous deficit trump the strict scrutiny? - not under the least restrictive means prong as the fiscal goal of reducing debt could be achieved by raising taxes especially in a more economically optimal way thus also satisfying the general welfare clause.

Only cuts narrowly tailored could proceed if they would impact wasteful spending.

The fact that deficit reduction is a compelling government interest would not impact the two other strict scrutiny prongs discussed above.

As you can see economic analysis is needed to be reviewed by the legislature, the courts and the public to better fine tune our democratic process which was designed in the first place as a gentlemen agreement where all parties work on achieving prosperity in respect for one another needs and individual situations applying the cost benefit analysis and budgeting.

A. Media Campaign and Overturning Precedent

Overturning court precedents can happen when societal and legal interpretations evolve. The courts slowly but surely evolved to protect vulnerable economically members of our society over the decades. A question before us now is whether courts would be ready to catch up by a technologically driven leap of considering economics as a certain science based on empirical big-data.

[23] "Medicaid Cuts Would Rip Away Health Coverage from Millions of Americans, disproportionately Harming People of Color." unidous.org 03.13.2025 https://unidosus.org/publications/medicaid-cuts-would-rip-away-health-coverage-from-millions-of-americans-disproportionately-harming-people-of-color/, See also "Medicaid at Risk: What Cuts Mean for People with Disabilities – and All of Us.", By Jackie Dilworth. thearc.org 01.03.2025 https://thearc.org/blog/media-memo-medicaid-at-risk/.

[24] See Dekker v. Weida 679 F.Supp. 3d 1271 N.D. Fla. 2023.

[25] San Antonio Independent School District v. Rodriguez 411 U.S. 1 (1973)

This of course will naturally be challenged by politics, justices preferences and customs but it shouldn't. Again, cost-benefit analysis is emphasized in the preamble of constitution and mandated by the spending clause.

Therefore, critical to this process will be a media campaign raising economic awareness and interest. Proving economic science reliability often against social media trends and intentional social media amplifications.

The justices would ultimately analyze whether review of objective economics would further politicize the court and if the lack of accountability due to justices life tenures vs terms of legislators would be optimal.

However, in authors opinion there would be less friction between political bias and empirically driven on big data economic science. This opinion needs to be seconded by all of us, however, biases already exist and what is lacking and would be truly transformative, is a scientific quality of economical legislation of the future.

B. Concerns Over Presidential Use of Tariffs as a Tax

Administration's use of tariffs as a broad revenue-raising mechanism exceeds presidential authority under current trade statutes and constitutional law. However, such use is functionally equivalent to taxation and thus subject to constitutional constraints, including the General Welfare Clause and the nondelegation doctrine. The author argues that testing such tariffs also under the General Welfare Clause would provide a more rational, constitutional, and policy-sound framework for evaluating tariff actions.

Under Article I, Section 8, Clause 1 of the U.S. Constitution, Congress has the exclusive power to "lay and collect Taxes, Duties, Imposts and Excises," and to "regulate Commerce with foreign Nations." Tariffs traditionally fall under this power when used to regulate international trade. Tariffs are a form of impost or duty imposed on imports, historically used for both revenue and regulatory purposes. Under Section 7 of the same article, "All Bills for Raising Revenue shall originate in the House of Representatives…". This reflects the framers' intent to keep the taxing power in Congress, not the President.

In U.S. v. Hvoslef, 237 U.S. 1 (1915), the Supreme Court recognized that tariffs may have both revenue and regulatory purposes. However, when tariffs function primarily to raise general revenue, they must conform to constitutional taxation rules, including those under the General Welfare Clause see United States v. Butler, 297 U.S. 1 (1936).

C. Tariffs Power Delegation

It has to be emphasized here that Congress has delegated only a limited tariff authority to the President through several statutes: Trade Expansion Act of 1962, 19 U.S.C. § 1862 (Section 232) – permits the President to impose tariffs if the Secretary of Commerce finds that imports threaten national security, Trade Act of 1974, 19 U.S.C. § 2411 (Section 301) – allows retaliatory tariffs in response to foreign unfair trade practices, and Tariff Act of 1930, 19 U.S.C. § 1336 (Section 336) – authorizes tariff adjustments to equalize costs of production.

These statutes have been upheld under the nondelegation doctrine because they include an intelligible principle to guide executive action. See J.W. Hampton Jr. & Co. v. United States, 276 U.S. 394 (1928). However, when executive action lacks a genuine connection to the statutory goals such as national security or trade retaliation the delegation may no longer be valid.[26] Broad, indiscriminate tariffs not tied to these principles risk violating the nondelegation doctrine.

D. Tax or Tariff

Courts assess whether a government measure is a tax based not on form but on function. In National Federation of Independent Business v. Sebelius, 567 U.S. 519 (2012), the Court emphasized that what matters is how the measure operates in practice, not how it is labeled.

If a tariff applies to all imports, without case-specific findings, is justified primarily by fiscal needs rather than trade goals, is projected as revenue in the federal budget and lacks a clear statutory link to national security or trade remedies then it functions as a tax and must be treated as such under the Constitution. The fact that the current administration tariff formula penalizes trade deficit has more to do with customer preferences or trading partners countries development stage and their wealth rather than national security concerns.

The General Welfare Clause, Article I, Section 8, Clause 1, limits federal taxation to purposes that promote general welfare. While the Court has afforded Congress broad discretion here (United States v. Butler, 297 U.S. 1 (1936)), executive actions do not enjoy the same deference particularly when Congress did not even expressly authorize the tax to be discussed in more detail below.

Since the administration uses tariffs to also generate revenue rather than only regulate trade, and in a way that imposes burdens on U.S. consumers and businesses, then the tariffs may not always meet the general welfare requirement which would consequently need to be evaluated. Moreover, Congress (and Courts) would have to be the body to determine (and review) based on facts whether such a tax promotes the general welfare not the President acting unilaterally.

Recent practice suggests that tariffs are being used as a de facto revenue tool. Office of Management and Budget projections include hundreds of billions of tariff revenues as part of the federal budget.[27] Public statements by officials, including the Presidents, frame tariffs as a means to fund U.S. programs and reduce deficits. Moreover, tariffs have been imposed indiscriminately across countries and goods, without clear security justification under discussed Section 232 or Section 301.

This pattern undermines any claim that the tariffs are narrowly tailored trade measures. Instead, the facts and circumstances strongly suggest that the executive is exercising general taxing power without even a delegation from Congress.

[26] https://www.law.cornell.edu/wex/nondelegation_doctrine?utm_source=chatgpt.com (Last visited 04.09.2025).

[27] See also “State of U.S. Tariffs: Week of April 7, 2025.” The Budget Lab. https://budgetlab.yale.edu/research/state-us-tariffs-week-april-7-2025?utm_source=chatgpt.com (Last visited 04.12.2025.)

If the President's tariff actions are primarily for revenue generation, this raises constitutional concerns: 1. Nondelegation Violation – Tariff imposition beyond intelligible statutory limits could invalidate the delegation itself. 2. Encroachment on Congressional Taxing Power – Only Congress can impose taxes; the President's actions may usurp that power. 3. Failure to Promote General Welfare – Taxes must serve a public interest, and these tariffs may fail that test. A challenge could invite the Supreme Court to revisit nondelegation, clarify when a tariff becomes a tax and whether it satisfies general welfare clause.

It is likely that if the Congress could explicitly authorize the President such use of tariffs in taxation i.e. revenue-raising activity, the President's delegation would be constitutional provided that the Congress lays down an intelligible principle under J.W. Hampton, Jr & Co. v. US. 276 U.S. 394 (1928). However, under the discussed Article 1 Section 8 and 7 revenue-raising is also treated as a core and exclusive function of the Congress thus the Congress cannot abdicate this power completely disallowing unbounded or standardless delegation under Skinner v. Mid-America Pipeline Co., 490 U.S. 212 (1989). Nevertheless, under United States v. Butler, 297 U.S. 1 (1936) the court ruled that the taxing power is limited to serving general welfare and as such revenue raising tariffs would be tested.

So whether such delegation be constitutional or not it would still very likely have to satisfy the general welfare clause. However, most recently, Jaime Diamond the CEO of JP Morgan Bank warned (together with many other economists) that tariffs will slow growth and increase prices on consumers thus clearly failing the general welfare clause.[28]

It is uncertain whether the Court would review presidential executive order imposing tariffs on all countries on a country-by-county basis and invalidate some and approve others that meet national security or trade retaliation statutory delegations.

It is more logical that courts would review tariffs on a country-by-country basis, would that, however, mean that such particular tariffs would need to satisfy the welfare clause? Author is of opinion that dual-purpose tariffs that are regulatory and revenue raising in nature, effect cannot run in contradiction of also the Welfare Clause in addition to other constitutional provisions like the Commerce Clause or Equal Protection.

They also must serve legitimate public interest, which as discussed earlier in this article is not general poverty creation, and of course not be arbitrary and purely punitive. This stems from the ruling in United States v.s. Butler where the court emphasized that Congresses power to impose tariffs is not unlimited and the ends must be constitutional and means appropriate.

When a tariff enacted by Congress serves not only a regulatory function under the Commerce Clause but also raises revenue, it implicates Congress's taxing power under Article I, Section 8, Clause 1. Because the taxing power is an independent constitutional grant, separate from and not limited by the scope of other enumerated

[28] "Jamie Dimon Warns Tariffs Will Raise Prices, Slow Growth - JPMorgan CEO says in annual letter he hopes for long-term benefits, but that many uncertainties exist." By Candice Choi, Updated April 7, 2025 https://www.wsj.com/economy/jamie-dimon-warns-tariffs-will-raise-prices-slow-growth-8b82baaf?st=3qjHCT

powers, any measure partially enacted under it—including dual-purpose tariffs—must satisfy the General Welfare Clause.[29]

Again, Congress may impose tariffs under the taxing power of Article I, Section 8, Clause 1 which gives Congress the power "to lay and collect Taxes, Duties, Imposts and Excises," so long as such measures promote the general welfare and are uniform. Although, the Commerce Clause of Article I, Section 8, Clause 3 allows Congress to regulate trade with foreign nations - tariffs are "duties" and "imposts," explicitly included within the scope of the taxing power. When tariffs raise revenue, even if they also regulate trade, they should trigger obligations under the General Welfare Clause.

The Supreme Court has made very clear that the taxing power stands on its own constitutional footing and is not dependent upon or limited by Congress's other enumerated powers. In United States v. Butler, 297 U.S. 1 (1936) the Court declared:

> "The power of taxation which is granted by the General Welfare Clause is not restricted by the limitations imposed on the use of other powers specifically granted."

In other words, Congress can tax for purposes of the general welfare even where it could not regulate directly under the Commerce Clause or another power. Note, the Butler Court invalidated a spending program because it used the taxing power to coerce state agricultural practices—thus violating federalism, even though the tax was for a claimed public benefit.

In NFIB v. Sebelius, 567 U.S. 519 (2012) the Court reaffirmed that a law valid under the taxing power need not be valid under the Commerce Clause. It upheld the individual mandate as a lawful tax, despite it failing under Commerce Clause scrutiny, emphasizing that the taxing power is subject only to its own internal limits, such as the General Welfare Clause. This is because the taxing power is not constrained by other powers, it is subject only to its own limits—most centrally, the requirement that it serve the general welfare.

When Congress imposes a tariff that: raises revenue, and serves a regulatory purpose (e.g., protecting industry, shaping trade behavior), it is exercising both its Commerce Clause and Taxing Clause powers. Because the taxing power is independent, the measure must independently satisfy the requirements of that power:

Under the general welfare clause the revenue raised must at least promote general welfare, if not overall create such general welfare and be used in service of the nation's general welfare, not for coercive or punitive ends.

Furthermore, Under the Uniformity Clause: Tariffs must be applied uniformly across the United States. However, more importantly under the Anti-Coercion Principle: If the tariff is so onerous as to function as a penalty rather than a tax, it may be struck down as a disguised regulation.[30] This prevents Congress or the President

[29] OpenAI ChatGPT. https://chat.openai.com. Response to author's guiding prompts. 04.19.2025.

[30] In Bailey v. Drexel Furniture Co., 259 U.S. 20 (1922), the Court struck down a tax imposed to regulate child labor, holding it was a penalty masquerading as a tax.

from using tariffs as a regulatory end-run, imposing economic burdens for policy purposes without meeting the constitutional standards of taxation.

To sum up the taxing power is not merely a support to other enumerated powers—it is a distinct and independent constitutional authority. As such, any federal measure—including tariffs—that partially relies on the taxing power must comply with the General Welfare Clause. This means that dual-purpose tariffs, which both regulate trade and raise revenue, cannot escape constitutional scrutiny. They must be structured and justified in a way that demonstrably at least promotes if not creates the general welfare of the United States, or they risk being declared unconstitutional under longstanding Supreme Court doctrine.

Regarding policy justification, this framework would (1) restrain any government branch overreach or populism, (2) ensure economically rational, welfare-enhancing tariff policy. and (3) It would also create judicially reviewable standards to curb erratic or politically motivated tariff actions.

The fact that the current back and forth inconsistent announcements of tariffs by President Trump led to a market crash, downward spiral in consumer spending sentiment, and over 50% - 70% likelihood of a recession provides a clear answer that the overall effects of these tariffs contradict the General Welfare Clause, serve an illegitimate public interest of general poverty creation, and many of them are simply punitive and they are unlikely to lead to increased manufacturing in the U.S.[31] It is important to note here that the President also attacked industrial policy championed by his predecessor's administration, a policy that has proven to increase domestic manufacturing, thus showing there is no consistent policy and genuine interest in manufacturing creation beyond predominantly revenue generation.[32]

Additionally, the steep percentage of the tariff levied tips the scale on their strong revenue-raising effect. In case of tariffs on China the very steep barrier could be simply cost prohibitive in using those inputs in U.S. manufacturing. [33] It would also invite due process concerns; The Fifth Amendment prohibits arbitrary government actions that lack a rational basis. and the embargo like 145% tariff imposed without economic justification could be challenged as an irrational and punitive act that "shocks the conscience."[34]

In conclusion, it appears that courts would likely rule that President tariffs violate the General Welfare Clause if challenged in court given the gravity of Presidents misguided policy or rather lack thereof and its financial ruinous effects. The General Welfare Clause should serve as an objective guiding principle for any branch of the

[31] Bailey v. Drexel Furniture Co., 259 U.S. 20 (1922).

[32] Over 33,000 jobs created by the Chips Act - see "Chips and Science Act: Breaking down the law's impact 2 years later." www.manufacturingdive.com by Joelle Anselmo. Published on 07.26.2024. https://www.manufacturingdive.com/news/semiconductor-chips-and-science-act-investments-impact/720235/

[33] 49% of the total imports from China that were subject to Section 301 tariffs were intermediate goods used in U.S. manufacturing in 2021. Reducing these tariffs would make US products more competitive and spur growth. See "Section 301 China Tariffs by End Use", By Tom Lee and Tori Smith. Published on 01.11.2023 https://www.americanactionforum.org/research/section-301-china-tariffs-by-end-use/?utm_source=chatgpt.com

[34] BMW of North America v. Gore*, 517 U.S. 559 (1996) illustrates how disproportionate economic penalties can violate substantive due process.

government and any leader whether on the left, center or right. The General Welfare Clause was analyzed here from perspective of spending and tax cuts, and tariffs, however, its role could be much broader and only our imagination could be its limit. Author wishes here of course the Supreme Court sees transformative potential this objective clause could have on quality of work coming from our representatives of course regardless of their party affiliation. This approach is both doctrinally sound and policy-wise prudent. It aligns constitutional interpretation with the realities of modern economic governance and reinforces the necessary checks and balances that preserve the rule of law. Author wishes also foreign scholars reading this article can draw comparative ideas on discussed U.S. policy mistakes (of the century) and rethink whether their laws have backstops against their homegrown economic populism.[35]

[35] The author acknowledges the use of OpenAI's ChatGPT for assistance in generating preliminary ideas and clarifying conceptual distinctions during the writing process, All interpretations and conclusions are author's own.

AN ANALYTICAL STUDY OF FEDERATED LEARNING-ENHANCED NATURAL LANGUAGE PROCESSING FOR PRIVACY-CENTRIC NATIONAL CYBERSPACE ID AUTHENTICATION

Zihan Zhang[*]

Abstract: The rapid growth of digital technology has amplified concerns about data privacy and cybersecurity, necessitating innovative solutions to protect personal information. This study explores the potential of Federated Learning (FL) to enhance privacy-focused national Cyberspace ID authentication. Recently, China put forward an initiative to utilize Cyberspace ID to address data privacy concerns, which has caused a heated debate about the legal and technological credibility of this project. To enable a better understanding of the possible risks and significance of Cyberspace ID, this essay first examines the legal landscape of data privacy, comparing frameworks such as the GDPR, U.S. sectoral laws, and China's cybersecurity policies. After that, this essay advocates for a possible solution to enhance Cyberspace ID system resilience by implementing Federated Learning algorithms and discusses how this aligns with the legal regulations. This interdisciplinary analysis highlights the potential of Federated Learning to advance cybersecurity and data privacy in the digital age.

Keywords: Data Privacy; Federated Learning; GDPR

[*] The University of Hong Kong, China.

Table of Contents

INTRODUCTION

The growth of digital technology has accelerated social, economic, and cultural developments but also brings complex ramifications for human rights that require careful consideration.[1] As data is becoming the centric resource of the digital economy and information society, data privacy and personal data protection have become increasingly significant. Digital footprints created by online activities, including social media interactions, internet searches, and online transactions, can be utilized to recognize, comprehend, and forecast unique behavioral patterns in people, putting individual privacy frequently at risk. In addition, data privacy issues are interconnected with cybersecurity. Personal data breaches may lead to fraudulent activities and criminal actions. In order to effectively manage the growing number of digital threats to human rights, it is imperative that the right to privacy, which is guaranteed by Article 12 of the Universal Declaration of Human Rights,[2] Article 17 of the International Covenant on Civil and Political Rights,[3] and numerous other international and regional human rights instruments, be respected and protected.[4] For example, the General Data Protection Regulation (GDPR) put forward by the EU is the toughest privacy and security law in the world.[5]

China is considering establishing the Cyberspace ID in order to protect data privacy and cybersecurity. The Cyberspace Administration of China and the Ministry of Public Security have released draft regulations that define Cyberspace ID as a series of encrypted numbers that serve as a user's identification for online authentication. Online service providers will no longer be able to access sensitive personal data and actual human identities since the national authentication platform will take over the authentication process and only return the identity verification findings. Evolved from real-name registration strategy in China, the initiative aims to protect citizens' personal information by implementing a trusted online identity strategy. The centralized data control approach, however, may increase the danger of data breaches, scalability issues, and Cyberspace ID re-identification.

Regarding the challenges that Cyberspace ID may face, this article will discuss how an advanced technological solution: Federated Learning (FL), can facilitate effective privacy protection by using a decentralized data processing method. Federated Learning is a distributed machine learning approach that collaboratively runs algorithms over several dispersed edge devices or servers while keeping the raw data on-device.[6] Google first proposed FL in 2016 to allow Android phone users to upgrade models locally without disclosing sensitive personal information.[7] After that, Google put in place an FL system designed to run federated average (FedAvg) algorithms on

[1] Riduan Siagian et al., Human Rights in the Digital Era: Online Privacy, Freedom of Speech, and Personal Data Protection, 2 Journal of Digital Learning and Distance Education 513-523 (2023).
[2] Universal Declaration of Human Rights, G.A. Res. 217A (III), U.N. GAOR, 3d Sess., pt. I, U.N. Doc. A/810 (1948)
[3] International Covenant on Civil and Political Rights, Dec. 16, 1966, 999 U.N.T.S. 171
[4] Office of the United Nations High Commissioner for Human Rights, *The Right to Privacy in the Digital Age* (Aug. 20, 2022), https://digitallibrary.un.org/record/3985679.
[5] Razieh Nokhbeh Zaeem & K. Suzanne Barber, The Effect of the GDPR on Privacy Policies: Recent Progress and Future Promise, 12 ACM Transactions on Management Information Systems (TMIS) 1-20 (2020).
[6] Jie Wen et al., *A Survey on Federated Learning: Challenges and Applications*, 14 International Journal of Machine Learning and Cybernetics 513-535 (2023).
[7] Id.

mobile devices. This system could be used to track statistics for large-scale cluster equipment without storing raw data on a cloud server. Since then, FL has emerged as one of the privacy computing industry's most concerning technologies. This article will discuss how FL can enhance the privacy protection of the Cyberspace ID system and conform with the data privacy regulations proposed by the current legal framework.

I. THE LEGAL LANDSCAPE OF DATA PRIVACY PROTECTION

Personal data (also used as "personal information") is the type of data that can not only be related to but also be used to recognize a specific individual.[8] The substantial factor of personal data is recognition, which emphasizes that anybody (not only the data controller) could adopt a rational methodology to recognize the identity of an individual.[9] The formal factor of personal data is digitally recorded and retrievable.[10]

The lawful rights of personal data are closely related to the field of property rights and privacy rights, but with distinct focuses and characteristics as well. Based on the monetary value of personal data and their similar attributes with property rights, some scholars argue the rationality of recognizing anonymized non-identifiable data as intangible assets.[11] In addition, there are assumptions of introducing the concepts of inalienability, user-transfer restriction, and opt-in default to propertize personal data.[12] From another perspective, personal data also evolves people's understanding about privacy in the information society. William L. Prosser has categorized four types of privacy torts: intrusion upon seclusion, publicity given to private life, false light publicity, and appropriation of name or likeness.[13] However, these four torts are limited and narrow when facing information privacy issues. While the original concern about privacy is the subjective factor that the parties do not want to disclose, the concern about the identification of personal information is whether the specific individual can be recognized objectively and does not involve the subjective factor of the parties.[14] The right to personal information not only has the right of elimination but also has the right to know, the right to correct, the right to delete, the right to block, and other positive functions that privacy rights do not have.[15]

The legal protection of personal information shows different characteristics across jurisdictions. In the U.S. legal system, personal information protection extends from the right to privacy. Alan F. Westin first defined the right to informational privacy as the right of a natural person to decide when, how and to what extent personal information will be disclosed to others.[16] This theory is strengthened by the case of Whalen v. Roe (1977), in which the Supreme Court extended the substantive due process protections of privacy to encompass informational privacy, thereby affirming

[8] Directive 95/46/EC of the European Parliament and of the Council of 24 October 1995 on the protection of individuals with regard to the processing of personal data and on the free movement of such data, 1995 O.J. (L 281) 31.

[9] Yuan He, *Data Law* 32-44 (1st ed. 2020).

[10] Id.

[11] Feng Xiong et al., Recognition and Evaluation of Data as Intangible Assets, 12 Sage Open 1-13 (2022).

[12] Paul M Schwartz, *Property, Privacy, and Personal Data*, 117 Harvard Law Review 1-10 (2020).

[13] William L. Prosser, *Privacy*, 48 California Law Review 389 (1960).

[14] Yuan He, *Data Law* 32-44 (1st ed. 2020).

[15] Id.

[16] Alan F Westin, *Privacy and Freedom* 7 (1st ed. 1967).

an individual's right to manage their personal information disclosure.[17] Consequently, the U.S. personal data protection framework has advanced from the scope of privacy to informational privacy, and then transformed to the constitutional right to informational privacy through the Federal Supreme Court's judicial interpretations of provisions of the Bill of Rights Act of the Constitution (He 2020). In addition to this bottom-up development process of personal information law, the United States common law legal system also features a sector-specific data protection that ranges from health, education, to finance, accompanied by consumer protection laws such as the California Consumer Privacy Act (CCPA).[18] For example, the Health Insurance Portability and Accountability Act (HIPAA) is a federal law that regulates the privacy of health information and clarifies financial penalties according to the level of culpability and types of violations.[19] In 2013, a malicious employee from Montefiore Medical Center, a non-profit hospital system, unlawfully accessed the medical records of 12,517 patients, copied their information and sold them to identity thieves.[20] Montefiore Medical Center was investigated and determined "failed to conduct an accurate and thorough risk analysis of the potential risks and vulnerabilities to the confidentiality, integrity, and availability of ePHI; failed to implement procedures to review records of activity in information systems, and failed to implement hardware, software, or procedural mechanisms to record and examine activity in the information system".[21] Montefiore was eventually fined $4.75 million in 2024 and monitored to implement a corrective action plan.[22] The case of Montefiore indicates how sectoral laws implement the principles of processing personal data lawfully, ensuring fairness, and maintaining transparency within a specific sector. However, the U.S.' multifaceted landscape of data privacy protection law is also doubted by lacking a comprehensive federal data privacy law and therefore relying on a mix of federate and state law, leading to fragmented consent requirements, data breach notifications, and enforcement and penalties.[23]

The data privacy protection landscape of EU members is different from the United States for bearing hybrid features of Common Law and Civil Law. The German legal system of personal data protection stems from the extension of general personality rights instead of the right to privacy. The German Federal Constitutional Court has interpreted Article 1, paragraph 1, of the German Basic Law, the "Human Dignity Clause", and Article 2, paragraph 1, of the German Basic Law, the "Free Development of the Personality", to give specific content to informational self-determination: the "general personality right" of the German Basic Law includes the protection of personal data from unrestricted extraction, storage, and continued transmission.[24] This fundamental right guarantees an individual the right to self-determination, disclosure, and use of their personal data only. Since then, the concept and term of informational

[17] Whalen v. Roe, 429 U.S. 589 (U.S. Supreme Ct. 1977).

[18] Vivek Krishnamurthy, *A Tale of Two Privacy Laws: The GDPR and the International Right to Privacy*, 114 AJIL Unbound 26-30 (2020).

[19] Seun Solomon Bakare et al., *Data Privacy Laws and Compliance: A Comparative Review of the EU GDPR and USA Regulations*, 5 Computer Science & IT Research Journal 528-543 (2024).

[20] Steve Alder, *Malicious Insider Incident at Montefiore Medical Center Results in $4.75 Million HIPAA Penalty*, The HIPAA Journal, (Feb. 7, 2024).

[21] Id.

[22] Id.

[23] Seun Solomon Bakare et al., *Data Privacy Laws and Compliance: A Comparative Review of the EU GDPR and USA Regulations*, 5 Computer Science & IT Research Journal 528-543 (2024). https://doi.org/ 10.51594/csitrj.v5i3.859.

[24] Yuan He, *Data Law* 32-44 (1st ed. 2020).

self-determination have flourished in European legal thought and have emerged as one of the conceptual underpinnings of the right to personal data protection ensured by Article 8 of the Charter of Fundamental Rights of the European Union.[25] Consistent with the rationale of informational self-determination, the European Union put forward the General Data Protection Regulation (GDPR) in 2018, which is the toughest privacy and security law in the world and represents some key data protection principles: Lawfulness, fairness, and transparency; Purpose limitation; Data minimization; Accuracy; Storage limitation; Integrity and confidentiality; Accountability. GDPR also clarifies three participant roles: Data Subject, Data Controller, and Data Processor, and assigns obligations for these roles to abide by the data protection law.[26] GDPR grants data subjects explicit rights to access, the right to rectification, and the right to erasure. In addition, under the concept of data protection by design and by default, data controllers and processors are obliged to a implement privacy control framework throughout the process.[27] For example, Google was alleged and fined roughly $57 million by CNIL, the French Data Protection Authority, for violating GDPR regulations. Google was criticized for not obtaining users' consent to process data for advertisement personalization, not clearly revealing the purpose of utilizing users' data, and failing to carry out de-referencing of sensitive data, which violates the GDPR principles of lawfulness, fairness, and transparency; purpose limitation; and the data subject's right to be forgotten. As indicated by the case, the EU GDPR provides a comprehensive approach to data privacy protection, emphasizes explicit consent and grants individuals a robust right to withdraw. Different from U.S. privacy law that places the default position of the law as "permit", GDPR presumes the default position of the law as "prohibit" and requires lawful consent from the users before the personal data may be collected, used, or disclosed.[28] Therefore, even though there is criticism about the ineffectiveness of GDPR in terms of its limited material scope,[29] GDPR still represents the toughest data privacy laws and forward-looking legal regulations in terms of data privacy protection, which evaluates the outcomes of other evolving personal data protection approaches.

II. PROPOSED CYBERSPACE ID SOLUTION IN DATA PRIVACY PROTECTION

A. Background: Cybersecurity Governance in China

Since 1994 when the Internet was introduced in China, China's cybersecurity policy development has generally undergone four stages each with distinct policy focuses [30]: (1) Initial stage (1994-1999): construction of Internet infrastructure; (2) Rapid development (2000-2004): multi-layered information service (3) Adjustment and

[25] Florent Thouvenin, *Informational Self-Determination: A Convincing Rationale for Data Protection Law?*, 12 JIPITEC 246-256 (2021).

[26] Razieh Nokhbeh Zaeem & K. Suzanne Barber, The Effect of the GDPR on Privacy Policies: Recent Progress and Future Promise, 12 ACM Transactions on Management Information Systems (TMIS) 1-20 (2020).

[27] Seun Solomon Bakare et al., *Data Privacy Laws and Compliance: A Comparative Review of the EU GDPR and USA Regulations*, 5 Computer Science & IT Research Jounal 528-543 (2024).

[28] Vivek Krishnamurthy, *A Tale of Two Privacy Laws: The GDPR and the International Right to Privacy*, 114 AJIL Unbound 26-30 (2020).

[29] Michaela Padden & Andreas Öjehag-Pettersson, *Protected How? Problem Representations of Risk in the General Data Protection Regulation (GDPR)*, 15 Critical Policy Studies 486-503 (2021).

[30] Cyberspace Administration of China, *20 Years of China Internet: Cyber-Security* (1st ed. 2014).

optimization (2005-2013): information security (4) In-depth improvement (2014-now): "Cyber Power": the trend of integrating cybersecurity agenda with the agenda of national strategic development [31].With the issuance of *National Cybersecurity Strategy* and *Chinese Cybersecurity Law* as important nodes, China views cybersecurity as a cornerstone of its national security, societal stability, and economic development.

China's approach to balancing individual rights and state power in the realm of cybersecurity and data governance differs from those in the EU and the U.S. The EU emphasizes individual rights and data protection, as exemplified by the GDPR. The GDPR enforces strict rules on data collection, processing, and storage to prioritize individual privacy over state control, reflecting a rights-based governance model. The U.S. adopts a decentralized, market-driven approach to data security by using a patchwork of federal and state laws rather than a unified framework like the GDPR. This approach leaves much of the responsibility to private companies and emphasizes economic freedom. In contrast to EU and U.S., China's governance model features state-centric control. From the Chinese government's perspective, cybersecurity is not only about protecting individuals but also about maintaining control over the cyberspace, which includes preventing the misuse of digital platforms for disinformation, dissent, or other activities perceived as threats to state security.

Due to the different emphasis on individual rights and public power, China's cyberspace protection methods differ from those of the EU and U.S., which derive the concept of personal data privacy from the right to privacy, but evolve the physical identity card to the real-name authentication in cyberspace. In order to promote a safer and healthier Internet and safeguard the public interest and social order from unlawful content, including libel, fraud, pornography, rumors, and vulgarity, Chinese national legislation has mandated since 2012 that the majority of online service providers use real-name registration [32]. According to the law entitled Decision of the Standing Committee of the National People's Congress on Strengthening Online Information Protection, network operators should require users to provide real identity information when providing corresponding network services [33], which elevates China's real-name policy to the level of national law. Chinese individuals typically give identity verification to online platforms in the form of mobile phone number verification because these numbers must be obtained and linked to a real name. As a result, the telecom operator controls the actual identity information, which other links' network services may utilize inadvertently for verification.

Although China's real-name registration policy aims to safeguard cybersecurity, it may be critiqued for harming individuals' privacy. Anonymity is a form of privacy protection that allows people to speak freely without having to submit to public identification. From another perspective, personal data privacy can be invaded by anonymous net citizens when they enjoy the freedom to express themselves with a low

[31] Zhengrong Li et al., *A Study of Chinese Policy Attention on Cybersecurity*, 69 IEEE Transactions on Engineering Management 3739-756 (2022).

[32] Jyh-An Lee & Ching-Yi Liu, *Real-Name Registration Rules and the Fading Digital Anonymity in China*, 25 Washington International Law Journal 1-33 (2016).

[33] Quangguo Renda Changweihui Guanyu Jiaqiang Wangluo Xinxi Baohu De Jueding [Decision of the Standing Committee of the National People's Congress on Strengthening Online Information Protection] (promulgated by the Standing Comm. Nat'l People's Cong., Dec. 28, 2012, effective Dec. 28, 2012), art. 6, (China), http://www.gov.cn/jrzg/201212/28/content_2301231.htm

sense of accountability [34]. The balance of personal data privacy and general governance of cybersecurity is a critical issue in China that requires careful consideration. Therefore, evolving from the background of real-name registration, Cyberspace ID tries to strike a more proper balance.

B. Mechanisms of Cyberspace ID

From the technology offering perspective, cyberspace ID uses both static and dynamic authentication methods to safeguard personal data privacy. The authentication process comprises three different entities: the claimant, the monitor, and the information system [35] .For Cyberspace ID, citizens would be the claimant who authenticates to the system in order to use the service. The national authentication app would be the monitor that checks the claimant's identity. Apps and online platforms would be the information systems that provide the services if the monitor correctly authenticates the claimant. In addition, the Cyberspace ID combines the usage of static and dynamic authentication protocols. While static authentication relies on fixed credentials for identity verification, dynamic authentication uses changing or one-time credentials to enhance security and reduce the risk of unauthorized access [36] . Cyberspace ID is a fixed set of generated numbers that are not associated with the identity information but can be matched with the individual. The static number can be displayed or reported when online service providers require to confirm that an individual is a user with authentic identity through the feedback of the national authentication platform. In offline scenarios, Cyberspace ID will adopt a dynamic authentication method by randomly generating dynamic two-dimensional code for identity verification to avoid screenshots and identity fraud.

The issuance of Cyberspace ID could advance the protection of personal data privacy compared with the current real-name registration system in China. Firstly, it centralizes the authentication process from telecom operators and online platforms to a national authentication platform. Online platforms only receive the identity authentication result instead of the actual identity of users, avoiding issues with online service providers collecting personal data beyond scope or retaining data for longer than required. It safeguards individuals' data privacy by ensuring data minimization, purpose limitation, and storage limitation. Secondly, compared with real-name registration, Cyberspace ID is not associated with the identity information of individuals. Generally, when analyzing user data to derive user behavioral characteristics, what Internet platforms and enterprises derive is only the behavioral characteristics themselves (some mathematical vectors) and cannot be backtracked to a specific individual. Even if it corresponds to an individual, it will only correspond to the cyberspace ID rather than the original identity information (e.g., ID card number, and biometrics information). Thirdly, compared to using phone numbers to trace back identity, Cyberspace ID is a more trustworthy verification certificate. In China, many telecom fraud gangs buy mobile phone numbers to open a large number of online accounts or registered companies for fraudulent purposes. If these online accounts are opened through a Cyberspace ID rather than tied to a phone number, it could curb

[34] Jyh-An Lee & Ching-Yi Liu, *Real-Name Registration Rules and the Fading Digital Anonymity in China*, 25 Washington International Law Journal 1-33 (2016).

[35] Syed Zulkarnain Syed Idrus et al., *A Review on Authentication Methods*, 7 Australian Journal of Basic and Applied Sciences 95-107 (2013).

[36] Id.

telecom fraud and raise the cost of crime.

However, there remain risks and challenges to implementing cyberspace ID. Firstly, there are inherent privacy risks of centralized governance of Cyberspace ID. In a centralized system, all data is stored in a single repository and under the control of one authority, which is more vulnerable to malicious attacks and a single point of failure. Centralized systems may also face scalability issues and struggle to handle billions of users efficiently. Secondly, threats of re-identification from Cyberspace ID remain. Even if cryptography is employed, there are still chances that personal information could be re-tracked by certain techniques. Users may be sorted into different target groups and be bothered by unethical advertisements. More severely, sensitive personal data can be leaked.

III. FUTURE ADVANCEMENT OF CYBERSPACE ID: DECENTRALIZED DATA PRIVACY GOVERNANCE

Decentralized data systems offer enhanced data privacy compared with centralized data systems. Instead of depending on a single central repository, decentralized data systems are made to disperse data processing and storage among several sites or nodes [37]. The distributed control of data empowers data subject to manage their data. By enforcing granular access controls, users can determine who can access their data and under what conditions [38]. This reduces the risk of unauthorized access and data breaches.

Apart from privacy concerns, decentralized data systems feature a stronger security proof. Decentralized systems lessen the possibility of a single point of failure by distributing data among multiple nodes. The overall system is improved since the remaining network continues to function even if one node is compromised or fails. In addition, mechanisms like consensus algorithms and encryption techniques are frequently incorporated into decentralized systems to guarantee that data is transparent and impenetrable [39]. Users' trust is strengthened since participants can confirm transactions and data integrity without having to rely on a central authority [40].

Therefore, in the context of Cyberspace ID, the proposed centralized data governance system can be enhanced into a decentralized method. However, decentralized data governance may lead to another problem: the fragmentation of databases (“data silos”) and the inefficiency of conducting user behavior analysis. Being closely related to the field of psychology, behavioral analysis initially centered on the study of human behavior, applying scientific methods to comprehend human conduct [41]. IT companies aggregate a large amount of data and derive general users' behavior patterns to drive decisions. These companies may also gather data across different platforms to develop user portraits and provide targeted advertisements. If the

[37] Moritz Platt, Ruwan J. Bandara, Andreea-Elena Drăgnoiu, & Sreelakshmi Krishnamoorthy, Information Privacy in Decentralized Applications, in Trust Models for Next-Generation Blockchain Ecosystems (Muhammad Habib ur Rehman et al. eds., EAI/Springer Innovations in Communication and Computing, 2021).

[38] Id.

[39] Id.

[40] Haleh Asgarinia et al., *“Who Should I Trust with My Data?” Ethical and Legal Challenges for Innovation in New Decentralized Data Management Technologies*, 14 Information 351 (2023).

[41] Alejandro G. Martin et al., *A Survey for User Behavior Analysis Based on Machine Learning Techniques: Current Models and Applications*, 51 Applied Intelligence 6029-6055 (2021).

data are controlled in a decentralized manner, integrating information to derive meaningful insights will become more challenging. To address this, Federated Learning (FL), a new technique in the field of NLP, could possibly balance privacy protection and user behavior analysis demand, contributing to the general cybersecurity issue and advancing Cyberspace ID deployment.

IV. FEDERATED LEARNING IN DECENTRALIZED PRIVACY PROTECTION AND USER BEHAVIOR ANALYSIS

Federated Learning is a distributed collaborative learning approach that enables joint modeling while safeguarding data privacy and security [42]. FL allows algorithms to be executed and trained on local nodes such as smartphones, laptops, and wearable devices, using local datasets stored only on that single device [43]. Every device first downloads a global model for local training, then refines the downloaded global model through several local training using individual device data, uploading the associated gradient information to the cloud, which then combines the averaged updates of local models to create a new global model that is sent back to the devices. This iterative process is repeated until the model reaches the target performance level. This assembles playing a Pictionary game. Each player (device) draws interpretations of user behaviors and shares drawings (parameters) with a guesser (central server). The guesser can aggregate drawings and make guesses without knowing the original prompt (user behavior data). Instead of transmitting the original user data to a central server, only the training results (the parameters) would be exchanged and used to calculate the global model in Federated Learning, which greatly protects privacy[44]. In general, Federated Learning advances machine learning by keeping the raw data in-device and extends the boundary of distributed learning as it could work with unbalanced and non-independent identically distributed data (non-IID)[45]. In this case, Federated Learning is a cross-disciplinary technique of computer science that enables data privacy and data sharing for decentralized devices.

Federated Learning could safeguard privacy while satisfying the need for user behavior analysis in a decentralized method. Different from centralized governance of Cyberspace ID stored and processed by the national authentication platform, a Federated-Learning-enhanced Cyberspace ID can function well in a decentralized method. By leveraging the computational power of user devices, identity verification and Cyberspace ID number can be generated on-device. In addition, information about user behaviors and interactions with online services could be gathered to train the local model and send updates to the central server, while the raw data are kept on-device and not revealed to the national authentication platform. In this case, the FL-enhanced Cyberspace ID system could derive insights from user behavior analysis, detect early threats from online platforms, and also safeguard personal privacy at the same time. For example, if multiple devices detect abnormal login patterns, the FL-trained model can learn from these patterns collectively without requiring centralized access to sensitive data. The re-identification risks of Cyberspace ID can also be lowered after iterative FL training and early threat detection. This strategy aligns with the general

[42] Id.

[43] Id.

[44] Nguyen Truong et al., *Privacy Preservation in Federated Learning: An Insightful Survey From the GDPR Perspective*, 110 Computers & Security 1-19 (2021).

[45] Id.

principle of cybersecurity by ethically protecting personal data privacy from IT companies and authorities.

As an advanced privacy protection technique, Federated Learning could comply with most of GDPR regulations about data privacy protection. Given that Federated Learning only aggregates locally trained parameters for global model updates and cannot be exploited for other purposes, Federated Learning complies with the principles of "purpose limitation". The data are all on-device, aligning with the principles of "data minimization" "storage limitation" and "accuracy". The integrated security techniques in Federated Learning, such as Secure Aggregation, Homomorphic Encryption, and other secure communications protocols, fulfill the requirements of "integrity and confidentiality" and "accuracy". The remaining requirement that Federated Learning has difficulty satisfying is "fairness and transparency" because, like other deep learning algorithms, Federated Learning is operated in a black-box feature [46]. However, it is a common problem for machine learning that there is limited understanding and transparency of how certain decisions are made. In general, Federated Learning is an advanced technique that could balance dynamic big data analysis and data privacy regulations. For example, Google is the first one to propose Federated Learning to comply with GDPR regulations. Google uses this technology to enhance its ads deployment on search engines and content recommendations.

In addition, Federated Learning can fit for the data privacy protection in the U.S. legal framework. For example, Federated Learning can preserve data privacy in the healthcare industry and comply with sectoral data privacy regulations such as HIPAA. Traditionally, collaborative healthcare research requires establishing generalizability and external validity by sharing patient data between institutions, which can violate the patients' right to their healthcare data privacy [47]. In contrast, by utilizing Federated Learning, it is possible to train the local models of different healthcare centers respectively by keeping the standardized health record data on device [48]. In this case, without sharing raw ePHI with other institutes, Federated Learning lowers the risk of leaking healthcare data during the transmission process and protects the individuals' rights to "direct a covered entity to transmit to a third party an electronic copy of their protected health information in an electronic health record", and the rights to "request corrections" as regulated by the HIPAA privacy rules. Given the evidence from the healthcare industry, implementing Federated Learning in different sectoral settings can similarly comply with the U.S. data privacy legal framework.

V. CASE STUDY: HOTEL RESERVATION PLATFORM MASSIVE DATA BREACH

Prestige Software's main product Cloud Hospitality is a channel manager that connects online reservation websites (e.g. Booking.com and Expedia) with hotels' software to enable online management of room availability and vacancy. In 2020, Website Planet revealed that Prestige Software has been exposing highly sensitive data

[46] Nguyen Truong et al., *Privacy Preservation in Federated Learning: An Insightful Survey From the GDPR Perspective*, 110 Computers & Security 1-19 (2021).

[47] Tyler J. Loftus et al., *Federated Learning for Preserving Data Privacy in Collaborative Healthcare Research*, 8 Digital Health 1-5 (2022).

[48] Id.

from millions of hotel guests worldwide since 2013 [49]. It was estimated that around 24.4 GB of data and totaling 10 million files have been exposed, covering customer data ranging from PII (Personal Identifiable Information), reservation details, to credit card and payment details [50]. Based in Spain, an EU country, Prestige Software must follow the regulations of GDPR and may face legal actions and huge fines because as a data processor, it violates the terms of storage limitation, integrity, and accountability. The high severity of this incident indicates the potential flaws in personal information protection in cyberspace and represents the requirements of advanced data protection techniques.

This massive data breach represents the potential risks of cloud storage security. Cloud Hospitality connects with various hotel booking websites and stores the data on Amazon Web Services (AWS) S3 bucket which provides cloud-based data storage, and the data leakage results from the misconfiguration of the AWS S3 bucket. Cloud storage data security includes static storage security and dynamic storage security, representing the cloud storage server security and data transmission confidentiality respectively. Given that data is transmitted through the IP network in the cloud storage, the cloud storage system will also be vulnerable to traditional network security threats such as data destruction, data theft, data tampering, etc. Furthermore, in cloud storage systems, users' data may be dispersed among several servers, and multiple users may share one server, raising the danger of unwanted unauthorized access [51]. There are various techniques to safeguard data security for cloud storage. For example, data encryption technology (identity-based encryption, attribute-based encryption, and homomorphic encryption etc.), data loss prevention (DLP) tools, and multi-factor authentication (MFA) are all the regular methods used to protect data security [52]. However, the massive data breach of Cloud Hospitality indicates the instability and vulnerability of cloud storage despite these protections. Therefore, along with the rise of cloud computing, it requires advancements in data protection techniques to tailor to the trend of increasing data exchanges and transmissions among different online servers.

Federated Learning can be a possible solution for this requirement. In the case of Cloud Hospitality, the platform doesn't employ any encryption or other security protection methods before transmitting and storing the data into AWS S3 bucket, making it highly vulnerable and risky to data breach. In contrast, utilizing Federated Learning can lower the risks of data leakage even when the cloud storage platform is misconfigured because the raw data are kept on-device. All the training processes such as updating the booking status of the hotel's rooms and analyzing hotel customers' preferences can be fulfilled by using the raw data on the hotel's own software. Only the training results instead of raw data (national ID, credit card numbers, and reservation details) will be transmitted to Cloud Hospitality and the cloud storage platform for model updates. In this case, even if the cloud storage platform faces data leakages, only the parameters which are some mathematical vectors will be exposed, instead of putting sensitive personal data at risk. This approach also conforms with the GDPR principles,

[49] Website Planet Security Team, *Report: Hotel Reservation Platform Leaves Millions of People Exposed in Massive Data Breach*, Website Planet (June 11, 2020), https://www.websiteplanet.com/blog/prestige-soft-breach-report/.

[50] Id.

[51] Pan Yang et al., *Data Security and Privacy Protection for Cloud Storage: A Survey*, 8 IEEE Access 131723-131737 (2020).

[52] Id.

especially in terms of data minimization and storage limitation.

The case of Cloud Hospitality echoes the current case of Cyberspace ID initiated by China’s government. Hotels preserve vast amounts of personal data, especially those that are highly sensitive and relevant to personal privacy. The personal identifiable information and credit card payment details in the case of Cloud Hospitality align with the protection object of Cyberspace ID initiative. In addition, characteristics of cloud storage platforms also reflect the increasing requirements of advanced data protection among frequent data transmissions in multi-platforms. Thus, this case study indicates the potential of Federated Learning-enhanced natural language processing as a prospective technique in the field of privacy protection.

CONCLUSION

In the information era, one of the basic human rights: privacy has evolved into the requirement of personal information protection. Legal regulations revolving around data privacy protection have been put forward with distinctive characteristics. Recognizing the significance of data privacy protection and aligning with legal requirements, Cyberspace ID is put forward as a possible solution. However, a national authentication strategy itself couldn't safeguard data privacy and requires further advancement. In this case, Federated Learning is a prospective technique that could safeguard personal privacy in a decentralized data control manner of Cyberspace ID. Further research is required to delve deeper into the detailed mechanics of implementing Federated Learning into Cyberspace ID. Firstly, legal concepts about the parameters incurred in Federated Learning need to be defined and it requires legal analysis and case studies to investigate whether these mathematical vectors should be regarded as privacy data. In addition, further experiments need to be conducted to verify the feasibility and reliability of a Federated Learning-enhanced national authentication platform. There may be risks that user behavior analysis via Federated Learning can lead to re-identification of users and invalidate the cyberspace ID. It will be an advancement in cybersecurity if this cross-disciplinary field combining legal regulations and machine learning can successfully develop.

DECIPHERING THE HERO VILLAIN NARRATIVE: A FUNCTIONALIST COMPARISON OF AI GOVERNANCE IN THE U.S. AND CHINA

Zhenzhen Zhan[*]

Abstract: The previous comparative studies on artificial intelligence (AI) governance between the U.S. and China have primarily focused on the differences between the two countries and their ideological antagonism. This paper aims to delve deeper into this issue by addressing the following questions: (1) what are the key differences in AI governance between the U.S. and China? (2) Are these differences rooted in fundamental distinctions such as ideology, or are they pragmatist responses to differing stages of AI development? To answer these questions, this study includes a broader range of policy documents related to AI governance from both countries for a more thorough comparison. The legal instruments compared include 36 federal and 25 state-level documents from the U.S., with a portion referenced in the annex, and 38 from China. Furthermore, this paper employs a functionalist comparative approach to analzse the policy documents included. In this vein, this paper categorizes the aforementioned legal instruments into three groups – facilitation, regulation, and international cooperation – based on the roles played by their rules, and examines the specific measures for AI governance in both countries. The findings demonstrate that the differences in the two countries' approaches can largely be attributed to their respective stages of technological development—the U.S. is focused on "maintaining leadership," while China is focused on "catching up." Despite these differences, both place considerable emphasis on the economic and strategic benefits brought by technological advancements, while relatively underestimating the potential risks.

Keywords: AI; Comparative Law; U.S.; China; Functionalist

[*] Civil, Commercial and Economic Law School, China University of Political Science and Law, China.

Table of Contents

INTRODUCTION

In recent years, the U.S. and China have emerged as major rivals of a full-blown competition in AI.[1] As the two primary leaders and competitors in AI, they adopt distinct strategies for AI development and governance.[2]

Regarding these distinctions, some scholars argue that the competition for AI leadership is an ideological confrontation,[3] with China's AI development seen as reinforcing authoritarian control[4] and threatening democracy and international security.[5] These narratives frame the U.S.-China AI race as a battle between democracy and authoritarianism,[6] portraying it as a clash of civilizations[7] where the U.S. must prevail to defend freedom and values.[8] Others compare the U.S. and China on ethical philosophies[9] and measures against AI-related challenges[10] in specific areas like military,[11] education[12] and technological development[13] and how they balance

[1] Because there is no universally accepted and authoritative definition of artificial intelligence, this paper does not aim to establish a definitive definition or outline the scope of comparison. Instead, it focuses on examining how AI is characterized and conceptualized within the policies of the two states. *See*: Haroon Sheikh, Corien Prins & Erik Schrijvers, *Artificial Intelligence: Definition and Backgro*und, i*n* MISSION AI: THE NEW SYSTEM TECHNOLOGY 15, 15 (Haroon Sheikh, Corien Prins & Erik Schrijvers eds.,2023); PETER NORVIG & STUART RUSSELL, ARTIFICIAL INTELLIGENCE: A MODERN APPROACH (4th ed. 2021).

[2] Graham Allison & Eric Schmidt, IS CHINA BEATING THE U.S. TO AI SUPREMACY?, THE BELFER CENTER FOR SCIENCE AND INTERNATIONAL AFFAIRS (Aug., 2020), https://www.belfercenter.org/publication/china-beating-us-ai-supremacy.

[3] Jing Cheng & Jinghan Zeng, *Shaping AI's Future? China in Global AI Governance*, 32 J. CONTEMP. CHINA 794, 807 (2023); Kerry McInerney, *Yellow Techno-Peril: The 'Clash of Civilizations' and Anti-Chinese Racial Rhetoric in the US–China AI Arms Race*, 11 BIG DATA SOC.1, 2 (2024).

[4] Karman Lucero, *Artificial Intelligence Regulation and China's Future*, 33 COLUMBIA J. ASIAN LAW 94, 114 (2019); Jinghan Zeng, *Artificial Intelligence and China's Authoritarian Governance*, 96 INT. AFF. 1441, 1441-42 (2020)

[5] Courtney Manning, *CODE WAR: How China's AI Ambitions Risk U.S. National Security*, AMERICAN SECURITY PROJECT 1, 10 (Oct. 17, 2023), https://www.americansecurityproject.org/perspective-code-war-how-chinas-ai-ambitions-threaten-u-s-national-security.

[6] Nike Retzmann, *'Winning the Technology Competition': Narratives, Power Comparisons and the US–China AI Race*, in COMPARISONS IN GLOBAL SECURITY POLITICS 237, 244-245 (Thomas Müller, Mathias Albert & Kerrin Langer eds., 2024).

[7] Kerry McInerney, *Yellow Techno-Peril: The 'Clash of Civilizations' and Anti-Chinese Racial Rhetoric in the US–China AI Arms Race*, 11 BIG DATA SOC.1, 2 (2024).

[8] Alfred D. Hull et al., *Why the U.S. Must Win the Artificial Intelligence (AI) Race*, 7 CYBER DEF. REV. 143, 150-151 (2022).

[9] Emmie Hine & Luciano Floridi, *Artificial Intelligence with American Values and Chinese Characteristics: A Comparative Analysis of American and Chinese Governmental AI Policies*, 39 AI SOC. 257, 268-70 (2024); Emmie Hine, *Governing Silicon Valley and Shenzhen: Assessing a New Era of Artificial Intelligence Governance in the U.S. and China*, 3 DIGIT. SOC., at 1, 15-18 (2024).

[10] Yoshija Walter, *Managing the Race to the Moon: Global Policy and Governance in Artificial Intelligence Regulation—A Contemporary Overview and an Analysis of Socioeconomic Consequences*, 4 DISCOV ARTIF INTELL 14 (2024).

[11] Maria Bega, *The New Arms Race between China and the US: A Comparative Analysis of AI-Powered Military and Economic Pursuits*, 17 EUR. CONTIN. CHANGE EUR. GOV. 75, 76-77 (2023).

[12] Dahlia Peterson, Kayla Goode & Diana Gehlhaus, *AI Education in China and the United States*, CENTER FOR SECURITY AND EMERGING TECHNOLOGY (Sep., 2021), https://cset.georgetown.edu/wp-content/uploads/CSET-AI-Education-in-China-and-the-United-States-1.pdf.

[13] Daniel Castro, *Who Is Winning the AI Race: China, the EU or the U.S.?*, CENTER FOR DATA INNOVATION (Aug. 19, 2019), https://datainnovation.org/2019/08/who-is-winning-the-ai-race-china-the-eu-or-the-united-states.

technological advancements and regulation.[14] These studies are often theme-specific, using selected legal instruments or frameworks for comparison, which fail to provide a comprehensive view. In a nutshell, existing studies fail to adequately address the rationale behind such differences. By emphasizing ideological opposition,[15] some studies unintentionally show inherent ideological biases.[16] The real question is whether there are any intrinsic ideological differences in their AI governance, and if not so, what the underlying causes are.

This paper aims to offer a more nuanced analysis of global AI governance by examining the policy frameworks and implementation practices of the two countries. It aims to mitigate the ideological opposition and potential biases, arguing that the regulatory differences between the U.S. and China are primarily due to technological disparities while highlighting the fundamental similarities in their regulatory strategies and key interests.

The paper broadens the scope of comparison to include various forms of "soft law." While not legally binding, these instruments have practical and legal effects, offer flexibility, and reflect social norms, especially as both countries adopt a gradual approach to AI governance.[17] Specifically, Chinese legal instruments include laws, administrative regulations, departmental rules, and influential "red-headed documents" issued by the central government, along with other policies.[18] U.S. legal instruments encompass state-level legislation, government agency rules, and guiding principles, including executive orders.[19] To gather data, the authors searched official U.S. federal and state government websites,[20] obtaining 36 federal documents and 25 state documents.[21] For Chinese legal instruments, the search was conducted through the "PKULaw" database (https://www.pkulaw.com/) and supplemented by the Compilation

[14] William Howey, *How Governments Are Looking to Regulate AI*, ECONOMIST INTELLIGENCE UNIT (July 21, 2023), https://www.eiu.com/n/how-governments-are-looking-to-regulate-ai/; Morgan Sullivan, *Global AI Regulation: A Closer Look at the US, EU, and China*, Data Privacy Infrastructure, https://transcend.io/blog/ai-regulation#china.

[15] Alfred D. Hull et al., *Why the U.S. Must Win the Artificial Intelligence (AI) Race*, 7 CYBER DEF. REV. 143, 145 (2022); Lucero, *supra* note 4, at 167-171; Manning, *supra* note 5, at 1-13.

[16] Hine and Floridi, *supra* note 9 at 268; Hine, *supra* note 9 at 9, 18.

[17] Francis Snyder, *The Effectiveness of European Community Law: Institutions, Processes, Tools and Techniques*, 56 THE MODERN LAW REVIEW 19, 32 (1993).

[18] Due to the substantial influence exerted by the Chinese government, normative documents that lack legal binding force play a crucial role in shaping government regulations and business practices in reality. Therefore, excluding such documents from the discussion would render the comparison almost meaningless.

[19] Although some U.S. bills, such as the *Testing and Evaluation Systems for Trusted Artificial Intelligence Act of 2023* and *Algorithmic Justice and Online Platform Transparency Act*, are still under review and have not yet come into effect, they are crucial for understanding the future trajectory of AI legislation and are therefore included.

[20] Using the keywords "AI" "Algorithm" and "Data Privacy", relevant policy documents and bills were searched on the official websites of the U.S. Government (https://www.congress.gov) and the White House (https://www.whitehouse.gov). Legislative reports from the National Conference of State Legislatures (https://www.ncsl.org) were also consulted as supplementary sources to ensure comprehensive information collection.

[21] Some state laws are listed in the annex, while the detailed analysis in the main text primarily focuses on federal regulations.

of Generative AI Laws,[22] yielding a total of 38 documents.

Concerning the theoretical framework, this paper employs *Legal Functionalism* for comparison,[23] which is grounded in the concept of "Functional Equivalence" introduced by Zweigert and Kötz. This concept suggests that while legal systems may differ in rules and procedures, they can be considered functionally equivalent if they serve the same social or legal purposes. Given the comparable challenges AI presents across national legal systems, this framework is well-suited to analyze how different legal systems respond to AI's disruptions. This functionalist perspective helps uncover the underlying logic behind these responses.

When examining the specific functions of AI legal responses, it becomes clear that they serve a dual role: facilitation and regulation[24]—— Legal adjustments facilitate AI development, while also addressing the risks and disruptions AI poses to the social order through regulation.[25] Furthermore, international coordination is discussed separately due to its distinct policy goals, particularly the prominent emphasis on national interests instead of the AI industry only. To clarify this distinction, this paper categorizes the AI-related legal responses into three components: facilitative law, regulatory law, and law of international coordination. Within this framework, legal instruments are further subdivided into areas such as infrastructure, human capital, ethics, and algorithm security, with each category explained to highlight its specific functional differences.

A closer analysis illustrates that the key differences lie in government roles, regulatory frameworks, and policy implementation. The underlying causes of the differences stem from the countries' respective stages of technological development. The U.S. focuses on maintaining its leadership position in AI, while China is determined to close the technological gap. These differences, therefore, are not rooted in abstract factors such as ideology, but rather in the respective stages of technological progress. However, both states prioritize technological progress over regulation in their governance strategies, with a stronger emphasis on growth than on security. It is also crucial to note that, if not carefully managed, the divergent approaches may push AI development beyond safe and acceptable limits, with far-reaching implications for international AI governance.

This paper is structured as follows: Sections II, III, and IV provide an in-depth exploration of the facilitative and regulatory laws, as well as strategies for international cooperationof both the U.S. and China, offering a comprehensive overview of their AI

[22] He Yaqi (贺雅琪), *Generative AI Laws, Regulations, and Policies Compilation Package* (生成式 AI 法律法规政策汇编大礼包), PKULAW WISDOM LEGAL PERSONNEL INSTITUTE (北大法宝智慧法务研究院) (Feb. 7, 2024), https://mp.weixin.qq.com/s/bX4fw-0THucfHDDICEWrpw.

[23] Konrad Zweigert & Hein Kötz, *An Introduction to Comparative Law* 34-35 (Tony Weir trans., 3d ed. 1998); Max Rheinstein, *Teaching Comparative Law*, 5 UNIV. CHIC. LAW REV. 615, 618 (1938) ("In spite of many national differences, modern civilization creates essentially the same problems everywhere.").

[24] Lucero, *supra* note 4, at 95.

[25] Angela Huyue Zhang, The Promise and Perils of China's Regulation of Artificial Intelligence, COLUM. J. TRANSNAT'L L. (forthcoming), https://ssrn.com/abstract=4708676; Pierre Lepaulle, The Function of Comparative Law with a Critique of Sociological Jurisprudence, 35 HARV. LAW REV. 838, 845 (1922) ("Law is, in one sense, a social medicine." "the legal machinery of a given society is very much like a living body with its reactions, its currents, its temperament, its prejudices; that it is extra-sensitive to certain things, blind to others.").

governance policies. These sections also present a comparative analysis of their policy preferences and development trends across three key areas. Section V examines the paths each country has chosen to achieve their AI governance objectives, analyzing the underlying reasons for these divergent approaches.

I. THE U.S. AND CHINA'S APPROACH TO AI FACILITATION

The facilitative law is defined as being positively contributed to the development of the AI industry, in terms of its intended objectives or actual outcomes. To fulfill its goal, the measures taken are multifaceted, including the formulation of comprehensive development plans, the promotion of AI applications across various scenarios, and the provision of support for talent, computing power, data, and other essential resources.

A. Analysis of Facilitative Legislation in the U.S. and China

1. Development Plans and Objectives

After realizing AI's potential, China has set overall goals and a developing roadmap, trying to reverse China's backward status. On July 8, 2017, China's State Council unveiled the *New Generation of Artificial Intelligence Development Plan*[26], the first systematic strategic plan for AI released by China since the turn of the century. This document starts by describing the strategic landscape of AI development, and then puts forward a three-step strategic goal:

> *(1) 2020: AI technologies and their applications will align with the world's advanced levels, and the AI industry will become a significant new engine of economic growth;*
>
> *(2) 2025: Significant breakthroughs in AI fundamental theories will be made, with several technologies and applications reaching the world's leading levels, and AI will serve as the main driving force for China's industrial upgrading and economic transformation;*
>
> *(3) 2030: AI theories, technologies, and applications will be world-leading, and China will emerge as a major global center of AI innovation.*

To materialize this goal, the *NGAIDP* puts forward primary tasks in three dimensions: technology, economy, and society. Meanwhile, it proposes the fundamental principles of systematic planning and a market-driven approach, balances the roles of government and market, and enables the government to play a better role in planning and guidance, policy support, security precautions, market supervision, environment building, and formulation of ethical laws and regulations. The aforesaid goals remain

[26] Xu Xuechen(许雪晨), Tian Kan (田侃) & Li Wenjun (李文军), *Xinyidai Rengongzhineng Jishu(AIGC): Fazhanyanjing*、*Chanyejiyu Ji Qianjingzhanwang*((新一代人工智能技术（AIGC）：发展演进、产业机遇及前景展望)[*New Generation of Artificial Intelligence Technology (AIGC): Development, Industrial Opportunities, and Future Outlook*], 2023 Chanye Jingji Pinglun(产业经济评论)[REV. IND. ECON.] no. 4 at 5, 6.

unchanged and have not been replaced by new ones so far.[27]

What is also particularly striking among China's legal instruments is that China is trying to make detailed arrangements focusing on critical theories and technological directions, targeting to gain advantages through the development of planned key directions. China has always attached great importance to the planning of research on fundamental theories and technologies and has facilitated the development by emphasizing critical directions for in-depth research.[28] The *Interim Measures for the Management of Generative Artificial Intelligence Services*, while primarily focused on AI safety, also place significant emphasis on the independent innovation of foundational technologies.[29]

Unlike China's approaches to setting specific stage-based targets, the U.S. emphasizes core goals for AI development in key documents and values like transparency, equity, accountability, and public trust, directing the government to balance advancing American innovation with protecting civil liberties, while minimizing barriers to AI adoption to drive innovation.[30] Similarly, the *National Artificial Intelligence Research and Development Strategic Plan*[31] highlights the need for equitable, transparent, and auditable AI technologies. The principles outlined in the above instruments are intended to guard against potential dangerous tendencies and ensure that AI develops in a positive direction, rather than to formulate a detailed roadmap for its development. This reflects a fundamental difference in the two countries' approaches and highlights the core distinction regarding government planning in fostering the development of AI.

2. Sector-Specific Applications of AI

A prominent feature of China's facilitative law is its strong focus on promoting AI across a wide range of application scenarios. To this end, China has issued a comprehensive array of legal instruments related to AI applications, 16 in total, accounting for 66.7% of all facilitative documents. As early as 2016, China issued the *Three-Year Action Plan for "Internet Plus" AI*, and proposed facilitating AI innovation

[27] Xinyidai Rengongzhineng Fazhan Guihua (新一代人工智能发展规划) [New Generation Artificial Intelligence Development Plan] (promulgated by the State Council, July 20, 2017), https://www.gov.cn/zhengce/content/2017-07/20/content_5211996.htm. [hereinafter *NGAIDP*]

[28] State Council, *supra* note 27; "Shisan Wu" Guojia Zhanlüexing Xinxing Chanye Fazhan Guihua ("十三五"国家战略性新兴产业发展规划) [The 13th Five-Year Plan for the Development of National Strategic Emerging Industries] (promulgated by the State Council, Nov. 29, 2016), CLI.2.286929 (Lawinfochina); Zhonghua Renmin Gongheguo Guomin Jingji he Shehui Fazhan Di Shishi Ge Wunian Guihua he 2035 Nian Yuanjing Mubiao Gangyao (中华人民共和国国民经济和社会发展第十四个五年规划和 2035 年远景目标纲要) [The Outline of the 14th Five-Year Plan (2021-2025) for National Economic and Social Development and Vision 2035 of the People's Republic of China] (promulgated by the Nat'L People's Cong., Mar. 11, 2021), CLI.1.353607 (Lawinfochina).

[29] Shengchengshi Rengongzhineng Fuwu Guanli Zanxing Banfa (生成式人工智能服务管理暂行办法) [Interim Measures for the Management of Generative Artificial Intelligence Services] (promulgated by the National Development and Reform Commission(NDRC) et al., July 10, 2023, effective August 15, 2023) Lawinfochina, CLI.4.5171165.

[30] Exec. Order No. 13,859, 84 Fed. Reg. 3967 (Feb. 14, 2019) [hereinafter *EO 13859*].

[31] *National Artificial Intelligence Research and Development Strategic Plan: 2023 Update*, Networking and Information Technology Research and Development Program (May 2023), https://www.whitehouse.gov/wp-content/uploads/2023/05/National-Artificial-Intelligence-Research-and-Development-Strategic-Plan-2023-Update.pdf . [hereinafter *AI Strategic Plan 2023*].

products in key sectors.[32] The *NGAIDP,* making systematic guidance for the application of AI,[33] gives a full picture of the Chinese government's future plan on AI utilization.

AI applications in the service industry cover a wide range of sectors, including education, healthcare, and elderly care,[34] with detailed plans proposed, e.g., the specific integration of AI with green and low-carbon industries in the energy sector.[35] Concerning social governance, the proposal advancing the "intelligent transformation of governance" entails enhancing administrative efficiency through AI, thereby facilitating a more responsive governance structure. This transformation encompasses the development of "smart government services," "smart courts," "smart cities," and "smart monitoring platforms" aiming to improve transparency, accountability, and public engagement.

Moreover, other specific provisions cover the application of AI in various sectors, including the food industry for the regulation of food safety, geological surveying to enhance earthquake disaster response,[36] forestry and grassland safety management, as well as disaster prevention and mitigation.[37] Given the frequency of document releases, AI application is the most thoroughly executed aspect of the *NGAIDP*,[38] reflecting the Chinese government's strong emphasis on it.

In a stark contrast, the U.S. executive orders call for practical use to ensure equitable AI accessibility, especially protecting vulnerable groups, but do not provide detailed requirements for widespread AI deployment. *Executive Order 14110 on the*

[32] "Hulianwang+" Rengongzhineng Sannian Xingdong Shishi Fang'an ("互联网+"人工智能三年行动实施方案) ["Internet Plus AI" Three-Year Action Implementation Plan] (promulgated by the NDRC et al., Ministry of Science and Technology, Ministry of Industry and Information Technology, and Cyberspace Administration of China, May 1, 2016) https://www.gov.cn/xinwen/2016-05/23/content_5075944.htm. [hereinafter *Internet Plus*]

[33] State Council, *supra* note 27.

[34] Guanyu Jiakua Changjing Chuangxin Yi Rengongzhineng Gaoshui Ping Yingyong Cujin Jingji Gao Zhi Liang Fazhan de Zhidao Yijian (关于加快场景创新以人工智能高水平应用促进经济高质量发展的指导意见) [Guiding Opinions on Accelerating Scenario Innovation and Promoting High-quality Economic Development with High-level Application of Artificial Intelligence] (promulgated by the Ministry of Science and Technology et al., Jul. 29, 2022), CLI.4.5132750 (Lawinfochina); Guanyu Zhichi Jianshe Xin Yidai Rengongzhineng Shifan Yingyong Changjing de Tongzhi (关于支持建设新一代人工智能示范应用场景的通知) [Notice on Supporting the Construction of New Generation Artificial Intelligence Demonstration Application Scenarios] (promulgated by the Ministry of Science and Technology, Aug. 12, 2022), CLI.4.5132812 (Lawinfochina).

[35] Guanyu Tuidong Nengyuan Dianzi Chanye Fazhan de Zhidao Yijian (关于推动能源电子产业发展的指导意见) [Guiding Opinions on Promoting the Development of the Energy Electronics Industry] (promulgated by the Ministry of Industry and Information Technology et al., Jan. 3, 2023), CLI.4.515037 (Lawinfochina).

[36] State Council,*supra* note 27.

[37] Guanyu Cujin Linye He Caoyuan Rengongzhineng Fazhan de Zhidao Yijian (关于促进林业和草原人工智能发展的指导意见) [Guiding Opinions on Promoting the Development of Artificial Intelligence in Forestry and Grassland] (promulgated by the National Forestry and Grassland Administration, Nov. 8, 2019), CLI.4.337391 (Lawinfochina); Fangzhen Jianzai Lingyu Rengongzhineng Fazhan Yanjiu Zhuanxiang Guihua (2023–2035 Nian) (防震减灾领域人工智能发展研究专项规划(2023—2035 年)) [Special Plan for the Development and Research of Artificial Intelligence in the Field of Earthquake Prevention and Disaster Reduction (2023–2035)] (promulgated by the China Earthquake Administration, Oct. 3, 2023), CLI.4.5183701 (Lawinfochina).

[38] State Council, *supra* note 27.

Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence[39] has outlined general promotion for using AI in areas like criminal justice, government services, and public healthcare service. However, it primarily emphasizes security, equity, and reliability, prioritizing ethical principles over specific actions, without stipulating specific implementation measures.[40] The U.S. emphasizes the potential applications of AI primarily within government services, focusing more on mitigating the risks that may arise from its use. Unlike China, the U.S. does not prioritize envisioning the broad, transformative potential of AI across industries, nor does it provide detailed, step-by-step measures to promote such widespread applications.

3. Talent Introduction and Cultivation Policies

As an important support for industry development, both countries have introduced various policies to promote talent introduction and cultivation, to attract more high-end human resources to enter the AI R&D field.

The U.S. government expedites the recruitment of AI talent and establishes expert working groups to address talent shortages.[41] The General Services Administration is mandated to collaborate with federal agencies and leverage the Presidential Innovation Fellows Program to attract AI experts.[42] Meanwhile, the federal fellowships and the promotion of AI education have been correspondingly prioritized.[43] These executive orders aim to strengthen AI development by improving talent acquisition, attracting external expertise, and advancing educational initiatives.

Recognizing the importance of human resources, Chinese facilitative law also emphasizes the cultivation and attraction of AI talents. The 2016 *Internet Plus* strongly encourages colleges and universities to provide training session of AI applications.[44] The 2017 *NGAIDP* further emphasizes on talent reserves, development, as well as intensifying efforts in workforce training.[45] The subsequent documents outline detailed measures for talent development, placing significant emphasis on attracting and recruiting top-tier global experts.[46] Furthermore, these documents establish specific goals to be achieved every five years from 2020 to 2030, with the aim that by 2030,

[39] Exec. Order No. 14,110, 88 Fed. Reg. 75191 (Nov. 1, 2023) **[hereinafter *EO 14110*]**.

[40] Although *Executive Order 14110* was formally revoked by President Trump on January 20, 2025, it remains a crucial point of reference in the analysis of U.S. AI regulation. The order played a foundational role in shaping AI policy initiatives, particularly in addressing national security risks and establishing frameworks for monitoring AI-related technologies. As such, it continues to be relevant in understanding the lasting impact on the regulatory framework and is therefore included in this analysis.

[41] *Id.*

[42] Exec. Order No. 13,960, 85 Fed. Reg. 78939 (Dec. 8, 2020) **[hereinafter *EO 13960*]**.

[43] *EO 13859, supra* note 30.

[44] NDRC et al., *supra* note 32.

[45] State Council, *supra* note 27.

[46] **Gaodeng Xuexiao Rengongzhineng Chuangxin Xingdong Jihua** (高等学校人工智能创新行动计划) [AI Innovation Action Plan for Institutions of Higher Education] (promulgated by the Ministry of Education (Jiaoyu Bu), Apr. 2, 2018), CLI.4.312949 (Lawinfochina); Guanyu “Shuang Yiliu” Jianshe Gaoxiao Cujin Xueke Ronghe Jiakuai Rengongzhineng Lingyu Yanjiusheng Peiyang de Ruogan Yijian (关于“双一流”建设高校促进学科融合加快人工智能领域研究生培养的若干意见) [Several Opinions on Promoting Interdisciplinary Integration and Accelerating Graduate Education in the Field of Artificial Intelligence at Universities Constructing “Double First-Class”] (promulgated by the Ministry of Education et al., Jan. 21, 2020), CLI.4.339960 (Lawinfochina).

Chinese higher education institutions will become a driving force behind the world's leading AI innovations, fueling the advancement of next-generation AI.[47]

4. Computing Power, Data and Other Essential Resources

Both countries acknowledge computing power and data as essential resources for AI development, with a shared emphasis on securing computing power, a crucial element of AI facilitative laws.

The U.S. attaches great importance to the sharing of data and computing resources to advance AI research while ensuring security, privacy, and confidentiality. Strategy 5 of *the AI Strategic Plan 2023* explicitly stipulates increasing investment in public resources for AI training and testing, granting researchers access to high-quality datasets.[48] Echoing this, All government agencies are required to review the usability of their Federal data and models and offer more opportunities for the non-Federal AI research community to access relevant data.[49]

China also recognizes that computing power and data are critical resources in AI technological development. Regarding data resources, innovation in data-driven AI technologies,[50] along with proactive planning and open access to data sets, are consistently suggested.[51] In terms of computing power, policies primarily focus on supporting centralized data processing, open access to computing platforms,[52] and the

[47] *Id.*

[48] *AI Strategic Plan 2023*, *supra* note 31.

[49] In 2019, the White House issued the *EO 13859*, which laid the foundation for the AI development strategy. It directed federal agencies to prioritize investments in AI R&D to ensure the United States maintains its global technological leadership. The order emphasized the importance of promoting international collaboration to ensure that global standards align with U.S. national interests.

[50] Cujin Dashuju Fazhan Xingdong Gangyao (促进大数据发展行动纲要) [The Action Outline for Promoting the Development of Big Data] (promulgated by the State Council, Aug. 31, 2015), CLI.2.256434 (Lawinfochina); Guanyu Jiakua Gongjian Quanguo Yitihua Dashuju Zhongxin Xietong Chuangxin Tixi de Zhidao Yijian (关于加快构建全国一体化大数据中心协同创新体系的指导意见) [Guiding Opinions on Accelerating the Construction of a Coordinated Innovation System for the National Integrated Big Data Center] (promulgated by the NDRC et al., Dec. 23, 2020), CLI.4.349469 (Lawinfochina); "Shi Si Wu" Dashuju Chanye Fazhan Guihua ("十四五"大数据产业发展规划) [14th Five-Year Plan for the Development of the Big Data Industry] (promulgated by the Ministry of Industry and Information Technology, Nov. 15, 2021), CLI.4.5111956 (Lawinfochina).

[51] NDRC et al., *supra* note 29.

[52] NDRC et al., *supra* note 29.

development of computing hubs in Western China.[53] According to the latest statistics, over 250 smart computing centers were either under construction or completed nationwide in the first half of 2024, with 791 bids for such centers, marking a 407% increase from the previous year. Over 20 cities have already established smart computing centers dedicated to AI model training.[54]

5. Market Competition and Private Sector R&D

In the U.S., market-driven forces play a central role in driving AI innovation, a stance deeply embedded in U.S. policy, reflecting the broader belief that the private sector is decisive in developing cutting-edge technologies. The *2020 Guidance for Regulation of Artificial Intelligence Applications* underscores the need to align AI regulatory frameworks with this market-driven approach, ensuring that U.S. companies maintain their competitive edge in the global market.[55] The executive order also highlights the importance of driving AI leadership through competitive forces, while advocating for the reduction of unnecessary regulations.[56]

While the Chinese government acknowledges that the market determines resource allocation, it emphasizes that the government should play a guiding role, particularly through policy support and market regulation.[57] In other words, it does not fully place its trust in the market's competitive mechanism, instead recognizing the need for government management to correct emerging issues in new fields. The *NGAIDP* and the following documents propose the establishment of platforms[58] that facilitate collaboration between industry, academia, and research institutions,

[53] Quanguo Yitihua Dashuju Zhongxin Xietong Chuangxin Tixi Suanli Shuniu Shishi Fang'an (全国一体化大数据中心协同创新体系算力枢纽实施方案) [Implementation Plan for the Computing Power Hub of the National Integrated Big Data Center Coordinated Innovation System] (promulgated by the NDRC et al., May 24, 2021), CLI.4.5013234 (Lawinfochina); Guanyu Jiakua Changjing Chuangxin Yi Rengongzhineng Gaoshui Ping Yingyong Cujin Jingji Gao Zhi Liang Fazhan de Zhidao Yijian (关于加快场景创新以人工智能高水平应用促进经济高质量发展的指导意见) [Guiding Opinions on Accelerating Scenario Innovation and Promoting High-quality Economic Development with High-level Application of Artificial Intelligence] (promulgated by the Ministry of Science and Technology et al., Jul. 29, 2022), CLI.4.5132750 (Lawinfochina); Suanli Jichu Sheshi Gao Zhi Liang Fazhan Xingdong Jihua (算力基础设施高质量发展行动计划) [Action Plan for the High-Quality Development of Computing Power Infrastructure] (promulgated by the Ministry of Industry and Information Technology et al., Oct. 8, 2023), CLI.4.5178324 (Lawinfochina); Guanyu Shenru Shishi "Dong Shu Xi Suan" Gongcheng Jiakuai Goucheng Quanguo Yitihua Suanli Wang de Shishi Yijian (关于深入实施"东数西算"工程加快构建全国一体化算力网的实施意见) [Implementation Opinions on Deepening the 'Eastern Data, Western Computing' Project and Accelerating the Construction of a National Integrated Computing Power Network] (promulgated by the NDRC et al., Dec. 25, 2023), CLI.4.5185823 (Lawinfochina).

[54] Woguo Jia Kuai Tuijin Suanli Jishu Biaozhunhua Jianshe (我国加快推进算力技术标准化建设) [*China Accelerates the Advancement of Computing Power Technology Standardization*], Xinhua News (Nov. 28, 2024), https://www.news.cn/tech/20241128/1454640d1e424c72a84f23292dac6315/c.html.

[55] Office of Mgmt. & Budget, Exec. Office of the President, Memorandum M-21-06, Guidance for Regulation of Artificial Intelligence Applications (Nov. 17, 2020) [hereinafter *AI Applications Guidance*]

[56] *AI Strategic Plan 2023, supra* note 31.

[57] State Council, *supra* note 27.

[58] State Council, *supra* note 27; Guojia Xin Yidai Rengongzhineng Kaifang Chuangxin Pingtai Jianshe Gongzuo Zhiyin (国家新一代人工智能开放创新平台建设工作指引) [Guidelines for Establishing National New-Generation AI Open Innovation Platforms] (promulgated by the Ministry of Science and Technology, Aug. 1, 2019), CLI.4.334682 (Lawinfochina).

emphasizing the importance of such partnerships.[59] Meanwhile, the pilot zones focus on exploring government policies and fostering interaction between AI and society, including social experiments and infrastructure development.[60] As of January 10, 2024,[61] 23 innovation platforms have been created, and by December 6, 2021, 17 pilot zones have been established.[62]

Such distinctions reflect the differing perspectives regarding the incentivizing function of AI law. China's issuance of numerous documents is premised on the belief that the government can guide industries towards more proactive development of AI applications in various scenarios, while the U.S.'s more hands-off approach seems to indicate a belief that the direction of market development should be left to market forces to determine.

6. Financial Support

Beyond their differing approaches to market competition, both countries recognize the need for substantial investment, but their allocation strategies reflect differing priorities.

In its report Driving U.S. Innovation in Artificial Intelligence, the National Security Commission on Artificial Intelligence emphasizes that to achieve its technology goals, the U.S. must invest at least $10 billion annually.[63] In line with this, the federal government is committed to consistently allocating the necessary funding for AI R&D.[64]

This substantial financial commitment is further supported by the various legislative measures, such as the *CHIPS and Science Act (2022)*,[65] which supports the semiconductor industry with tax credits for domestic manufacturing. The NSF also provides competitive awards to support AI research institutions and nonprofit

[59] Lucero, *supra* note 4 at 124.

[60] Guojia Xin Yidai Rengongzhineng Chuangxin Fazhan Shiyanqu Jianshe Gongzuo Zhiyin (Xiudingban) (国家新一代人工智能创新发展试验区建设工作指引（修订版）) [Guidelines for the Construction of National New-Generation AI Innovation and Development Pilot Zones (Revised Edition)] (promulgated by the Ministry of Science and Technology, August 29, 2019, effective August 29, 2019; rev'd by the Ministry of Science and Technology, September 29, 2020) Lawinfochina, CLI.4.351354.

[61] 25 Ge Chuangxin Pingtai, Guojiadui Shengdui Qi Baodao! Guangdong Zheyang Buju AI Xin Saidao (25 个创新平台，国家队省队齐报到！广东这样布局 AI 新赛道) [25 Innovation Platforms Join Forces! National and Provincial Teams Set the Stage for AI Development in Guangdong], the official website of Guangdong Provincial Department of Science and Technology (Jan. 10, 2024, 10:52 AM), https://gdstc.gd.gov.cn/kjzx_n/gdkj_n/content/post_4329590.html.

[62] *Guojia Xin Yidai Rengong Zhinen Chuangxin Fazhan Shiyanqu Yi Da 17 Ge* (国家新一代人工智能创新发展试验区已达 17 个) [*The National New Generation Artificial Intelligence Innovation Development Pilot Zones Have Reached 17*], The State Council of the People's Republic of China (Dec. 6, 2021), https://www.gov.cn/xinwen/2021-12/06/content_5657953.htm.

[63] Nat'l Sec. Comm'n on Artificial Intelligence, *Final Report* (2021), https://reports.nscai.gov/final-report/.

[64] Nat'l Inst. of Standards & Tech., A Plan for Federal Engagement in Developing Technical Standards and Related Tools (Aug. 9, 2019). https://www.nist.gov/system/files/documents/2019/08/10/ai_standards_fedengagement_plan_9aug2019.pdf.

[65] *CHIPS and Science Act*, Pub. L. No. 117-167, 136 Stat. 1366 (2022).

organizations.[66] According to the 2024 government budget, the federal government has allocated over $3 billion to support departments in developing AI technologies to achieve the multi-disciplinary R&D goals.[67] This amount is expected to grow further. In May 2024, bipartisan senators called for a significant increase in government funding for AI research, proposing no less than $32 billion annually for AI innovation in non-defense sectors.[68] While the U.S. government has made substantial strides in AI funding, China has similarly committed significant financial resources to foster innovation across both academic and industrial sectors.

China has similarly committed substantial financial resources to foster innovation across both academic and industrial sectors. Key initiatives, such as the *Internet Plus* and *NGAIDP*, provide significant support for AI advancement. Fiscal and tax policies, including tax incentives for high-tech companies and additional deductions for R&D expenses, further stimulate AI development.[69] Moreover, the National Natural Science Foundation of China established a separate discipline code (F06) for AI research projects in 2018. In 2020, the Ministry of Industry and Information Technology of China added three new AI-related secondary discipline codes under the "Artificial Intelligence" research.[70] Between 2018 and 2023, a total of 3,343 projects were funded, with a total funding of 1.862 billion RMB.[71]

7. Public-Private Collaboration

Public-private collaboration is a cornerstone of U.S. AI policy. The U.S. government's approach prioritizes fostering strong partnerships between academia, industry, and government, aiming to enhance synergies across various sectors. It also incentivizes private sector involvement through government contracts, thereby driving technological progress while ensuring the responsible commercialization of AI.

[66] AI research funding programs aim to tackle both technical and ethical challenges. The Cyber Physical Systems Program focuses on developing secure, trustworthy, and interpretable AI systems with an emphasis on safety and transparency. The Secure and Trustworthy Cyberspace Program supports cybersecurity and privacy research for automated systems. The Formal Methods in the Field Program prioritizes formal verification to ensure AI reliability. The Designing Accountable Software Systems Program funds research into methodologies for developing software that complies with legal and regulatory standards.

[67] The White House, *Fact Sheet: The President's Budget Advances President Biden's Unity Agenda* (Mar. 11, 2024), https://www.whitehouse.gov/briefing-room/statements-releases/2024/03/11/fact-sheet-the-presidents-budget-advances-president-bidens-unity-agenda/.

[68] David Shepardson, *U.S. Senators Unveil AI Policy Roadmap, Seek Government Funding Boost*, Reuters (May 15, 2024), https://www.reuters.com/world/us/us-senators-unveil-ai-policy-roadmap-seek-government-funding-boost-2024-05-15/.

[69] State Council, *supra* note 27; NDRC et al., *supra* note 32.

[70] Wu Guozheng(吴国政) et al., *Qianxi Rengong Zhinen Xueke Jijin Xiangmu Shenqing Zizhu Qingkuang Ji Zhanwang* (浅析人工智能学科基金项目申请资助情况及展望) *[A Brief Analysis and Prospect of Artificial Intelligence Discipline Fund Project Applications and Funding Situations]*, 46 Zidonghua Xuebao 自动化学报 *[Acta Automatica Sinica]* No. 12 2711, 2712.

[71] F06. Rengong Zhinen, Zidonghuasuo Diyi, Jiexialai Jingzheng Jilie, Shui Shi Yajun! Guojia Ziran Kexue Jijin Erji Xueke Remen Yituo Danwei TOP20 (F06.人工智能，自动化所第一，接下来竞争激烈，谁是亚军！国家自然科学基金二级学科热门依托单位 TOP20) [*F06. Artificial Intelligence: Automation Institute Ranked First, Fierce Competition Ahead—Who Will Be the Runner-Up? Top 20 Popular Supporting Institutions for Secondary Disciplines of the National Natural Science Foundation*], Inquire Research (Dec. 18, 2023), https://mp.weixin.qq.com/s/JZOH8lKujCDQtiATQvloHw.

The U.S. government emphasizes that AI innovation and commercialization, as well as risk mitigation, should be driven by robust partnerships between the public and private sectors.[72] This approach is highlighted across U.S. AI strategy documents, which stress the importance of collective efforts in developing AI standards,[73] expanding access to resources, and fostering real-world applications of AI technology.[74] For instance, the National Science Foundation's *2024 National AI Research Resource Pilot* offers AI researchers and educators access to vital computational resources, data, software, and models, supporting innovation in AI.[75] Meanwhile, the U.S. government has also stimulated private sector development through AI-related procurement contracts,[76] with contract amounts increasing annually.[77] These contracts, which have seen a marked rise in funding—from $261 million to $675 million in just one year—help promote AI industry growth by establishing clear responsibilities for private businesses while aligning with national interests.

Moreover, the U.S. government has further cemented its supportive role in AI industry growth through the establishment of the AI Center of Excellence, designed to enhance AI's effectiveness in federal operations.[78] As outlined in the final report of *National Security Commission on Artificial Intelligence* report, these public-private partnerships are key to transforming scientific innovations into economic value,[79] with close collaboration among market participants, venture capital, and key stakeholders. This approach ensures that market participants, particularly AI companies, play a central role in both the formation and implementation of policy. This collaborative model not only accelerates technological advancements but also ensures that the resulting policies are informed by the needs and insights of the industry.

As for China, consistent with the section on "Fostering Market Competition," the Chinese government plays a crucial role in shaping policies, ensuring safety, and regulating the industry while fostering a business-friendly environment. China

[72] Hereinafter *EO 14110, supra* note 39.

[73] National Institute of Standards and Technology, *U.S. LEADERSHIP IN AI: A Plan for Federal Engagement in Developing Technical Standards and Related Tools* (Aug. 9, 2019) https://www.nist.gov/system/files/documents/2019/08/10/ai_standards_fedengagement_plan_9aug2019.pdf.

[74] *AI Strategic Plan 2023, supra* note 31.

[75] U.S. Nat'l Sci. Found., National Artificial Intelligence Research Resource (NAIRR) Pilot, https://nairrpilot.org (last visited Jan. 22, 2025).

[76] U.S. Gov't Accountability Off., *A Snapshot of Government-Wide Contracting for FY 2023* (2024), https://www.gao.gov(For technology companies, although AI contract suppliers remain relatively dispersed, the number of high-value government contracts has increased significantly. In 2023, among newly signed government contracts, 205 suppliers secured deals worth over $10 million each, with six exceeding $50 million per contract.).

[77] Mark Muro & Jacob Whiton, *The Evolution of Artificial Intelligence Spending by the U.S. Government*, BROOKINGS INST. (Sept. 12, 2024), https://www.brookings.edu/articles/the-evolution-of-artificial-intelligence-ai-spending-by-the-u-s-government/ (Between August 2022 and August 2023, U.S. federal government funding for AI hardware and software contracts rose from $261 million to $675 million, while the potential award value increased nearly 1,200%, from $355 million to $4.561 billion. In the national defense sector, AI-related spending grew from $269 million (76% of all federal funding) to $4.323 billion in 2023 (95% of all funding). Meanwhile, AI contract spending in the aviation and healthcare sectors increased by 25% to 30%, respectively.).

[78] **H.R. 2575, 116th Cong. (2019).**

[79] *Final Report: National Security Commission on Artificial Intelligence*, National Security Commission on Artificial Intelligence, https://reports.nscai.gov/final-report/(last visited Feb 28, 2025).

emphasizes balancing government and market forces,[80] with the government playing a central role in policy support, regulation, and fostering a favorable business environment.

Legal frameworks set forth detailed guidelines for key industry tasks, including the development of AI products, critical components like sensor chips, and essential infrastructure.[81] These tasks are often addressed through competitive mechanisms designed to select the most promising candidates,[82] whether research institutes, private companies, or other organizations. Once selected,[83] these initiatives receive substantial funding and policy support, facilitating their growth. This can be seen as a "wish list" for AI progress,[84] with clear directives guiding industry implementations guided by a series of supporting documents.[85] Through these efforts, China aims to position itself as a leader in AI innovation while ensuring that the necessary infrastructure and talent development are in place to sustain its growth.

B. Comparative Analysis of the U.S.-China AI Facilitation Law

The key difference between China and the U.S. in the realm of AI facilitative law lies in their governance structure and role of government. The U.S. adopts a more decentralized, market-driven model, where the government's role is largely to facilitate innovation by setting ethical guidelines, ensuring public trust, and providing access to resources without direct control over industry development. In contrast, China's top-

[80] State Council, *supra* note 27.

[81] Cujin Xin Yidai Rengongzhineng Chanye Fazhan Sannian Xingdong Jihua (2018-2020 Nian) (促进新一代人工智能产业发展三年行动计划（2018-2020 年）) [Three-Year Action Plan for Promoting the Development of a New Generation of Artificial Intelligence Industry (2018-2020)] (promulgated by the Ministry of Industry and Information Technology, Dec. 13, 2017), CLI.4.306740 (Lawinfochina).

[82] Xinyi Dai Rengongzhineng Chanye Chuangxin Zhongdian Renwu Jiebang Gongzuo Fang'an (新一代人工智能产业创新重点任务揭榜工作方案) [Work Plan for the Key Tasks of "Ranking and Tackling" in the New-Generation Artificial Intelligence Industry Innovation] (promulgated by the General Office of the Ministry of Industry and Information Technology, Nov. 8, 2018), CLI.4.326219 (Lawinfochina).

[83] E.g., Gongxin Bu Gongbu Shoupi 48 Ge AI Chanye Chuangxin Jiebang You Sheng Chengguo (工信部公布首批 48 个 AI 产业创新揭榜优胜成果) *[The Ministry of Industry and Information Technology Announces the First Batch of 48 Winning Results in AI Industry Innovation "Ranking and Tackling"]*, China News Service (May 21, 2021), https://www.chinanews.com/it/2021/05-21/9482886.shtml.

[84] Matt Sheehan, *How China's Massive AI Plan Actually Works*, MACROPOLO (Feb.12, 2018), https://macropolo.org/analysis/how-chinas-massive-ai-plan-actually-works/.

[85] Renxing Jiqiren Chuangxin Fazhan Zhidao Yijian (人形机器人创新发展指导意见) [Humanoid Robot Innovation Development Guidance] (promulgated by the Ministry of Industry and Information Technology, Oct. 20, 2023), CLI.4.5181600 (Lawinfochina); Guanyu Tuidong Weilai Chanye Chuangxin Fazhan de Shishi Yijian (关于推动未来产业创新发展的实施意见) [Implementation Opinions on Promoting the Innovation and Development of Future Industries] (promulgated by the Ministry of Industry and Information Technology, Ministry of Education, Ministry of Science and Technology, Ministry of Transport, Ministry of Culture and Tourism, State-owned Assets Supervision and Administration Commission of the State Council, and Chinese Academy of Sciences, Jan. 8, 2024), CLI.4.5187548 (Lawinfochina); Guanyu Jiakuai Chuantong Zhizao Ye Zhuanxing Shengji de Zhidao Yijian (关于加快传统制造业转型升级的指导意见) [Guiding Opinions on Accelerating the Transformation and Upgrading of Traditional Manufacturing Industry] (promulgated by the Ministry of Industry and Information Technology, National Development and Reform Commission, Ministry of Education, Ministry of Finance, People's Bank of China, State Taxation Administration, National Financial Regulatory Administration, and China Securities Regulatory Commission, Dec. 28, 2023), CLI.4.5185686 (Lawinfochina).

down approach involves strong state involvement in setting goals, promoting applications, and securing resources, with the government playing a central role in guiding AI growth in various areas, from talent development to infrastructure.

Despite these differences, both countries recognize the critical importance of talent, computing power, data, and public-private collaboration, and both have established robust funding mechanisms to support AI innovation. Building on the aforementioned overview, the following section will specifically analyze the differences in the strategies of both countries in facilitative law for AI.

1. The U.S.: Market-Led Facilitation

Upon reviewing the facilitative laws, it is clear that the U.S. takes a different approach than China. Unlike China's policy documents, which set clear, quantifiable development metrics for AI, U.S. policies focus more on promoting AI applications without setting specific development goals or providing detailed provisions for extensive application scenarios. This difference reflects a broader perspective in the U.S. that prioritizes market forces over government intervention in driving AI technology development. By doing so, the U.S. model allows AI companies greater autonomy in determining the direction and pace of innovation based on market demand, offering more room for flexible, bottom-up development.

The *National AI Strategy* further reinforces this by emphasizing the importance of robust market competition and the creation of a fair, open market. This underscores the U.S.'s commitment to private-sector-driven progress, where market forces lead the way in determining the pace and direction of AI development.

Consistent with the principle of free competition, the U.S. government provides resources in a supportive manner, focusing primarily on creating an environment conducive to innovation with essential resources—such as talent, data, and computing power—thereby establishing a robust ecosystem where industry leaders and stakeholders can collaborate freely. The U.S. government also actively encourages collaboration with private companies as a core element of its AI governance strategy. One example of this is the establishment of the *National Artificial Intelligence Initiative Office*, which facilitates communication with stakeholders, including private enterprises, ensuring that market participants play a crucial role in shaping policy recommendations.

Beyond the collaborations mentioned, the U.S. government also supports AI technology development through targeted investments that strike a balance—providing support without attempting to dominate the industry. Guided by the Executive Orders, the U.S. government established partnerships with market players by entering into procurement contracts with AI tech companies.[86] With total AI contract spending projected to reach $32 billion by 2026, this approach not only fosters collaboration but is also reinforced by a sharp increase in government funding for the AI industry.[87] From 2022 to 2023, federal funding for AI-related contracts increased by over 150%, with

[86] *AI Strategic Plan 2023*, *supra* note 31; *EO 13960*, *supra* note 42.

[87] Indermit Gill, *Whoever Leads in Artificial Intelligence in 2030 Will Rule the World until 2100*, BROOKINGS (Jan. 17, 2020), https://www.brookings.edu/articles/whoever-leads-in-artificial-intelligence-in-2030-will-rule-the-world-until-2100/.

the value of potential contracts soaring by nearly 1,200%. Notably, 95% of AI contract spending comes from the Department of Defense, highlighting the significant military and defense sector focus. It also provides direct funding support and tax incentives for the semiconductor industry, which clearly demonstrates the government's efforts to accelerate AI development and maintain its technological edge.[88]

2. China: Government-Led Facilitation

The Chinese government has set forth detailed legal instruments on AI development, including master planning, direction of technological development, data and computing power, and promotion of application scenarios, among others. To support these efforts, China is establishing professional institutions, such as an *AI planning and promotion office* under the Ministry of Science and Technology, which is dedicated to advancing the implementation of AI strategies. In addition, an AI strategic advisory committee is created to provide guidance and evaluate key AI decisions.[89] Furthermore, the *NGAIDP* underscores that the overall AI development is led, planned, and coordinated by the National Leading Group for Science and Technology System Reform and Innovation System Development.[90]

At the heart of China's AI development strategy is a government-led model, where the government plays a central role not only as an investor but also as a strategic guide for resource allocation. By setting priorities, establishing clear objectives, and directing the course of AI advancement, the government plays a key driver for industry growth and innovation through hands-on involvement. This government's deep engagement reflects China's recognition of AI's transformative potential to optimize manufacturing processes, automate tasks, and enhance business operations, particularly in sectors such as manufacturing, electronics, and technology.[91]

China's facilitative approach is focused on pooling existing resources to achieve key breakthroughs in designated areas, ensuring that research outcomes are effectively translated into industrial and economic benefits. This government-led model is consistent with China's broader economic development system. In November 2013, the Third Plenary Session of the 18th Central Committee of the Communist Party of China proposed a dual approach—"enabling the market to play a decisive role in resource allocation, while ensuring the government effectively guides industry development priorities". The government transitions from directly allocating market resources to "guiding the priorities of industry development" and "setting the goals for industry development." Government departments at all levels are tasked with their own development targets and policy objectives for the sectors under their management..[92]

The Chinese government's deep participation in the AI industry could find the

[88] *CHIPS and Science Act, supra* note 65.

[89] State Council, *supra* note 27.

[90] State Council, *supra* note 27.

[91] Gu Feng, *Corpus-based Critical Discourse Analysis on AI Policy: A Comparison Between North America and Developing Countries in East Asia*, 8 ASIAN JOURNAL OF SOCIAL SCIENCE STUDIES 14 (2023).

[92] In terms of phase-based goals, the Chinese government has issued a series of documents, including the *Next Generation Artificial Intelligence Development Plan*, the *13th Five-Year National Science and Technology Innovation Plan*, and the *14th Five-Year Plan for National Economic and Social Development and the Vision for 2035*, to clearly define the development objectives for the AI industry every three to five years.

answer in its critical scientific research work system proposed by China in recent years, namely "new system for mobilizing the resources nationwide". This concept was first outlined in the Fifth Plenary Session of the 18th Central Committee of the CPC in 2015. It was further comprehensively elaborated in September 2022, that is, "We need to effectively bring the government, market, and society together ... target a number of critical areas.... We need to reinforce the centralized and unified leadership of the Party Central Committee, and establish a firm but fair decision-making and commanding system".[93] On June 24, 2024, President Xi Jinping re-emphasized that the Party Central Committee shall exercise the centralized and unified leadership over science and technology work.[94] In a nutshell, it is an operational mechanism where to achieve technological development, the government makes plans for, mobilizes and allocates national resources from all sectors to accomplish major tasks in technological development.

Recognizing the immense potential of AI to fuel economic growth, the Chinese government acknowledges that AI development requires substantial financial investment and resources. While China still faces numerous challenges and deficiencies—particularly in fundamental theories, technologies, human resources, and industry infrastructure—the government continues to leverage its centralized system to focus resources and pursue systematic planning. These align with the objectives of "the new system for mobilizing nationwide resources," which emphasizes strategic coordination and resource allocation to address existing gaps and accelerate progress in the "key technology with first-mover advantages and foundational frontier technology that leads future development".[95] In short, China's AI development is characterized by government-led planning, with the government playing an active role in guiding and facilitating the sector's growth through strategic intervention.

In this context, local governments also play a critical role in this government-led model driving AI innovation. Local governments play an active role in implementing regulatory provisions, and for another, they take active steps in policy implementation per the goals of industry development and offer funding and policy incentives for AI technological innovation in their respective region.[96] Following the

[93] *Xi Jinping Zhuchi Zhaokai Zhongyang Quanmian Shenhua Gaige Weiyuanhui Di Ershiqi Ci Huiyi Qiangdiao Jianquan Guanjian Hexin Jishu Gongguan Xinxing Juguo Tizhi Quanmian Jiaqiang Ziyuan Jieyue Gongzuo* (习近平主持召开中央全面深化改革委员会第二十七次会议强调健全关键核心技术攻关新型举国体制 全面加强资源节约工作) [*Xi Jinping Chairs the 27th Meeting of the Central Commission for Comprehensively Deepening Reform, Emphasizing the Improvement of a New National System for Tackling Core Technologies and Strengthening Resource Conservation*], Xinhua News (Sep. 6, 2022), https://www.news.cn/politics/leaders/2022-09/06/c_1128981539.htm.

[94] Xi Jinping (习近平), Zai Quanguo Keji Dahui, Guojia Kexue Jishu Jiangli Dahui, Liangyuan Yuanshi Dahui Shang de Jianghua (在全国科技大会、国家科学技术奖励大会、两院院士大会上的讲话) [Speech at the National Science and Technology Conference, National Science and Technology Awards Conference, and the Academician Conference of the Two Academies], Xinhua News (Jun. 24, 2024), https://www.news.cn/politics/leaders/20240624/16741a201e564d8d8775ffb1450ecf29/c.html.

[95] Lu Feng (路风) & He Pengyu (何鹏宇), Xinxing Juguo Tizhi: Zhongguo Zhengzhi Lingdaoceng Litu Wancheng Zhongda Biange de Renwu Tizhi (新型举国体制：中国政治领导层力图完成重大变革的任务体制) [The New National System: A Task System for China's Political Leadership to Achieve Major Reforms], Zhili Yanjiu (治理研究) [Governance Studies], no. 4. 2024, at 7-8.

[96] Huw Roberts et al., *The Chinese Approach to Artificial Intelligence: An Analysis of Policy, Ethics, and Regulation*, 36 AI & Soc 59, 61–62 (2021).

performance assessment reforms introduced in October 2020,[97] local governments are incentivized to prioritize AI development, as the achievements in technological innovation now directly influence the career prospects of government officials.[98] As local governments compete to attract tech investments, they offer local funding and policy incentives to AI companies. The resulting cluster effect, where technological R&D achievements in one region attract further investments, accelerates the growth of AI-related industries in those areas. In this highly competitive environment, local governments seek to align with the central government's policy initiatives,[99] making AI development a high priority to enhance their growth prospects and secure resources for technological innovation.[100] This eventually contributes to the creation of a highly systematic system for investment attraction that encompasses land development,[101] industrial planning, and project running.[102]

II. THE U.S. AND CHINA'S APPROACH TO AI REGULATORY

A. Analysis of Regulatory Law in the U.S. and China

As AI technologies rapidly evolve, both countries are working to establish frameworks to manage Al risks. Al regulatory law encompasses rules, policies, and regulations aimed at governing AI development and addressing risks like ethical concerns, safety, data privacy, and algorithmic bias. The first sub-section compares the regulatory laws of the U.S. and China, followed by a deeper analysis of how each country tackles these challenges within their Al governance frameworks.

1. Fundamental Principles and Policy Objectives

Regarding policy objectives, the U.S. government underscores that the successful application of AI depends on public trust and recognition, highlighting the

[97] Guanyu Gaijin Tuidong Gao Zhiliang Fazhan de Zhengji Kaohe de Tongzhi (关于改进推动高质量发展的政绩考核的通知) [Notice on Improving Performance Evaluation to Promote High-Quality Development] (promulgated by the Organization Department of the CPC Central Committee, Oct. 24, 2020), CLI.16.347642 (Lawinfochina).

[98] Song Di (宋笛), Difang Zhengfu De "Keji Zhaoshang" Zhan (地方政府的"科技招商"战) [The "Tech Investment Promotion" Battle of Local Governments], The Economic Observer (Jun. 16, 2018, 9:29 AM), https://m.eeo.com.cn/2018/0616/330459.shtml.

[99] Hongbin Li, *Political turnover and economic performance: the incentive role of personnel control in China*, 89 JOURNAL OF PUBLIC ECONOMICS 1743 (2005).

[100] Matt Sheehan, *How China's Massive AI Plan Actually Works*, MACROPOLO (Feb. 12, 2018), https://macropolo.org/analysis/how-chinas-massive-ai-plan-actually-works/ (last visited Oct 26, 2024).

[101] Chen Shuyun (陈淑云) & Zeng Long (曾龙), Difang Zhengfu Tudi Churang Xingwei Dui Chanye Jiegou Shengji Yingxiang Fenxi—Jiyu Zhongguo 281 Ge Diji Ji Yishang Chengshi de Kongjian Jiliang Fenxi (地方政府土地出让行为对产业结构升级影响分析——基于中国 281 个地级及以上城市的空间计量分析) [An Analysis of the Impact of Local Government Land Transfer Behavior on Industrial Structure Upgrading—A Spatial Econometric Analysis Based on 281 Prefecture-Level and Above Cities in China], Chanye Jingji Yanjiu (产业经济研究) [Industrial Economics Research], no. 6, 2017, at 89, 100.

[102] Lv Yuxia (吕玉霞), Hou Linke (侯麟科) & Wan Xueying (万学焴), Jingji Kaifaqu Zhaoshang Yinzi de Zuzhi Dongyuan he Chanye Jiju Celue—Jiyu Weiguan Qiye Shuju de Fenxi (经济开发区招商引资的组织动员和产业集聚策略——基于微观企业数据的分析) [Organizational Mobilization and Industrial Agglomeration Strategies for Investment Promotion in Economic Development Zones—An Analysis Based on Micro-Level Firm Data], Chanye Jingji Pinglun (产业经济评论) [Industrial Economics Review], no. 4, 2017, at 5, 8-9.

need for reliable, robust, and trustworthy AI technologies to boost public confidence.[103] In its strategic focus on national security, the government prioritizes AI as a critical component, the "covered national security technologies and products",[104] emphasizing its importance alongside other emerging technologies. Key policies stress addressing security risks to AI systems, fostering responsible innovation, and encouraging international collaboration to maintain global leadership in AI.[105] Furthermore, the government mandates that AI applications used within federal agencies be lawful,[106] transparent, accountable, and aligned with national values, with ongoing monitoring and safeguards to ensure these standards are met.[107] At the state level, the principles are swiftly responded. For example, California committed to examining and incorporating these principles into its legislation regulating the use and deployment of automated systems.[108]

China's approach to AI regulation is underpinned by a set of fundamental principles and policy objectives aimed at ensuring the safe and ethical development of AI technologies.

The *NGAIDP* makes generalized provisions of establishing an ethical and moral framework to ensure healthy AI development.[109] To this end, *Governance Principles for a New Generation of Artificial Intelligence — Developing Responsible AI*, alongside the *Ethical Norms for a New Generation of Artificial Intelligence,* lay down eight principles: harmony and friendliness, equity and justice, inclusiveness and sharing, safety and reliability, shared responsibility, openness and collaboration, and agile governance.[110] These principles aim to "enhance human well-being, promote equity and justice, protect privacy and safety, ensure reliability and trustworthiness, strengthen accountability and responsibility, and improve ethical literacy".[111]

Ethical and moral initiatives serve as a crucial component of China's regulatory intentions, signaling the government's emphasis on responsible innovation. However, it is important to note that these ethical initiatives are not legally binding. Their implementation largely depends on the voluntary compliance of research institutions,

[103] *CHIPS and Science Act, supra* note 65.

[104] Exec. Order No. 14,105, 3 C.F.R. 54867 (2023). [hereinafter *EO 14105*]

[105] *EO 14110, supra* note 39.

[106] The White House issued Executive Order 13960, *Promoting the Use of Trustworthy AI in the Federal Government*, underscoring the responsible use of AI within federal agencies. The order outlined principles such as privacy, civil liberties, accountability, and transparency, aiming to foster public trust in government AI systems and ensure their alignment with national values.

[107] *EO 13960*, *supra* note 42.

[108] S. Con. Res. 17, 2023–2024 Leg., Reg. Sess. (Cal. 2023) (enacted).

[109] State Council, *supra* note 27.

[110] *Xin Yidai Rengong Zhinen Zhili Yuanze—Fazhan Fu Zeren de Rengong Zhinen* (新一代人工智能治理原则——发展负责任的人工智能) *[Governance Principles for a New Generation of Artificial Intelligence—Developing Responsible AI]*, released by the National New Generation AI Governance Professional Committee, Ministry of Science and Technology Official website (June 17, 2019), https://www.most.gov.cn/kjbgz/201906/t20190617_147107.html.

[111] *Xin Yidai Rengong Zhinen Lunli Guifan* (新一代人工智能伦理规范) *[Ethical Norms for a New Generation of Artificial Intelligence]*, released by the National New Generation AI Governance Professional Committee (Sept. 25, 2021), Ministry of Science and Technology Official Website, https://www.most.gov.cn/kjbgz/202109/t20210926_177063.html.

businesses, and users, which raises questions about the effectiveness of these measures in driving long-term, systemic change.

2. Algorithmic Safety and Risk Mitigation

a. The U.S. Approach to Algorithmic Safety Regulation

The U.S. approach to AI regulation is characterized by a strong focus on safety oversight, privacy protection, algorithmic transparency, and anti-discrimination efforts.

A key characteristic of U.S. policy lies in the integration of systemic safety oversight, with safety measures such as the implementation of red-teaming procedures and mandatory reporting on foreign transactions involving AI models.[112] Under such framework, the supplementary documents, for example, the *Blueprint for an AI Bill of Rights*, further emphasize that automated systems should be developed in consultation with different stakeholders and undergo ongoing monitoring.[113] Operationally, the *AI Risk Management Framework* offers a reference framework for AI regulation that includes four core functions[114]: GOVERN[115], MAP[116], MEASURE[117], and MANAGE[118], to ensure the reliability and safety of AI technologies across different domains of application. This comprehensive approach also calls for AI systems to undergo regular risk assessments to ensure they remain effective and secure.[119]

Another important aspect is privacy protection. The *Blueprint for an AI Bill of Rights*[120] emphasizes data security, particularly regarding personally identifiable information. The use of AI in sensitive fields like healthcare is tightly regulated,

[112] *EO 13960*, *supra* note 42.

[113] Office of Sci. & Tech. Policy, The White House, Blueprint for an AI Bill of Rights: Making Automated Systems Work for the American People (Oct. 2022), https://www.whitehouse.gov/wp-content/uploads/2022/10/Blueprint-for-an-AI-Bill-of-Rights.pdf. [hereinafter *AI Blueprint*]

[114] National Institute of Standards and Technology, *Artificial Intelligence Risk Management Framework* (Jan. 26, 2023), https://doi.org/10.6028/NIST.AI.100-1. [hereinafter *RMF*]

[115] GOVERN：A cross-cutting function that is infused throughout AI risk management and enables the other functions of the process. Aspects of GOVERN, especially those related to compliance or evaluation, should be integrated into each of the other functions. Attention to governance is a continual and intrinsic requirement for effective AI risk management over an AI system's lifespan and the organization's hierarchy.

[116] MAP：The MAP function establishes the context to frame risks related to an AI system. The information gathered while carrying out the MAP function enables negative risk prevention and informs decisions for processes such as model management, as well as an initial decision about appropriateness or the need for an AI solution.

[117] MEASURE：The MEASURE function employs quantitative, qualitative, or mixed-method tools, techniques, and methodologies to analyze, assess, benchmark, and monitor AI risk and related impacts.

[118] MANAGE：The MANAGE function entails allocating risk resources to mapped and measured risks on a regular basis

[119] In October 2022, the U.S. federal government released the *Blueprint for an AI Bill of Rights*, a non-binding guidance document intended to supplement the *Executive Order on the Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence*. The blueprint articulated five key principles—safety and effectiveness, non-discrimination, data privacy, transparency, and human alternatives and control—highlighting the primary concerns of AI regulation in the United States, including directives for federal agencies: the Federal Trade Commission (FTC) to promote transparency and fairness in algorithmic systems; the Equal Employment Opportunity Commission (EEOC) to enforce anti-discrimination principles in AI-driven hiring processes; and the Consumer Financial Protection Bureau (CFPB) to supervise AI-assisted credit applications.

[120] *AI Blueprint, supra* note 113.

ensuring compliance with privacy laws.[121] As for State-level regulations, they focus on protecting consumers from abuse by algorithmic platforms, limiting the platforms' use of and access to data, and upholding consumer opt-out rights.[122]

Algorithmic transparency is another key focus of regulatory policies, with clear mandates for developers to provide easily understandable explanations of AI system functions.[123] This principle emphasizes the importance of transparency,[124] especially in high-stakes decision-making, as reflected in legislation requiring internet platforms using generative AI to disclose information about AI-generated content to users.[125] Moreover, federal and state regulations are gradually incorporating AI into existing frameworks,[126] ensuring that AI-generated content adheres to the same standards of fairness and accountability as other commercial practices.[127]

The U.S. also prioritizes the mitigation of algorithmic bias and discrimination. Policies aim to ensure AI systems foster equity,[128] with specific initiatives helping businesses avoid discrimination in AI-assisted hiring processes.[129] Deep synthesis technology also raises ethical concerns,[130] with various legal instruments stressing the urgency to address such risks.[131] U.S. policymakers have been actively addressing the risks through legislation, particularly in areas like deepfakes and AI-generated voices.[132] Meanwhile, the government departments are extending regulations to cover AI-generated voices to safeguard individuals' rights.[133] Legislative efforts also focus

[121] The *Georgia Control of Hazardous Conditions Act* requires that the use of AI in the course of medical care should conform to the *Georgia Telehealth Act*, and the *Health Insurance Portability and Accountability Act* and other privacy protection laws. Connecticut passed the *Artificial Intelligence Automated Decision* to regulate AI's use of private data and required the Office of Policy and Management to conduct a security check of all AI systems. *See*: Ga. Code Ann. § 31-12 (2022); S.B. 1103, 2023 Gen. Assemb., Reg. Sess. (Conn. 2023).

[122] *See* [Annex 2].

[123] *AI Applications Guidance, supra* note 55.

[124] *CHIPS and Science Act, supra* note 65.

[125] AI Transparency and Accountability Act，S. 3312, 118th Cong. (2024).

[126] Federal Trade Commission Act §5, 15 U.S.C. § 45 (2024).

[127] The Federal Trade Commission has taken legal actions against companies that use AI to mislead consumers through "Operation AI Comply". *See*: Federal Trade Commission, *FTC Announces Crackdown on Deceptive AI Claims and Schemes* (Sep. 25, 2024), https://www.ftc.gov/news-events/news/press-releases/2024/09/ftc-announces-crackdown-deceptive-ai-claims-schemes.

[128] *EO 13960*, *supra* note 42; *AI Blueprint*, *supra* note 113.

[129] U.S. Dep't of Labor, Office of Disability Emp. Policy, *AI & Inclusive Hiring Framework* (Sep. 24, 2024), https://www.dol.gov/sites/dolgov/files/ODEP/pdf/AI-Inclusive-Hiring-Framework.pdf.

[130] The report from Department of Homeland Security highlights that over 100,000 publicly accessible AI-generated nude images are available online without the consent or knowledge of the women depicted, some of which involve child pornography. *See:* U.S. Dep't of Homeland Sec., *Increasing Threats of Deepfake Identities* (Oct 26, 2021), https://www.dhs.gov/sites/default/files/publications/increasing_threats_of_deepfake_identities_0.pdf.

[131] *EO 14110, supra* note 39.

[132] Democratic U.S. Senators Chris Coons and Amy Klobuchar, and Republican Senators Marsha Blackburn and Thom Tillis co-proposed the *Nurture Originals, Foster Art, and Keep Entertainment Safe Act* (NO FAKES Act), which aims to protect individuals' voices from unauthorized use by generative AI. See: S. 4875, 118th Cong., 2d Sess. (2024).

[133] In February 2024, the FCC extended the Telephone Consumer Protection Act (TCPA) to cover AI-generated voices, regulating the use of AI-generated "artificial or prerecorded voice" in communications to protect public rights. See: Fed. Commc'ns Comm'n, *Implications of Artificial Intelligence Technologies on Protecting Consumers from Unwanted Robocalls and Robotexts*, FCC 24-96 (Sept. 10, 2024), https://docs.fcc.gov/public/attachments/DOC-404036A1.pdf.

on transparency in political campaigns, with bills requiring clear labeling of AI-generated content in political advertisements.[134]

Consistent with the approaches in AI facilitation, the U.S. government fosters extensive public engagement in AI regulation, which is evident in the emphasis on public input in the *AI Applications Guidance*[135] and the government's collaborations with tech giants like Amazon, Google, and Meta.[136] These partnerships aim to ensure that AI technologies are developed and deployed responsibly, with shared risk management strategies in place to address the societal impacts of AI.[137]

b. China's Approach to Algorithmic Safety Regulation

The Chinese government primarily regulates AI through lower-level departmental regulations and various legal instruments, with no higher-level comprehensive legislation introduced as of yet.[138] In addition, there are certain regulatory provisions concerning AI within broader laws, such as the *Personal Information Protection Law* which addresses privacy concerns, and the *Regulation on*

Likewise, the Securities and Exchange Commission has filed charges under t*he Securities Exchange Act of 1934 Section 18* against investment advisers for using AI false advertising to protect investors from being deceived by misleading statements made by AI. See:Press Release, U.S. Sec. & Exch. Comm'n, *SEC Charges Two Investment Advisers with Making False and Misleading Statements About Their Use of Artificial Intelligence* (Mar. 18, 2024), https://www.sec.gov/newsroom/press-releases/2024-36.

[134] *REAL Political Advertisements Act*, H.R. 3044, 118th Cong. (2023); *AI Transparency in Elections Act*, H.R. 3044, 118th Cong. (2024); *New York State Political Artificial Intelligence Disclaimer Act,* S. 3875, 118th Cong. (2024);

[135] *AI Applications Guidance***,** *supra* note 55.

[136] The first batch, announced on July 21, 2023, included Amazon, Anthropic, Google, Inflection, Meta, Microsoft, and OpenAI. The second batch, announced on September 12, 2023, included Adobe, Cohere, IBM, Nvidia, Palantir, Salesforce, Scale AI, and Stability AI.

[137] The White House, *Biden-Harris Administration Secures Voluntary Commitments from Leading Artificial Intelligence Companies to Manage the Risks Posed by AI* (July 21, 2023), https://www.whitehouse.gov/briefing-room/statements-releases/2023/07/21/fact-sheet-biden-harris-administration-secures-voluntary-commitments-from-leading-artificial-intelligence-companies-to-manage-the-risks-posed-by-ai/.

[138] E.g., Guanyu Jiaqiang Hulianwang Xinxi Fuwu Suanfa Zonghe Zhili de Zhidao Yijian (关于加强互联网信息服务推荐综合治理的指导意见) [Guiding Opinions on Strengthening Overall Governance of Internet Information Service Algorithms] (promulgated by the Cyberspace Administration of China, Publicity Department of CPC Central Committee, Ministry of Education, Ministry of Science and Technology, Ministry of Industry and Information Technology, Ministry of Public Security, Ministry of Culture and Tourism, State Administration for Market Regulation, National Radio and Television Administration, Sep. 17, 2021), CLI.4.5077312 (Lawinfochina);Hulianwang Xinxi Fuwu Suanfa Tuijian Guanli Guiding (互联网信息服务算法推荐管理规定) [Provisions on the Administration of Algorithm-generated Recommendations for Internet Information Services] (promulgated by the Cyberspace Administration of China, Ministry of Industry and Information Technology of the People's Republic of China, Ministry of Public Security, State Administration for Market Regulation, Mar. 1, 2022), CLI.4.5113084 (Lawinfochina); Hulianwang Xinxi Fuwu Shendu Hecheng Guanli Guiding (互联网信息服务深度合成管理规定) [Provisions on the Administration of Deep Synthesis Internet Information Services] (promulgated by the Cyberspace Administration of China et al., Jan. 10, 2023), CLI.4.5145526 (Lawinfochina); NDRC et al., *supra* note 29.

the Protection of Minors in Cyberspace (2023), which includes provisions on online safety for minors.[139]

The legal instruments outlined above exhibit three key characteristics. First, China prioritizes the development of security technologies to ensure the feasibility of regulation. This emphasis is on creating methods and technologies for security testing and evaluation. The government aims to establish a comprehensive security assurance mechanism with essential capabilities, such as AI security testing and evaluation systems, threat information sharing, and automated response mechanisms.[140]

Second, the instruments adopt an engineering mindset, typically focusing on "prioritizing immediate needs with a focus on practicality and effectiveness".[141] In the early stages of technological development, only non-binding opinions were issued to guide regulatory direction.[142] Later, binding regulations were introduced for only limited application scenarios such as recommendation algorithms, deep synthesis, and generative AI services,[143] following a problem-solving approach rather than seeking to create a comprehensive legal framework for all potential risks and issues. Furthermore, these regulations are primarily departmental in nature. While they hold a lower legal hierarchy, they are swiftly enacted and responsive, bypassing the lengthy legislative process and facilitating the accumulation of regulatory experience.

Third, China's regulatory framework tailors duty provisions to specific entities, reflecting the different roles and functions that each entity plays within the AI ecosystem. These regulations take into account the perspectives of various stakeholders and assign differentiated rights, powers, and obligations to researchers, developers, service providers, service users, and regulators. This ensures that each party is held accountable for its actions within the broader scope of AI development. Therefore, China's regulatory framework can be analyzed from the perspective of the varying obligations assigned to different entities.

For researchers and developers, there are relatively few provisions imposing obligations, including those related to data security and personal information

[139] Geren Xinxi Baohu Fa (个人信息保护法) [Personal Information Protection Law] (promulgated by the Standing Committee of the Standing Comm. Nat'l People's Cong., Aug. 20, 2021, effective Nov. 1, 2021), art. 62, CLI.1.5055321 (Lawinfochina); Weichengnianren Wangluo Baohu Tiaoli (未成年人网络保护条例) [Regulation on the Protection of Minors in Cyberspace] (promulgated by the State Council of the People's Republic of China, Oct. 16, 2023, effective Jan. 1, 2024), art. 26, CLI.2.5180814 (Lawinfochina).

[140] Ministry of Industry and Information Technology, *supra* Note 31; Nat'L People'S Cong., *supra* Note 154.

[141] Zhang Linghan (张凌寒), Zhongguo Xuyao Yibu Zenyang de "Rengong Zhinen Fa"?—Zhongguo Rengong Zhinen Lifa de Jiben Luoji yu Zhidu Jiagou (中国需要一部怎样的《人工智能法》?——中国人工智能立法的基本逻辑与制度架构) [*What Kind of "Artificial Intelligence Law" Does China Need?—The Fundamental Logic and Institutional Framework of China's AI Legislation*], 42 *Science of Law* (Journal of Northwest University of Political Science and Law) no. 3 at 3, 6-7 (2024).

[142] Cyberspace Administration of China et.al., Guiding Opinions on Strengthening Overall Governance of Internet Information Service Algorithms, *supra* note **错误!未定义书签。**.

[143] Zhang Lu (张璐), *Tongyong AI Fengxian Zhili yu Jianguan Chutan—ChatGPT Yinfa de Wenti yu Tiaozhan* (通用 AI 风险治理与监管初探——ChatGPT 引发的问题与挑战) *[An Initial Exploration of General AI Risk Governance and Regulation—Issues and Challenges Raised by ChatGPT]*, 2023 *Electronic Government*, no. 9 at 14, 16-17.

protection.[144] Instead, the regulations primarily target service providers, focusing on key issues such as ethics, the prevention of false information, labeling of algorithm-generated or synthesized content, security assessments and legal liability. With respect to ethics, the regulations stress adherence to mainstream values and the proper political orientation.[145] Security assessment is a consistent regulatory approach in China,[146]which mandates service providers to register essential information with government authorities for swift accountability in case of harm. These assessments are mainly directed at products and applications with public opinion attributes or social mobilization capabilities, though the regulations do not specify how such assessments should be conducted.

Additionally, the regulations require the establishment of an internal control system encompassing various obligations,[147] such as algorithm review, data security, personal information protection, emergency response to security incidents, and identity authentication. Overall, these regulations provide a comprehensive "task list" for companies but leave the specifics of implementation and enforcement to the discretion of service providers. Regarding liability, the regulations primarily prescribe penalties such as warnings, public criticism, and orders for rectification within a specified time frame. For non-compliance or severe violations, penalties may include service suspension or relatively minor fines (ranging from 10,000 to 100,000 RMB). Given the critical role of continuous service provision for major internet platforms in maintaining market competitiveness, the shift from warnings, public criticism, and minor fines to service suspension appears to result in an imbalanced allocation of liability. Furthermore, the regulations fail to clearly define what constitutes a "severe violation".

As the regulator, the government is also tasked with obligations to monitor algorithmic risks and establish a tiered and categorized security management system[148]. However, the provisions concerning these obligations remain somewhat vague and generalized, lacking specific and clear directives for implementation.[149]

[144] Wangluo Anquan Biaozhun Shijian Zhinan — Rengongzhineng Lunli Anquan Fengxian Fangfan Zhiyin (网络安全标准实践指南——人工智能伦理安全风险防范指引) [Cybersecurity Standards Practice Guide - Guidelines for Ethical and Security Risk Prevention in Artificial Intelligence] (promulgated by the Secretariat of the National Information Security Standardization Technical Committee(NISSTC), Jan. 5, 2021), CLI.4.349998 (Lawinfochina).

[145] State Council,*supra* note 27.

[146] Cyberspace Administration of China et.al., Provisions on the Administration of Deep Synthesis Internet Information Services, *supra* note 138.;State Council,*supra* note 27.

[147] Weichengnianren Wangluo Baohu Tiaoli (未成年人网络保护条例) [Regulation on the Protection of Minors in Cyberspace] (promulgated by the State Council of the People's Republic of China, Oct. 16, 2023, effective Jan. 1, 2024), art. 26, CLI.2.5180814 (Lawinfochina).

[148] Cyberspace Administration of China et.al., Guiding Opinions on Strengthening Overall Governance of Internet Information Service Algorithms, *supra* note 138; Cyberspace Administration of China et.al.,Provisions on the Administration of Algorithm-generated Recommendations for Internet Information Services, *supra* note 138; Cyberspace Administration of China et.al., Provisions on the Administration of Deep Synthesis Internet Information Services, *supra* note 138; NDRC et al., *supra* note 29.

[149] Cyberspace Administration of China et.al., Guiding Opinions on Strengthening Overall Governance of Internet Information Service Algorithms, *supra* note 138.

3. Standards Development

Considering the complexity of AI, the implementation of regulatory requirements must be supported by technical standards.

The U.S. government has placed a strong emphasis on establishing clear criteria and mechanisms for identifying and evaluating AI use cases across federal agencies, which is part of a broader strategy to develop and implement comprehensive AI evaluation techniques and technical standards, ensuring uniformity in the adoption of AI technologies.[150] As part of this initiative, the government has outlined the need for robust AI risk management processes, including the setting of uniform standards and methodologies, to help AI practitioners manage risks effectively and consistently.[151]

To further support this approach, ongoing efforts focus on the development of technical standards that will guide AI deployment in both the public and private sectors.[152] These standards are designed to ensure that AI technologies are deployed in a manner that is safe, effective, and aligned with best practices for managing potential risks such as bias and security concerns.[153] This comprehensive framework reflects the U.S. government's commitment to fostering responsible AI development while addressing the challenges posed by emerging technologies.

China also attaches great importance to achieving governance purposes through the formulation and implementation of standards. The current regulatory framework reveals that China is still an administration-led system in standard setting.[154] The Chinese government places a high value on setting standards in the AI sector. It not only specifically mentions developing an AI standard system, but also promulgates

[150] *EO 13960, supra* note 42.

[151] National Institute of Standards and Technology, *U.S. Leadership in AI: A Plan for Federal Engagement in Developing Technical Standards and Related Tools* (Aug. 22, 2019), https://www.nist.gov/system/files/documents/2019/08/10/ai_standards_fedengagement_plan_9aug2019.pdf.

[152] National Institute of Standards and Technology, *Towards a Standard for Identifying and Managing Bias in Artificial Intelligence (NIST Special Publication 1270)* (Mar. 24, 2022), https://nvlpubs.nist.gov/nistpubs/ai/NIST.AI.600-1.pdf.

[153] *Id.*

[154] the *Standardization Law of China* and the *Outline of the National Standardization Development* Biaozhunhua Fa (标准化法) [Standardization Law] (promulgated by the Standing Comm. Nat'l People's Cong., Nov. 4, 2017, effective Jan. 1, 2018), art 5-6,10-11.CLI.1.304266 (Lawinfochina); Guojia Biaozhunhua Fazhan Gangyao (国家标准化发展纲要) [Outline of the National Standardization Development] (promulgated by the Central Committee of the Communist Party of China and the State Council, Oct. 10, 2021), CLI.16.5077460 (Lawinfochina)(Although the Outline of the National Standardization Development proposes that "by 2025, the standard supply mechanism will shift from being government-led to placing equal emphasis on both government and market forces").

more documents later.[155] The Chinese government also sets quantitative goals in this regard.[156] To be specific, by 2026, more than 50 new national and industry standards will have been formulated, and more than 1,000 companies will have advocated and implemented these standards and have engaged in formulating over 20 international standards.[157]

The latest standard system includes eight primary standards and corresponding secondary standards,[158] covering areas like industry, technology, application, products, and services. In terms of concrete actions, China continues to release relevant technical standards on an ongoing basis. As of October 26, 2024, a search for "Artificial Intelligence" on the National Public Service Platform for Standard Information reveals 10 currently effective national recommended standards and 27 standards in draft or under public consultation.[159] These standards span various aspects of AI system development.

As noted above, the Chinese government views the establishment of an AI standard system as essential for the sustainable development of related industries, while also enabling China to exert greater influence in global AI governance. However, based on the available information and the authors' own experience within AI companies, there is a gap between the eagerness of these companies to contribute to the standard formulation and their actual commitment to implementing these standards. This is primarily because the standards they help create lack mandatory enforcement mechanisms.

4. Establishment of Specialized Regulatory/Research Institutions

Both the U.S. and China have recognized the importance of establishing dedicated regulatory and research institutions to manage AI's growth. These efforts also reflect a shared understanding of the need for inter-agency coordination, specialized expertise,

[155] Guojia Xin Yidai Rengongzhineng Biaozhun Tixi Jianshe Zhinan (国家新一代人工智能标准体系建设指南) [Guidelines for the Development of the National New Generation Artificial Intelligence Standard System] (promulgated by the State Administration for Market Regulation, Cyberspace Administration of China, National Development and Reform Commission, Ministry of Science and Technology, and Ministry of Industry and Information Technology, July 27, 2020), CLI.4.344973 (Lawinfochina); Xinxihua Biaozhun Jianshe Xingdong Jihua (2024–2027 Nian) (信息化标准建设行动计划（2024—2027 年）) [Action Plan for the Development of Informatization Standards (2024–2027)] (promulgated by the Cyberspace Administration of China, State Administration for Market Regulation, and Ministry of Industry and Information Technology, May 2024), CLI.16.5193749 (Lawinfochina); Guojia Rengongzhineng Chanye Zonghe Biaozhunhua Tixi Jianshe Zhinan (2024 Ban) (国家人工智能产业综合标准化体系建设指南（2024 版）) [Guidelines for the Construction of the National Comprehensive Standardization System for the Artificial Intelligence Industry (2024 Edition)] (promulgated by the Ministry of Industry and Information Technology et al., June 5, 2024), CLI.4.5196174 (Lawinfochina).

[156] Ministry of Industry and Information Technology et.al., *supra* note 155.

[157] *Id.*

[158] *Id.* Including eight sections: "A Basic Commonalities," "B Supporting Technologies and Products," "C Basic Software and Hardware Platforms," "D Critical General Technologies," "E Key Domain Technologies," "F Products and Services," "G Industry Applications," and "H Security/Governance."

[159] Quanguo Biaozhun Xinxi Gonggong Fuwu Pingtai (全国标准信息公共服务平台) [National Public Service Platform for Standard Information], https://std.samr.gov.cn/ (last visited Dec. 14, 2024) (Use "人工智能" as the key word).

and long-term strategic planning.

The U.S. government is working on establishing an inter-agency committee that is dedicated to algorithmic transparency and addresses risks. Under the *National Artificial Intelligence Initiative Act* of 2020,[160] the National Artificial Intelligence Initiative Office (NAIIO) was created as the central body for coordinating federal AI efforts, fostering inter-agency collaboration, as well as facilitating communication with stakeholders. The National Artificial Intelligence Advisory Committee, established under the leadership of the Office of Science and Technology Policy, provides policy guidance to the U.S. president and relevant federal agencies on AI-related issues.[161] In recent years, the U.S. has further honed its AI governance capabilities through a variety of specialized agencies. For example, in June 2023, the NIST AI Public Working Groups, an initiative under the National Institute of Standards and Technology (NIST), was launched to build upon the *RMF*.[162]

In contrast, China does not have a dedicated agency to regulate AI. Instead, it relies on the coordination of government departments to address emerging risks and challenges. This decentralized approach may result in gaps in oversight, potentially affecting the efficiency and effectiveness of AI regulation.

B. Comparative Analysis of the U.S.-China AI Regulatory Approaches

1. U.S. Approach: Broad Guidelines with Limited Binding Effect

In the approach to AI regulation, the U.S. advocates for building up public trust, prioritizing responses to security risks, setting algorithmic security requirements for privacy protection and algorithm transparency, and encouraging the entire society to engage in establishing a regulatory framework and creating dedicated regulatory agencies. However, it is important to note that the U.S. regulations are more suggestive than mandatory, and prioritize a hands-off approach to avoid stifling industry growth.

The introductory section and preamble of those regulatory documents establish "promoting technological advancement" and "safeguarding the interests of the United States and its people" as the core messages and fundamental principles of regulation. Meanwhile, even for regulatory laws mentioned above, those legal instruments consistently highlight protecting the freedom of scientific research for individuals and businesses, enabling the U.S. to tap the full potential of AI development and innovation.[163]

The current AI regulatory framework in the U.S. tends to be decentralized,

[160] The bill mandates inter-agency collaboration to coordinate AI research, development, and deployment, aiming to drive innovation and maintain U.S. leadership in the field through shared knowledge and resources. The updated version of the bill is currently under review. *See* National Artificial Intelligence Initiative Act of 2020, Pub. L. No. 116-283, div. E, §§ 5001–5509, 134 Stat. 3388, 4523–4560 (2021).

[161] H.R. 6216, 116th Cong. (2020).

[162] Nat'l Inst. of Standards & Tech., *NIST Public Working Group on AI* (June 2023), https://www.nist.gov/artificial-intelligence/nist-public-working-group-ai.

[163] Huw Roberts et al., *Achieving a "Good AI Society": Comparing the Aims and Progress of the EU and the US*, 27 SCI. ENG. ETHICS 68 (2021); William Howey, *How Governments Are Looking to Regulate AI*, ECONOMIST INTELLIGENCE UNIT (2023), https://www.eiu.com/n/how-governments-are-looking-to-regulate-ai/ (last visited Oct 26, 2024).

evolving incrementally to keep pace with the rapid advancements in AI.[164] At the federal level, regulatory progress in the U.S. has been slow, despite the gradual development of AI-related regulatory initiatives in recent years. To date, the U.S. has yet to establish a comprehensive federal regulatory framework for AI, with agencies instructed to avoid actions that could hinder AI innovation.[165] This slow pace can be attributed to several factors, including political polarization, the fragmented nature of the federal system, and the technical complexities inherent in the legislative process.[166] The AI legislation is primarily implemented at the state level, with notable disparities in the legislative progress across states.

Consequently, the U.S. is adopting a rolling legislative approach, with AI-related laws being advanced through the progressive introduction and review of smaller, targeted bills.[167] Given the federal government's limitations in quickly enacting large-scale legislation, state-level regulations are expected to play a more prominent role in shaping AI governance. [168]

2. China’s Approach: Detailed Regulations with Limited Enforcement Clarity

Although China has not issued a large volume of AI-related regulatory documents, its regulatory strategy is nonetheless focused and clear, with an emphasis on specific issues such as recommendation algorithms, deep synthesis technologies, and content services provided by generative AI. The primary concern of these regulations is the use of data by non-state actors, alongside ensuring national security, social stability, and the protection of core values.

It is also worth noting that, China’s AI regulations, while mandatory, are not particularly stringent due to their inherent ambiguity. Although the country has established detailed provisions in several areas, many of these regulations are framed in results-oriented terms, which provide the government with significant flexibility in enforcement. For instance, provisions related to service transparency and reliability are often worded loosely, such as requiring “effective precautions” rather than specifying clear punitive measures. This ambiguity allows the government to adjust regulations as needed, either tightening or loosening them based on emerging circumstances and even rolling back stricter requirements from earlier drafts.[169] Such flexibility underscores a regulatory approach that is reactive,[170] addressing risks as they materialize rather than

[164] “A basic national belief that society will benefit and innovation and creativity will flourish in a system that is free from government control but strengthened through essential governmental participation via effective public-private partnerships.”

[165] *EO 13960, supra* note 42.

[166] Maia Cook, Lobbying on AI Reaches New Heights in 2024, OpenSecrets (June 2024), https://www.opensecrets.org/news/2024/06/lobbying-on-ai-reaches-new-heights-in-2024/

[167] The Bipartisan Senate AI Working Group, *Driving U.S. Innovation in Artificial Intelligence* (May 17, 2024), https://www.schumer.senate.gov/imo/media/doc/Roadmap_Electronic1.32pm.pdf.

[168] N. Turner Lee & J. Turner, *Can California Fill the Federal Void on Frontier AI Regulation?*, BROOKINGS INST. (June 24, 2024), https://www.brookings.edu/articles/can-california-fill-the-federal-void-on-frontier-ai-regulation/.

[169] Zhang, *supra* note 25.

[170] Zhang, *supra* note 25.

preemptively setting rigid rules.[171]

This adaptive nature of China's AI governance reflects the country's prioritization of fostering AI growth and innovation over stringent regulatory control. While this flexibility facilitates rapid development in the AI sector, it may come at the cost of regulatory accountability, potentially leading to inconsistency in enforcement and delays in addressing emerging risks. The lack of precise enforcement mechanisms also means that regulatory compliance may vary, depending on the priorities and interpretations of the authorities at any given time.

To conclude, China's regulatory system is primarily administration-driven, with the central government playing a key role in setting standards. Despite the lack of detailed enforcement provisions, practical obligations such as labeling, filing, and security assessments remain substantial. Ultimately, China's approach to AI regulation balances the need for innovation with a regulatory framework that is flexible and responsive, albeit with some trade-offs in terms of consistency and accountability.

III. THE U.S. AND CHINA'S APPROACH TO INTERNATIONAL COOPERATION AND COMPETITION

The AI landscape is shaped by both international cooperation and competition. Cooperation is crucial for addressing global challenges, ensuring that AI benefits are shared while minimizing harm. However, intense competition also exists as nations race for technological dominance and economic advantage, particularly in strategic sectors like healthcare and defense. While this competition can drive rapid innovation, it may also result in fragmented policies and inconsistent regulatory frameworks across borders.

Strategy for managing this balance of cooperation and competition is often reflected in legal instruments. To navigate these complexities, states must find a balance between cooperation and competition. The following section will examine how the U.S. and China address this balance through their respective legal frameworks and policies concerning international cooperation and competition in AI.

A. Analysis of AI International Cooperation and Competition in the U.S. and China

1. Approaches to International Cooperation

Regarding cooperation on technological research and development, the U.S. has been proactive in fostering global partnerships to advance AI innovation. In May 2019, the U.S. signed the OECD Recommendation on AI, aiming to foster common principles and drive technological innovation.[172] This was followed by the initiation of the *Global Partnership on Artificial Intelligence* at the G7 Science and Technology Ministerial Meeting in June 2020, further promoting AI development in alignment with shared

[171] Huw Roberts et al., *Governing Artificial Intelligence in China and the European Union: Comparing Aims and Promoting Ethical Outcomes*, 39 THE INFORMATION SOCIETY 79 (2023).

[172] OECD, Recommendation of the Council on Artificial Intelligence, OECD Legal Instruments, No. OECD/LEGAL/0449 (May 22, 2019), https://legalinstruments.oecd.org/en/instruments/oecd-legal-0449.

values.[173] More recently, it advocates for AI research collaboration, while leading a UN resolution urging governments to support inclusive AI development in 2024.[174]

Meanwhile, China has adopted a similarly active approach to international AI relations on multiple fronts. It values international cooperation through ethical rules and initiatives, upholding a people-centered approach and promoting AI for good.[175] China also actively expands its overseas market through its policies,[176] supporting mergers, acquisitions, and R&D centers, and accelerating AI applications. It also proposes accelerating the promotion and application of AI technologies in countries along the "Belt and Road", enhancing China's global presence in AI technology and fostering technological growth in developing nations.[177]

Regarding international standard-setting, the U.S. has focused on its leadership in global AI governance. The 2024 *NIST Global Engagement Plan for AI Standards* reinforces this approach, outlining U.S. efforts to promote AI standardization based on its own framework.[178] China also places great emphasis on engagement in formulating international standards. It promotes policies to encourage research institutions and businesses to contribute to the formulation of global AI standards, driving the international adoption of its technologies.[179]

Both the U.S. and China promote international cooperation in AI through talent exchange programs, recognizing the importance of cross-border cooperation to enhance research and development. In September 2020, the Trump administration signed a declaration with the U.K. to foster exchanges and collaborations between researchers and students, reflecting a shared commitment to driving innovation and leadership in AI.[180] Similarly, China has implemented initiatives supporting AI professionals in engaging in academic exchanges abroad.[181] Additionally, China encourages the establishment of joint labs for international AI cooperation, particularly within its higher education institutions, to foster global research collaboration and innovation.[182]

[173] Muhammed Can & Halid Kaplan, *Transatlantic Partnership on Artificial Intelligence: Realities, Perceptions and Future Implications, 6* GLOBAL AFFAIRS, *537-550 (2024).*
[174] G.A. Res. A/78/L.49, Seizing the Opportunities of Safe, Secure and Trustworthy Artificial Intelligence Systems for Sustainable Development, (Mar. 11, 2024).
[175] Quanqiu Rengongzhineng Zhili Changyi (全球人工智能治理倡议) [Global AI Governance Initiative] (promulgated by CAC, Oct. 18, 2023), CLI.4.5180312 (Lawinfochina).
[176] the *NGAIDP*, the *Three-Year Action Plan for "Internet Plus" AI* and the *Three-Year Action Plan for Promoting the Development of a New Generation of Artificial Intelligence Industry (2018-2020)*
[177] State Council,*supra* note 27; NDRC et al., *supra* note 32; Ministry of Industry and Information Technology, *supra* note 31.
[178] National Institute of Standards and Technology, *A Plan for Global Engagement on AI Standards* (Aug. 2024), https://nvlpubs.nist.gov/nistpubs/ai/NIST.AI.100-5.pdf.
[179] the documents such as the *Three-Year Action Plan for "Internet Plus" AI* and the *Interim Measures for the Management of Generative Artificial Intelligence Services* NDRC et al., *supra* note 32; NDRC et al., *supra* note 29.
[180] U.S. Dep't of State, *Declaration of the United States of America and the United Kingdom of Great Britain and Northern Ireland on Cooperation in Artificial Intelligence Research and Development: A Shared Vision for Driving Innovation and Leadership in Artificial Intelligence* (2024), https://www.state.gov/declaration-of-the-united-states-of-america-and-the-united-kingdom-of-great-britain-and-northern-ireland-on-cooperation-in-artificial-intelligence-research-and-development-a-shared-vision-for-driving/.
[181] NDRC et al.,*supra* note 32; Ministry of Education, *supra* note 46.
[182] State Council,*supra* note 27; Ministry of Education, *supra* note 46.

2. Approaches to International Competition

At the international level, the U.S. places significant emphasis on maintaining and securing its technological leadership. The U.S. emphasizes that it must maintain leadership in AI and shape global AI development in accordance with its own values and priorities,[183] with repeated stress on leadership in all AI-related areas.[184] This focus on leadership is a central theme in various policy frameworks, such as the *CHIPS and Science Act*, which aims to solidify U.S. advantages in technological competition, particularly against China, by channeling significant investments into key areas of research and development.[185] Similarly, the *National Artificial Intelligence Initiative Act* proposes to strengthen U.S. global leadership in AI through technological breakthroughs and multi-sectoral synergies to protect national interests.[186] A more radical proposal is the "AGI Manhattan Project," advocated by the *U.S.-China Economic and Security Review Commission* in its 2024 Annual Report. This initiative aims to expedite the development of artificial general intelligence (AGI) to secure a decisive edge over China in the AI race.[187] For specific actions, the U.S. Department of Justice unveiled the "China Initiative" in 2018 and underscored the importance of protecting core technology.[188] In the same year, the National Institutes of Health initiated hundreds of investigations into scientists and researchers, thus causing a drop in U.S.-China research collaborations in science and technology.[189]

Both countries' legislative and executive frameworks surrounding AI are marked by a proactive approach to national security. The Department of State is mandated to assess the risks, with regular updates on AI defenses against potential military threats from adversaries.[190] In line with this, the U.S. continues to invest in military applications of AI to enhance defense capabilities.[191] The *House Select Committee on the Strategic Competition between the United States and the Chinese Communist Party* has been established to investigate the technology competition between the two countries,[192] focusing particularly on ensuring that U.S. investments do not inadvertently support adversarial interests. On the defense front, China's AI strategy emphasizes civil-military integration, fostering the dual-use transformation of AI technologies and encouraging civilian research to support national defense innovations.[193] Moreover, China is focusing on building a unified AI technology

[183] *EO 13859, supra* note 30.
[184] *AI Strategic Plan 2023, supra* note 31.
[185] *CHIPS and Science Act*, *supra* note 65.
[186] *EO 13960, supra* note 42.
[187]U.S.-China Econ. & Sec. Review Comm'n, 2024 Annual Report to Congress (2024), https://www.uscc.gov/sites/default/files/2024-11/2024_Annual_Report_to_Congress.pdf.
[188] U.S. DEP'T OF JUSTICE, INFORMATION ABOUT THE DEPARTMENT OF JUSTICE'S CHINA INITIATIVE AND A COMPILATION OF CHINA-RELATED PROSECUTIONS SINCE 2018 (2021).
[189] Jia R, Roberts ME, Wang Y, Yang E. *The impact of US-China tensions on US science: Evidence from the NIH investigations,* 19 PROC NATL ACAD SCI U S A., May 7, 2024, at 121.
[190] U.S. Nat'l Sci. Found., s*upra* note 75.
[191] National Defense Authorization Act for Fiscal Year 2024, Pub. L. No. 117-31, 137 Stat. 1092 (2023).
[192] H.R. Res. 11, 118th Cong. (2023).
[193] State Council, *supra* note 27; Ministry of Education, *supra* note 46; Nat'L People's Cong., *supra* note 28.

standard system that applies to both the military and civilian sectors, with an emphasis on fostering innovation through joint technology platforms.[194]

As part of its strategy to maintain technological leadership, the U.S. has imposed strict controls to prevent the unauthorized dissemination of technology that could undermine the country's technological advantages.[195] While AI is not classified as a distinct export category, the U.S. regulates AI-related technologies through restrictions on integrated circuits, semiconductors, AI design and development software, chip manufacturing equipment, and other related technologies. Executive orders further emphasize the need to monitor AI-related transactions that could pose national security risks,[196] particularly focusing on preventing strategic competitors from acquiring critical AI technologies.[197] New initiatives, such as the *Protecting Americans' Data from Foreign Adversaries Act (2024),* also aim at safeguarding sensitive personal data from foreign adversaries and regulating illegal data broker activities.[198] Subsequently, from 2020 to 2024, the U.S. imposed multiple rounds of sanctions and pressured other countries to cease exporting AI-related technology to China.[199]

As for China, China's export controls on AI technologies, while less extensive than the U.S., target specific areas such as algorithmic recommendation services.[200] These restrictions are more narrowly focused on regulating exports rather than imposing comprehensive bans.

B. Comparative Analysis of the U.S.-China AI International Policies

Both the U.S. and China recognize the crucial role of international cooperation in advancing AI, but they adopt different strategies to shape global AI governance.

The U.S. seeks to maintain its leadership in AI by pursuing formal partnerships and playing a dominant role in shaping international standards. By emphasizing its leadership position in the global AI landscape, the U.S. not only advocates for ethical AI practices but also safeguards its technological supremacy. Approaches such as export controls and oversight of international AI transactions are central to this strategy, enabling the U.S. to control the dissemination of key technologies that could potentially undermine its position.

[194] State Council, *supra* note 27; Ministry of Education, *supra* note 46; Nat'L People's Cong., *supra* note 28.

[195] U.S. export controls do not classify artificial intelligence as a distinct category but regulate it indirectly through related technologies and hardware, including integrated circuits, semiconductors, AI design and development software, chip manufacturing equipment, and associated technologies.

[196] Enhancing National Frameworks for Overseas Restriction of Critical Exports Act, H.R. 8315, 118th Cong., 2d Sess. (2024).

[197] Exec. Order No. 14,117, 89 Fed. Reg. 15,421 (2024).

[198] Protecting Americans' Data from Foreign Adversaries Act, H.R. 7520, 118th Cong., 2d Sess. (2024).

[199] Additions and Modifications to the Entity List; Removals From the Validated End-User (VEU) Program, 89 Fed. Reg. 96830 (Dec. 5, 2024), https://www.federalregister.gov/public-inspection/2024-28267/additions-and-modifications-to-the-entity-list-removals-from-the-validated-end-user-program

[200] Guanyu Gongbu "Zhongguo Jinzhi Chukou Xianzhi Chukou Jishu Mulu" de Gonggao (关于公布《中国禁止出口限制出口技术目录》的公告) [Announcement on the Publication of the Catalogue of Technologies Prohibited or Restricted for Export in China] (promulgated by the Ministry of Commerce and the Ministry of Science and Technology, Dec. 21, 2023), CLI.4.5185071 (Lawinfochina).

In contrast, China aims to strengthen its global AI influence through a different approach, one that emphasizes expanding its presence, particularly in developing countries. Central to China's strategy is the development of a unified AI standard system that bridges different sectors and promotes dual-use technologies, ensuring alignment between technological innovation and national priorities. China's focus on international exchanges, collaborations, and partnerships—especially with emerging economies—reflects its broader ambition to enhance its influence in global AI governance.

When it comes to AI competition, the U.S. and China adopt distinct strategies. The U.S. enforces strict export controls on a range of AI-related hardware and development tools, taking a proactive approach to monitoring and restricting transactions involving these technologies. This ensures that core U.S. technologies do not flow to adversaries who could use them to catch up quickly. Conversely, China does not impose similar export restrictions. This contrast in approach may be linked to the technological status of the two nations. As a technological leader, the U.S. seeks to prevent the rapid global dissemination of its advanced AI technologies to maintain its competitive edge. China, as a latecomer in AI development, seeks to learn from other countries, enriching its own technological base by promoting international exchanges of knowledge and technology.

IV. DISCUSSION: THE "MAINTAINING LEADERSHIP VS. CATCHING UP" DYNAMICS

Having explored the similarities and differences of the legal instruments of the two countries, the next question arises: What factors have shaped these divergent approaches? This section will delve into the underlying causes of these differences, offering crucial insights into the distinct paths the U.S. and China have taken.

A. AI Development: Diverging Stages of Progress

The divergent approaches to AI development and governance between the U.S. and China are rooted in the distinct stages of their technological development. Technologically, the U.S. and China are situated in a dynamic of maintaining leadership vs. catching up—the U.S. is positioned as the global leader in AI technology, whereas China is still in the process of catching up.

In 2023, the U.S. private investment reached $67.2 billion, 8.7 times greater than China's. Additionally, the U.S. had 61 machine learning models, four times that of China.[201] These figures highlight the U.S.'s dominant position in both the scale and scope of AI development. Moreover, the U.S. leads in the more innovative aspects of AI technology,[202] such as foundational theories, original algorithm research, and the development of high-end devices,[203] leaving China at a disadvantage in these critical

[201] Nestor Maslej et al., *The AI Index 2024 Annual Report*, AI INDEX STEERING COMMITTEE, INSTITUTE FOR HUMAN-CENTERED AI, STANFORD UNIVERSITY (2024), https://aiindex.stanford.edu/wp-content/uploads/2024/04/HAI_2024_AI-Index-Report.pdf.

[202] AlShebli, B., Cheng, E., Waniek, M. et al., *Beijing's central role in global artificial intelligence research*, 12 SCI REP, 21461 (2022).

[203] Wen Gao, *Analysis of Cutting-Edge Technologies in Artificial Intelligence and High-Quality Development*, https://www.pcl.ac.cn/html/943/2023-12-30/content-4361.html (last visited Oct 26, 2024).

areas.

The majority of China's AI research is more focused on practical applications and functionalities rather than on fundamental advancements. Chinese R&D has been concentrated on areas like AI hardware production lines and specific AI application scenarios. [204] This approach has contributed to China's underperformance in foundational algorithm models and patent applications.

Furthermore, the discrepancy between the U.S. and China is also reflected in the composition of AI talent. China's talent pool remains significantly less diversified than that of the U.S.[205] In the U.S., talent is distributed more evenly across foundational (22.8%), technical (37.3%), and application (39.9%) tiers. In contrast, China's talent composition is heavily skewed toward application (61.8%), with much lower percentages in foundational (3.3%) and technical (34.9%) roles.[206] Moreover, China faces challenges in retaining AI talent also due to limited incentives, inadequate resources, and poor coordination within the AI R&D ecosystem[207] These factors contribute to the loss of AI professionals, hindering China's progress in this area.[208]

Despite these challenges, China has been making significant strides in closing the technological gap with the U.S. Over time, China's increasing investments in AI research and development, as well as its rising role in global science, have allowed it to contest the U.S.'s decades-long dominance.[209] In 2012, the U.S. led global AI research investment with $656 billion (27% of the total), while China invested $526 billion (22%).[210] By 2024, China contributed 32% of global semiconductor output, compared to the U.S.'s 18%, illustrating China's growing capacity in high-tech industries that are central to AI development.[211]

In summary, the differences in the stages of AI development in the U.S. and China reflect not only the technological capabilities of each country but also the strategic approaches they adopt in AI governance. The U.S., as a leader, focuses on securing its position and ensuring technological supremacy, while China, as a latecomer, emphasizes practical applications to bridge the gap.

[204] Daniel Zhang et al., *The AI Index 2021 Annual Report*, AI INDEX STEERING COMMITTEE, INSTITUTE FOR HUMAN-CENTERED AI, STANFORD UNIVERSITY (2021), https://aiindex.stanford.edu/wp-content/uploads/2021/11/2021-AI-Index-Report_Master.pdf.

[205] Wen Gao, *How Artificial Intelligence is Changing Us*, CHINESE ACADEMY OF ENGINEER (Nov. 16, 2023), https://www.cae.cn/cae/html/main/col35/2023-11/16/20231116195656494726698_1.html.

[206] Zhejiang University & Baidu, *2022 Plans for China AI Talent Development*, XINHUANET (Jan. 25, 2022), https://www.xinhuanet.com/info/20220125/9fa3a71c89124f9fb0e4ba3cf9d7ce72/c.html.

[207] Wei Zhang, Yuhua Cai & Qixiang Cai, *On the Application and Challenges of AI in China*, 15 CHINESE JOURNAL OF TECHNOLOGY AND LAW (2018) (China).

[208] Qian Peng, Hualing Li, *A Closer Look at the Five Major Shortcomings of China's AI Talent System*, XIN HUA (Aug. 28, 2019, 17:31), https://www.gov.cn/xinwen/2019-08/28/content_5425310.htm.

[209] Zhou, P. & Leydesdorff, L., *The Emergence of China as a Leading Nation in Science*, 35 RESEARCH POLICY 83 (2006); Marginson, S., *"All Things Are Influx": China in Global Science*, 83 HIGHER EDUCATION 881 (2022).

[210] National Science Foundation, *Production and Trade of Knowledge- and Technology-Intensive Industries,* https://ncses.nsf.gov/pubs/nsb20226/enabling-technologies (2022).

[211] National Science Foundation, *Production and Trade of Knowledge- and Technology-Intensive Industries,* https://ncses.nsf.gov/pubs/nsb20247/introduction (2024).

B. Government's Role in AI Facilitation

The differences in facilitative laws between the U.S. and China are heavily influenced by their stages of technological development, with the U.S. focused on "maintaining leadership" and China aiming to "catch up." The U.S. has firmly established itself as a leader in the AI industry, with its leadership emerging from a market-driven model where government intervention is often minimal, as over-regulation could be counterproductive. In contrast, China, as a latecomer, faces relative resource constraints and has therefore adopted a strategy to bridge the gap through extensive investment in AI development. This catch-up strategy relies on systematic planning, to optimize resource efficiency and maximize developmental effectiveness, aiming to not only ultimately catch up with but potentially surpass existing AI leaders.

In the U.S., the government employs a less interventionist approach, mainly focusing on setting broad policy frameworks. The U.S. government conveys specific technological needs to the private sector, which responds by aligning its solutions accordingly through government contracts, technical standards, and policy guidance. This model is built upon the confidence that the private sector, particularly major tech companies, can effectively drive AI development given their existing technological expertise and substantial investment capabilities. Another reason is that, given the U.S.'s clear leadership in AI, many U.S. tech giants possess a deep understanding of both AI technology and the AI sector, coupled with greater investment capabilities in the field. Hence, rather than overly intervening with tech giants in this sector, it is more appropriate to align with and respect market forces and to have confidence in the capabilities of the private sector, while keeping a close eye on the movements of major competing countries and respond accordingly. This approach allows the U.S. to leverage its private sector's flexibility and cutting-edge technology to maintain its leadership role in the AI ecosystem.

In contrast, China's "new system for mobilizing resources nationwide" and similar policies are not exclusive to any ideology but a necessity, for later-mover countries, drawing on past experiences. Japan and South Korea, for example, used deep government intervention post-World War II to boost industrial growth. Japan achieved this by importing technologies, providing financial support, and implementing government-led policies to nurture key industries.[212] South Korea and others similarly used deep government intervention to drive industrial growth.[213] China's early stage in AI industrialization necessitates a more active government role. The Chinese government recognizes the need for substantial backing and policy-driven incentives to foster technological maturity, aiming to address the gaps and imbalances. The government's involvement extends to local governments, holding them accountable through performance assessments and pushing national AI goals. As such, the Chinese government has a central role in shaping the AI industry's development, aiming to address existing gaps.

The distinction between these "market-led" and "government-led" approaches highlights how each country's circumstances—particularly their stage in technological

[212] Hiroyuki Odagiri & Akira Goto, *Technology and Industrial Development in Japan: Building Capabilities by Learning*, INNOVATION AND PUBLIC POLICY, 44–51 (1996).
[213] Kwan S. Kim, *The Korean Miracle (1962-1980) Revisited*, 12(Kellogg Inst. for Int'l Stud., Working Paper No. 166, 1991).

development—shape their facilitative laws. While ideological differences certainly play a role, the primary driver of these disparities is the different technological landscapes and the strategies each country employs to position itself within the global AI competition.

C. Regulatory Flexibility as a Strategic Concession to AI Development

The dynamic of "maintaining leadership vs. catching up" plays a significant role in shaping the regulatory strategies of both the U.S. and China. Both countries prioritize laws that enable technological development, downplaying the regulatory function to avoid stifling innovation. This shared strategic approach underscores a broader commitment to securing competitive advantages in the global AI race.

In the U.S., the regulatory framework emphasizes "light-touch regulation" and is designed to minimize barriers to AI development and promote continuous innovation. The primary concern is to preserve technological leadership by avoiding regulatory measures that could impede the rapid pace of AI advancement. Rather than imposing heavy-handed regulations, the U.S. seeks to foster an environment conducive to AI innovation, ensuring that it stays at the forefront of AI advancements. This is exemplified by Executive Order 14179, titled "*Removing Barriers to American Leadership in Artificial Intelligence*," issued on January 23, 2025, which calls for the revocation of existing policies that impede AI innovation and underscores the U.S.'s commitment to eliminating regulatory obstacles that could hinder its dominance in the sector.[214] Through this, the U.S. aims to maintain its position as a global leader in AI by focusing on flexibility and market-driven innovation, rather than rigid regulatory frameworks that could stifle progress.

For China, the government's approach similarly emphasizes the facilitation of AI development, with an emphasis on integrating safety and innovation. The "*Interim Measures for the Management of Generative Artificial Intelligence Services*" prescribes the principles of "*equal focus on development and safety*," alongside "a commitment to integrating innovation with legal governance." The comparison of 37 legal instruments shows that 23 are focused on facilitating AI development, while only 14 are regulatory in nature. These instruments are geared more toward incentivizing innovation than imposing strict regulation, with considerable attention given to fostering technological applications and breakthroughs.[215] Premier Li Qiang stated, "On the basis of ensuring security, we should actively pursue inclusive and prudent regulation and grant new technologies sufficient room for innovation and also necessary room for trial and error,"[216] reflecting China's approach to ensuring that AI technologies are given room to develop, while also mitigating potential risks. This demonstrates China's commitment to fostering an environment where technological progress can flourish, but without neglecting the need for prudent safety measures.

[214] Executive Order No. 14179, 86 Fed. Reg. 35617 (July 9, 2021).

[215] Hine, E., Floridi, L., *Artificial intelligence with American values and Chinese characteristics: a comparative analysis of American and Chinese governmental AI policies*, 39 AI & SOC, 257–278 (2024).

[216] Wei Zou, Li Qiang*, Emphasizes During His Research in Beijing the Need to Promote the Deep Integration of Technological and Industrial Innovation, Accelerating the Creation of New Drivers and Advantages for High-Quality Development*, XIN HUA (Mar. 13, 2024, 22:26), http://politics.people.com.cn/n1/2024/0313/c1024-40195316.html.

In summary, both the U.S. and China share a strategic choice prioritizing facilitation. This approach reflects their mutual recognition of the potential benefits of rapid technological progress. By refraining from premature regulatory actions, both countries aim to foster innovation while cautiously addressing the risks associated with AI technologies.

D. Emphasis on Global Competition for AI Technological Superiority

The aforementioned "maintaining leadership vs. catching up" dynamic also prominently shapes how the U.S. and China approach international cooperation and the development of international competitive law. The U.S. prioritizes efforts to maintain its leadership in AI by strengthening alliances and promoting global standards that reflect its values while taking active measures to safeguard its technological edge. In contrast, China adopts a strategy of collaboration and participation, focusing on enhancing its global standing through partnerships and international cooperation.

For the U.S., international coordination is strategically focused on ensuring that AI technologies developed in collaboration with its allies align with American values, while simultaneously safeguarding its technological dominance. This approach is consistently evident in policy documents. For example, *EO 13859* underscores the "continued American leadership in AI" as vital to both national and economic security. Its hostility toward China is clearly outlined in the *National Security Commission on Artificial Intelligence's 2022 report* stresses the need to "win the AI competition that is intensifying strategic competition with China".[217] AI is viewed as a key pillar for maintaining global influence, justifying proactive measures like export controls and sanctions to curtail the technological advancements of competitors, particularly China. These measures align with the broader aim of preserving U.S. leadership and economic security, emphasizing protectionism to secure technological superiority.

On the other hand, China's strategy focuses more on "catching up" and enhancing its standing within the global AI landscape. Unlike the U.S., which uses aggressive protectionist measures, the U.S. is not specifically mentioned as an adversary in any of China's current AI policy documents. China's strategy encourages international cooperation, aligning with global technological trends and fostering partnerships to drive its AI ambitions. As President Xi Jinping highlighted, "Whoever can seize the opportunities of AI and big data will be at the forefront of the times,"[218] AI is identified as a critical area of international competition. The *Interim Measures for the Management of Generative Artificial Intelligence Services* prescribes that "it is encouraged to engage in establishing international regulations concerning generative AI".[219] The absence of export controls in China's policies also indicates that it is more inclined to leverage collaboration as a path to technological growth, rather than curbing

[217] NAT'L SEC. COMM'N ON ARTIFICIAL INTELLIGENCE, FINAL REPORT OF THE NATIONAL SECURITY COMMISSION ON ARTIFICIAL INTELLIGENCE (2021).

[218] YuChen Duan, *Empowering High-Quality Development with Artificial Intelligence*, QIUSHI THEORY (Apr. 13, 2024, 11:39 AM), http://www.qstheory.cn/dukan/hqwg/2024-04/13/c_1130108914.htm.

[219] NDRC et al., *supra* note 29.

others' progress.[220] These documents suggest that while China is keenly aware of AI's critical role in global competition, its approach remains more oriented toward long-term international partnerships, rather than immediate leadership dominance.

In conclusion, the differing approaches to international cooperation and competitive law are reflective of the dynamic of "maintaining leadership vs. catching up". The U.S. pursues a protectionist strategy, focusing on safeguarding its leadership, while China adopts a more collaborative, long-term approach, aiming to catch up and establish itself as a competitive force on the world stage. Both nations recognize AI's strategic importance, but they navigate the global competition in fundamentally different ways,[221] shaped by their contrasting priorities and stages of technological development.

CONCLUSION

Adopting a functionalist perspective, this paper systematically analyzes how the U.S. and China have developed and regulated emerging AI technologies. At first glance, it might seem that the U.S. and China have adopted fundamentally different approaches to AI facilitation and regulation, driven by distinct underlying philosophies. However, when viewed through the lens of the different stages of AI development and considering the broader context of government-market dynamics, the underlying cause of these differences appears to be more about their respective technological development stages than ideological conflicts. In other words, the U.S. and China are actually responding to the pressures of the "maintaining leadership vs. catching up" dynamic.

The U.S. takes a measured approach to facilitative law, without heavy involvement or specific development goals, reflecting its AI leadership and the desire to maintain its technological edge. In contrast, China prioritizes long-term AI growth, actively guiding market actions to accelerate development. Regarding regulation, the U.S. prefers non-binding measures to avoid stifling innovation, whereas China's mandatory but often ambiguous regulations are designed to expedite technological progress. On the international stage, the U.S. seeks to maintain leadership while China focuses on strengthening itself through collaboration and partnerships, embodying a more cautious, catch-up mindset. As a leader in AI technology, the U.S. is able to maintain its leadership with minimal intervention and a market-respecting approach, whereas China, as a newcomer in AI, needs to leverage government influence to align technological development and industry resources in order to catch up. Both countries, fueled by their desire for development, have taken a similar stance that prioritizes facilitation over regulation. The attention both countries give to securing international competitive advantages further validates this mindset of competing through development.

When disregarding the differences in technological development stages and economic policies, the two countries exhibit a similar attitude toward AI: prioritizing

[220] The *New Generation AI Development Plan* states, "Major developed countries around the world regard the development of AI as a significant strategy to enhance national competitiveness and safeguard national security... striving to gain a dominant position in the new round of international technological competition."

[221] Zaidan, E., Ibrahim, I.A., *AI Governance in a Complex and Rapidly Changing Regulatory Landscape: A Global Perspective*, 11 HUMANITIES AND SOCIAL SCIENCES COMMUNICATIONS 11-21 (2024).

development over regulation due to the emphasis on international competitive advantages. The shared mindset may complicate efforts for global cooperation in AI governance, as both the U.S. and China aim to dominate the regulatory discourse by creating systems and technical standards that align with their own discourse. Such competition could also provoke backlash from other countries, particularly the EU, and hinder efforts to establish a cohesive international regulatory framework.

The findings prompt rethinking about exacerbating global regulatory conflicts, which may contribute to the "Collingridge" dilemma[222]. Countries, in their rush to gain a first-mover advantage in emerging technologies, may underestimate the risks posed by AI, resisting effective regulation and international governance. This challenge—more perilous than ideological confrontations—calls for a deeper reflection on how global AI standards and regulations can be shaped to prevent one country from imposing its model on the rest. Given the increasing dominance of the U.S. and China in the AI narrative, the imbalance of power may escalate tensions, underscoring the need for a balanced approach that promotes both national interests and global cooperation. Excessive competition could lead to divisiveness, undermining efforts to establish universal technical standards that foster fair and safe AI development worldwide. In light of these challenges, a broader reflection on how to balance national interests with global cooperation becomes critical in addressing the future of AI governance.

[222] DAVID COLLINGRIDGE, THE SOCIAL CONTROL OF TECHNOLOGY 19 (1980).

[Annex 1] The U.S. Policy Documents Referenced and Discussed

No	Title	Issuing Authority	Date of Issue	Category	Note
1	Maintaining American Leadership in Artificial Intelligence	The White House	02/11/2019	Executive Order 13859	
2	Recommendation of the Council on Artificial Intelligence	OECD	05/22/2019	OECD Legal Instruments	
3	U.S. LEADERSHIP IN AI: A Plan for Federal Engagement in Developing Technical Standards and Related Tools	National Institute of Standards and Technology	08/09/2019	NIST Publications Supplementary Document of Executive Order 13859	
4	National Artificial Intelligence Initiative Act of 2020	House of Representatives, U. S.	03/12/2020	Federal Law	Referred to the House Committee on Science, Space, and Technology.
5	A Shared Vision for Science and Technology in Responding to the Pandemic, Protecting Human Health, and Promoting Social and Economic Recovery	G7 Conference	05/28/2020	G7 Conference Document	
6	AI in Government Act of 2020	House of Representatives, U. S.	09/15/2020	Federal Law	Received in the Senate. Read twice.

7	America and the United Kingdom of Great Britain and Northern Ireland on Cooperation in Artificial Intelligence Research and Development: A Shared Vision for Driving Innovation and Leadership in Artificial Intelligence	U.S. Department of State	09/25/2020	International Document	
8	Guidance for Regulation of Artificial Intelligence Applications	The White House, Office of Management and Budget	11/17/2020	Memorandum	
9	Promoting the Use of Trustworthy Artificial Intelligence in the Federal Government	The White House	12/03/2020	Executive Order 13960	
10	AI Training Act	House of Representatives, U. S.	08/04/2021		Became Public Law No: 117-207.
11	CHIPS and Science Act	House of Representatives, U. S.	08/09/2022	Federal Law	Became Public Law No: 117-167.
12	Ensuring Robust Consideration of Evolving National Security Risks by the	The White House	09/15/2022	Executive Order 13960 EO 14803	

	Committee on Foreign Investment in the United States				
13	Blueprint for an AI Bill of Rights	White House Office of Science and Technology Policy	10/12/2022	Supplementary Document of Executive Order 14110	
14	Policy Statement Regarding the Scope of Unfair Methods of Competition Under Section 5 of the Federal Trade Commission Act: Unfair or Deceptive Acts or Practices	Federal Trade Commission	11/10/2022	Federal Law	
15	Establishing the Select Committee on the Strategic Competition Between the United States and the Chinese Communist Party	House of Representatives, U. S.	01/10/2023	Resolution	Agreed to in House
16	Artificial Intelligence Risk Management Framework (AI RMF 1.0)	National Institute of Standards and Technology	01/26/2023	NIST Publications	
17	REAL Political Advertisements Act	House of Representatives, U. S.	05/02/2023	Federal Law	Referred to the House Committee on House

					Administration.
18	National Artificial Intelligence Research and Development Strategy Plan 2023 Update	Select Committee on Artificial Intelligence of the National Science and Technology Council	05/04/2023	Policy Report	
19	Executive Order on Addressing United States Investments in Certain National Security Technologies and Products in Countries of Concern	The White House	08/09/2023	Executive Order 14105	
20	Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence	The White House	10/03/2023	Executive Order 14110	
21	National Defense Authorization Act for Fiscal Year 2024	House of Representatives, U. S.	12/22/2023	Federal Law	Became Public Law No: 118-31.
22	Preventing Access to Americans' Bulk Sensitive Personal Data and United States Government-Related Data by Countries of Concern	The White House	02/28/2024	Executive Order 14117	Preventing Access to Americans’ Bulk Sensitive Personal Data and United States Government-Related

					Data by Countries of Concern
23	Implications of Artificial Intelligence Technologies on Protecting Consumers from Unwanted Robocalls and Robotexts,	Federal Communications Commission	02/28/2024	Declaratory Ruling	
24	Part II, Chapter 11 of the National Security Commission on Artificial Intelligence's final report	National Security Commission on Artificial Intelligence	03/2024	Policy Report	
25	Seizing the opportunities of safe, secure and trustworthy artificial intelligence systems for sustainable development	Resolutions of the 78th Session - UN General Assembly	03/11/2024	U.N General Assembly Resolution	
26	Protecting Americans' Data from Foreign Adversaries Act of 2024 (PADFA)	House of Representatives, U. S.	04/24/2024	Federal Law	Received in the Senate and Read twice and referred to the Committee on Commerce, Science, and Transportation.
27	Enhancing	House of	05/08/2024	Federal Law	Ordered to

	National Security Through Exports Control Framework （ENFORCE Act）	Representatives, U. S.			be Reported (Amended)
28	AI Transparency in Elections Act of 2024	House of Representatives, U. S.	05/15/2024	Federal Law	Placed on Senate Legislative Calendar under General Orders. Calendar No. 389.
29	FAA Reauthorization Act of 2024	House of Representatives, U. S.	05/16/2024	Federal Law	Became Public Law No: 118-63.
30	Driving U.S. Innovation in Artificial Intelligence: A Roadmap for Artificial Intelligence Policy in the United States Senate	The Bipartisan Senate Artificial Intelligence (AI) Working Group	05/17/2024	Policy Report	
31	Artificial Intelligence Risk Management Framework: Generative Artificial Intelligence Profile (NIST-AI-600-1)	National Institute of Standards and Technology	07/25/2024	NIST Publications	
32	A Plan for Global Engagement on AI	National Institute of Standards and Technology	07/26/2024	NIST Publications	

	Standards				
33	Artificial Intelligence Research, Innovation, and Accountability Act of 2023	House of Representatives, U. S.	07/31/2024	Federal Law	Committee on Commerce, Science, and Transportation. Ordered to be reported with an amendment.
34	The Nurture Originals, Foster Art, and Keep Entertainment Safe Act（NO FAKES Act）	House of Representatives, U. S.	07/31/2024	Federal Law	Read twice and referred to the Committee on the Judiciary.
35	*AI & Inclusive Hiring Framework*	U.S. Department of Labor	09/24/2024	Supplementary Document	
36	U.S.-China Economic and Security Review Commission, 2024 Annual Report to Congress.	U.S.-China Economic and Security Review Commission	11/19/2024	Policy Report	
State law					
1	An Act Concerning Artificial Intelligence, Automated Decision-Making, and Personal Data Privacy	Connecticut General Assembly	06/07/2023	State law	Enacted
2	New York Political Artificial	New York General	05/10/2023	State law	Amend and recommit to

	Intelligence Disclaimer Act（PAID Act）	Assembly			election law
3	California SCR 17, Dodd. Artificial Intelligence	California General Assembly	08/23/2023	State law	Chaptered by Secretary of State. Res. Chapter 135, Statutes of 2023.

[Annex 2] U.S. Enacted State-Level Laws on Transparency & Privacy Protection

No	Title	Effective Date	Category
1	SB-1001 California Bots: disclosure	09/28/2018	BOT makes it unlawful for a person or entity to use a bot to communicate or interact online with a person in California in order to incentivize a sale or transaction of goods or services or to influence a vote in an election without disclosing that the communication is via a bot.
2	The Virginia Consumer Data Protection Act	03/02/2021	The VCDPA allows individuals to opt out of profiling used for decisions that have legal or significant effects on them. This provides consumers the right to protect their information from algorithmic profiling.
3	Colorado Protecting Consumers from Unfair Discrimination in Insurance Practices	07/06/2021	The law applies to insurers' use of external consumer data and information sources (ECDIS), as well as algorithms and predictive models that use ECDIS in "insurance practices," that "unfairly discriminate" based on race, color, national or ethnic origin, religion, sex, sexual orientation, disability, gender identity, or gender expression.
4	Indiana Consumer Privacy Law	05/01/2023	The bill establishes rules for profiling and automated decision-making, granting individuals the right to opt-out of profiling that results in decisions with legal or similarly significant effects on the consumer.
5	Tennessee Information Protection Act	05/11/2023	The bill mandates data protection assessments for profiling activities that pose a foreseeable risk of unfair or deceptive treatment, unlawful disparate impact, financial, physical, or reputational harm, or invasion of privacy that would be offensive to a reasonable person, ensuring that profiling is carefully evaluated when it may lead to substantial harm.
6	Montana the Consumer Data Privacy Act	05/19/2023	The law requires data controllers to limit the collection of personal data to what is necessary for the purposes disclosed to the consumer, implement adequate security

			measures to protect the data, and provide a mechanism for consumers to revoke their consent. Data processors must adhere to the controller's instructions and assist in fulfilling the controller's obligations.
7	Connecticut Privacy Act (CTPA)	07/01/2023	Controllers must also perform data risk assessments prior to processing consumer data when such processing presents a “heightened risk of harm.” These situations involve profiling that poses a foreseeable risk of unfair treatment, unlawful impact, financial or reputational harm, invasion of privacy, or other significant injury to consumers.
8	Colorado Privacy Act (CPA)	07/01/2023	Consumers have the right to opt-out of the processing of their personal data for purposes of profiling that results in legal or similarly significant effects. The CPA also mandates that data controllers conduct a Data Protection Impact Assessment (DPIA) if the processing of personal data creates a heightened risk of harm to consumers.
9	Oregon Consumer Data Privacy Act	07/18/2023	The bill follows the Virginia Consumer Data Protection Act and establishes rules for profiling and automated decision-making. It specifically allows individuals to opt out of processing that involves profiling for decisions with legal effects or similarly significant consequences.
10	Delaware Personal Data Privacy Act	09/11/2023	Controllers are required to conduct data protection assessments when data processing poses a “heightened risk of harm,” such as when profiling may result in unfair treatment, financial or reputational harm, privacy invasions, or other substantial injury to consumers.
11	New Jersey Data Protection Act	01/15/2024	Consumers are granted rights to access, correct, delete, and transfer their data, as well as opt-out of certain processing activities. The definition of sensitive data includes financial information. Controllers must provide consumers with a universal opt-out mechanism for targeted advertising within six months of the law’s enactment. Additionally, special opt-in consent is

			required for processing personal data of children between the ages of 13 and 17.
12	AB-1836 California Use of likeness: digital replica.	02/16/2024	Prohibits commercial use of digital replicas of deceased performers in films, TV shows, video games, audiobooks, sound recordings, etc., without first obtaining the consent of those performers' estates.
13	New Hampshire Consumer Data Privacy Bill	03/06/2024	The law requires the Secretary of State to establish secure and reliable methods for consumers to exercise their privacy rights and set standards for privacy notices. It also specifies that personal information maintained for compliance with the federal Controlled Substances Act (21 U.S.C. section 830) and information included in a limited data set as outlined in 45 C.F.R. 164.514(e) are subject to specific usage, disclosure, and maintenance requirements as defined in that regulation.
14	Colorado Artificial intelligence consumer protection bill	05/17/2024	Developers are presumed to have exercised reasonable care if they make relevant information and documentation available to deployers for completing impact assessments, publish a statement outlining the types of high-risk systems developed and how discrimination risks are managed, and disclose foreseeable risks of discrimination to the Attorney General (AG) and deployers within 90 days of discovery.
15	Minnesota Consumer Data Privacy Law	05/19/2024	The Act grants consumers the unique right to question profiling, request profiling results, and challenge inaccurate information. Controllers are required to provide a conspicuous opt-out link if they sell personal data, process it for targeted advertising, or engage in profiling, offering a way for consumers to opt-out outside of the privacy notice.
16	Colorado Candidate Election Deepfake Disclosures	05/24/2024	The act regulates the use of deepfakes created with generative artificial intelligence in political communications about candidates for elective office. It prohibits the distribution of communications containing undisclosed or

			improperly disclosed deepfakes, with knowledge or reckless disregard for their deceptive nature.
17	Texas Date Privacy and Security Act	07/01/2024	Create similar requirements enabling individuals to opt-out of “profiling” that produces a legal or similarly significant effect concerning the individual. Controllers must also perform a data protection assessment for high-risk profiling activities.
18	AB-2839 California Elections: deceptive media in advertisements	09/17/2024	The bill prohibits any individual, committee, or other entity from knowingly distributing advertisements or other election communications containing materially deceptive content within 120 days before an election in California (or, in specified cases, 60 days after an election), subject to specified exemptions.
19	AB-2602 California Contracts against public policy: personal or professional services: digital replicas	09/17/2024	A provision in an agreement between an individual and any other person for the performance of personal or professional services is unenforceable only as it relates to a new performance, by a digital replica of the individual of the voice or likeness of an individual in lieu of the work of the individual.
20	AB-2355 California Political Reform Act of 1974: political advertisements: artificial intelligence	09/17/2024	Electoral advertisements that use AI-generated or significantly altered content must include a disclosure stating that the material has been altered.
21	SB-942 California AI Transparency Act	09/19/2024	The law applies to businesses providing a generative AI system with over 1 million monthly visitors within a 12-month period and that is publicly accessible within the state’s geographic boundaries. These businesses are required to develop an AI detection tool that enables users to query which content was created by a generative AI system.
22	AB-2013 California	09/28/2024	The law applies to AI developers, which is defined broadly to mean any person,

	Generative artificial intelligence: training data transparency		government agency, or entity that either develops an AI system or service or "substantially modifies it". The law aims to ensure that Californians have access to clear documentation regarding the data driving AI systems, promoting transparency and accountability in AI development.

[Annex 3] The China Policy Documents Referenced and Discussed

No	Title	Issuing Authority	Date of Issue	Category
1	Action Outline for Promoting the Development of Big Data《促进大数据发展行动纲要》	The State Council of the People's Republic of China（国务院）	8/31/2015	administrative normative document
2	“Internet Plus” Artificial Intelligence Three-Year Action Plan《“互联网+”人工智能三年行动实施方案》	National Development and Reform Commission, Ministry of Science and Technology, Ministry of Industry and Information Technology, Cyberspace Administration of China（国家发展改革委、科技部、工业和信息化部、中央网信办）	5/18/2016	administrative normative document
3	13th Five-Year Plan for Developing National Strategic and Emerging Industries《“十三五”国家科技创新规划》	The State Council of the People's Republic of China（国务院）	7/28/2016/	administrative normative document
4	New Generation of Artificial Intelligence Development Plan《新一代人工智能发展规划》	The State Council of the People's Republic of China（国务院）	7/8/2017	administrative normative document
5	Three-Year Action Plan for Promoting the Development of a New Generation of Artificial Intelligence Industry（2018-2020）《促进新一代人工智能	Ministry of Industry and Information Technology（工业和信息化部）	12/13/2017	administrative normative document

	产业发展三年行动计划（2018-2020年）》			
6	AI Innovation Action Plan for Institutions of Higher Education《高等学校人工智能创新行动计划》	Ministry of Education（教育部）	4/2/2018	administrative normative document
7	Governance Principles of a New Generation of Artificial Intelligence: Developing Responsible AI《新一代人工智能治理原则——发展负责任的人工智能》	Ministry of Science and Technology（科技部）	6/17/2019	administrative normative document
8	the Work Guidelines for the Construction of National Open Innovation Platforms for the New Generation Artificial Intelligence《国家新一代人工智能开放创新平台建设工作指引》	Ministry of Science and Technology（科技部）	8/1/2019	administrative normative document
9	Guiding Opinions on Promoting the Development of Artificial Intelligence in Forestry and Grassland《关于促进林业和草原人工智能发展的指导意见》	National Forestry and Grassland Administration（国家林业和草原局）	11/8/2019	administrative normative document

10	Several Opinions on Promoting Interdisciplinary Integration and Accelerating Graduate Education in the Field of Artificial Intelligence at Universities Constructing 'Double First-Class'《关于“双一流”建设高校促进学科融合加快人工智能领域研究生培养的若干意见》	Ministry of Education, National Development and Reform Commission, Ministry of Finance（教育部、国家发展改革委、财政部）	1/21/2020	administrative normative document
11	Guidelines for Building New Generation AI Standard System《国家新一代人工智能标准体系建设指南》	Standardization Administration of China, Cyberspace Administration of China, National Development and Reform Commission, Ministry of Science and Technology, Ministry of Industry and Information Technology（国家标准化管理委员会、中央网信办、国家发展改革委、科技部、工业和信息化部）	7/27/2020	administrative normative document
12	Guidelines for the Construction of National New-Generation AI Innovation and Development Pilot Zone(Revision)《国家新一代人工智能创新发展试验区建设工作指引（修订版）》	Ministry of Science and Technology（科技部）	9/29/2020	administrative normative document

13	Guiding Opinions on Accelerating the Construction of a Coordinated Innovation System for the National Integrated Big Data Center《关于加快构建全国一体化大数据中心协同创新体系的指导意见》	National Development and Reform Commission, Cyberspace Administration of China, Ministry of Industry and Information Technology, National Energy Administration（国家发展改革委、中央网信办、工业和信息化部、国家能源局）	12/23/2020	administrative normative document
14	Cybersecurity Standards Practice Guide - Guidelines for Ethical and Security Risk Prevention in Artificial Intelligence《网络安全标准实践指南——人工智能伦理安全风险防范指引》	Secretariat of the National Information Security Standardization Technical Committee（全国信息安全标准化技术委员会秘书处）	1/5/2021	administrative normative document
15	The Outline of the 14th Five-Year Plan (2021-2025) for National Economic and Social Development and Vision 2035 of the People's Republic of China 中华人民共和国国民经济和社会发展第十四个五年规划和 2035 年远景目标纲要	National People's Congress（全国人大）	3/11/2021	Policy document
16	Implementation Plan for the Computing Power Hub of the National Integrated Big Data Center Coordinated	National Development and Reform Commission, Cyberspace Administration of China, Ministry of	5/24/2021	administrative normative document

	Innovation System《全国一体化大数据中心协同创新体系算力枢纽实施方案》	Industry and Information Technology, National Energy Administration（国家发展改革委、中央网信办、工业和信息化部、国家能源局）		
17	Personal Information Protection Law《个人信息保护法》	tanding Committee of the National People's Congress（全国人大常委会）	8/20/2021	law
18	Guiding Opinions on Strengthening Overall Governance of Internet Information Service Algorithms《关于加强互联网信息服务算法综合治理的指导意见》	Cyberspace Administration of China, Publicity Department of CPC Central Committee, Ministry of Education, Ministry of Science and Technology, Ministry of Industry and Information Technology, Ministry of Public Security, Ministry of Culture and Tourism, State Administration for Market Regulation, National Radio and Television Administration（国家互联网信息办公室、中共中央宣传部、教育部、科学技术部、工业和信息化部、公安部、文化和旅游部、国家市场监督管理总局、国家广播电视总局）	9/17/2021	administrative normative document
19	New Generation Artificial Intelligence Ethical Code《新一代人工智能伦理规范》	National New Generation Artificial Intelligence Governance Specialist Committee（国家新一代人工智能治理专	9/25/2021	administrative normative document

		业委员会）		
20	14th Five-Year Plan for the Development of the Big Data Industry《“十四五”大数据产业发展规划》	Ministry of Industry and Information Technology（工业和信息化部）	11/15/2021	administrative normative document
21	Provisions on the Administration of Algorithm-generated Recommendations for Internet Information Services《互联网信息服务算法推荐管理规定》	Cyberspace Administration of China, Ministry of Industry and Information Technology of, Ministry of Public Security, State Administration for Market Regulation（国家互联网信息办公室、中华人民共和国工业和信息化部、中华人民共和国公安部、国家市场监督管理总局）	3/1/2022	ministerial rule
22	Guiding Opinions on Accelerating Scenario Innovation and Promoting High-quality Economic Development with High-level Application of Artificial Intelligence《关于加快场景创新以人工智能高水平应用促进经济高质量发展的指导意见》	Ministry of Science and Technology, Ministry of Education, Ministry of Industry and Information Technology, Ministry of Transport, Ministry of Agriculture and Rural Affairs, National Health Commission（科技部、教育部、工业和信息化部、交通运输部、农业农村部、国家卫生健康委）	7/29/2022	administrative normative document
23	Notice on Supporting the Construction of a New Generation of Artificial	Ministry of Science and Technology（科技部）	8/12/2022	administrative normative document

	Intelligence Demonstration Application Scenarios《关于支持建设新一代人工智能示范应用场景的通知》			
24	Guiding Opinions on Promoting the Development of the Energy Electronics Industry《关于推动能源电子产业发展的指导意见》	Ministry of Industry and Information Technology, Ministry of Education, Ministry of Science and Technology, People's Bank of China, China Banking and Insurance Regulatory Commission, National Energy Administration（工业和信息化部、教育部、科学技术部、中国人民银行、中国银行保险监督管理委员会、国家能源局）	1/3/2023	administrative normative document
25	Provisions on the Administration of Deep Synthesis Internet Information Services《互联网信息服务深度合成管理规定》	Cyberspace Administration of China, Ministry of Industry and Information Technology, Ministry of Public Security（国家互联网信息办公室、工业和信息化部、公安部令）	1/10/2023	ministerial rule
26	Interim Measures for the Management of Generative Artificial Intelligence Services《生成式人工智能服务管理暂行办法》	Cyberspace Administration of China, National Development and Reform Commission, Ministry of Education, Ministry of Science and Technology, Ministry of Industry and Information Technology, Ministry	8/15/2023	ministerial rule

		of Public Security, National Radio and Television Administration（国家互联网信息办公室、国家发展改革委、教育部、科学技术部、工业和信息化部、公安部、国家广播电视总局）		
27	Action Plan for the High-Quality Development of Computing Power Infrastructure《算力基础设施高质量发展行动计划》	Ministry of Industry and Information Technology, Cyberspace Administration of China, Ministry of Education, National Health Commission, People’s Bank of China, State-owned Assets Supervision and Administration Commission of the State Council（工业和信息化部、中央网信办、教育部、国家卫生健康委员会、中国人民银行、国务院国有资产监督管理委员会）	10/8/2023	administrative normative document
28	Special Plan for the Development and Research of Artificial Intelligence in Earthquake Prevention and Disaster Reduction (2023-2035)《防震减灾领域人工智能发展研究专项规划(2023—2035 年)》	China Earthquake Administration（中国地震局）	10/13/2023	administrative normative document

29	Regulation on the Protection of Minors in Cyberspace《未成年人网络保护条例》	The State Council of the People's Republic of China（国务院）	10/16/2023	ministerial rule
30	Global AI Governance Initiative《全球人工智能治理倡议》	Cyberspace Administration of China（中央网信办）	10/18/2023	administrative document
31	Guiding Opinions on the Innovative Development of Humanoid Robots《人形机器人创新发展指导意见》	Ministry of Industry and Information Technology（工业和信息化部）	10/20/2023	administrative normative document
32	Guiding Opinions on Accelerating the Transformation and Upgrading of Traditional Manufacturing《关于加快传统制造业转型升级的指导意见》	Ministry of Industry and Information Technology, National Development and Reform Commission, Ministry of Education, Ministry of Finance, People's Bank of China, State Administration of Taxation, National Financial Regulatory Administration, China Securities Regulatory Commission（工业和信息化部、国家发展改革委、教育部、财政部、中国人民银行、税务总局、金融监管总局、中国证监会）	12/1/2023	administrative normative document
33	Announcement on the Publication of the Catalogue of Technologies Prohibited or Restricted for Export in China 关于公布《中国禁止出口限	Ministry of Commerce, Ministry of Science and Technology(商务部、科技部)	2023/12/21	administrative normative document

	制出口技术目录》的公告			
34	Implementation Opinions from the National Development and Reform Commission and Other Departments on Deepening the 'Eastern Data, Western Computing' Project and Accelerating the Construction of a National Integrated Computing Power Network《国家发展改革委等部门关于深入实施"东数西算"工程加快构建全国一体化算力网的实施意见》	National Development and Reform Commission, National Data Bureau, Cyberspace Administration of China, Ministry of Industry and Information Technology, National Energy Administration（国家发展和改革委员会、国家数据局、中央网信办、工业和信息化部、国家能源局）	12/25/2023	administrative normative document
35	Implementation Opinions on Promoting the Innovative Development of Future Industries《关于推动未来产业创新发展的实施意见》	Ministry of Industry and Information Technology, Ministry of Education, Ministry of Science and Technology, Ministry of Transport, Ministry of Culture and Tourism, State-owned Assets Supervision and Administration Commission of the State Council, Chinese Academy of Sciences（工业和信息化部、教育部、科学技术部、交通运输部、文化和旅游部、国务院国有资产监督管理委员会、中国科学院）	1/8/2024	administrative normative document

36	Action Plan for the Construction of Informationization Standards (2024-2027)《信息化标准建设行动计划（2024-2027 年）》	Cyberspace Administration of China, Ministry of Industry and Information Technology, State Administration for Market Regulation（中央网信办、工业和信息化部、国家市场监督管理总局）	5/29/2024	administrative normative document
37	Guidelines for the Construction of the National Comprehensive Standardization System for the Artificial Intelligence Industry (2024 Edition)国家人工智能产业综合标准化体系建设指南（2024 版）	Ministry of Industry and Information Technology, Cyberspace Administration of China, National Development and Reform Commission, Standardization Administration of China（工业和信息化部、中央网信办、国家发展和改革委员会、国家标准化管理委员会）	6/5/2024	administrative normative document
38	AI Safety Governance Framework (V1.0)《人工智能安全治理框架》1.0 版	National Cybersecurity Standardization Technical Committee（全国网络安全标准化技术委员会）	9/9/2024	administrative normative document

Hallucinations in Legal Practice: A Comparative Case Law Analysis

Dr. Bakht Munir[*]

Abstract: The employment of Artificial Intelligence (AI) in legal operations raised concerns about ethical challenges and their potential consequences. Among other issues, hallucinations refer to a phenomenon whereby AI systems generate plausible but inaccurate or fabricated responses. In legal matters, where precision and compliance with authorities are paramount, inconsistency with legal doctrines and judicial precedents may lead to wrong legal advice or decisions. AI tools such as ChatGPT and Lexis +AI exhibit human-like intelligence. Still, their fabricated responses could lead to real-world consequences such as professional misconduct resulting in civil liabilities. This article contributes to the following aspects: it compares judicial scholarship evolved on AI hallucinations in the USA, Pakistan, UK, Australia, and Canada. It examines the standing orders and policy guidelines set by the bar and bench constituting patchwork with competing outcomes. The article emphasizes uniform and comprehensive policy guidelines for the responsible use of generative AI tools in legal operations.

Keywords: Cases of Hallucination; Standing Orders on Hallucinations; Generative AI; AI and Malpractices; AI and Civil Liabilities

[*] University of Kansas School of Law, US.

Table of Contents

INTRODUCTION

The recent integration of AI into legal operations offers unparalleled opportunities and poses critical challenges.[1] Though AI models are instrumental in performing legal tasks, their adoption is hampered by crucial concerns such as producing incorrect or deceptive outcomes, commonly known as hallucinations.[2] Modern AI solutions are transforming the legal fields, including legal education, research, and practice.[3] Within a few months of its public release in November 2022, ChatGPT secured itself as the fastest-ever growing consumer application in human history.[4] Embracing the trend, recent studies have found that generative AI witnessed remarkable performance in law school exams, Bar exams, and other legal analyses.[5] AI enables machines to mimic human intelligence, empowering them to learn, solve problems, and make decisions. The employment of AI in various spheres is driving transformative changes and has the potential to revolutionize legal operations. Lawyers are utilizing AI in legal operations to augment legal services. 41 of the top 100 US law firms have initiated AI in their legal services.[6] According to a study by LexisNexis, 80% of Fortune 1000 executives desire their external counsels to enhance efficiency by leveraging AI capabilities. However, these tools are not risk-free and constitute ethical challenges such as bias, copyright, data invasion, fabricated responses, and information security, posing ultimate liability to corroborate their outcomes.[7]

[1] Darla Wynon Kite-Jackson, *2023 Artificial Intelligence (AI) TechReport*, AM. BAR ASS'N (Jan 15, 2024).

[2] Matthew Dahl et al., *Large Legal Fictions: Profiling Legal Hallucinations in Large Language Models*, 16 J. LEGAL ANALYSIS 64, (2024), https://academic.oup.com/jla/article/16/1/64/7699227.

[3] Jonathan H. Choi and Daniel Schwarcz, 2024. *AI Assistance in Legal Analysis: An Empirical Study*. J. LEGAL EDUC. doi: 10.2139/ssrn.4539836. (forthcoming), https://elsevier-ssrn-document-store-prod.s3.amazonaws.com/23/08/13/ssrn_id4539836_code499486.pdf; *See also*, Michael A. Livermore, Felix Herron, & Daniel Rockmore, *Language Model Interpretability and Empirical Legal Studies*. J. INSTITUT. THEORETI. ECON., forthcoming (2024), https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4599212; *See also*, Ian Rodgers, John Armour, & Mari Sako, *How Technology Is (or Is Not) Transforming Law Firms, 19* ANN. R. LAW SOCIAL SCI. *299–317* (2023), https://www.annualreviews.org/content/journals/10.1146/annurev-lawsocsci-111522-074716.

[4] See Krystal Hu, *ChatGPT Sets Record for Fastest-Growing User Base*, REUTERS (Feb. 2, 2023), https://www.reuters.com/technology/chatgpt-sets-record-fastest-growing-user-base-analyst-note-2023-02-01/.

[5] Jonathan H. Choi, Kristin E. Hickman, Amy B. Monahan, & Daniel Schwarcz, *ChatGPT Goes to Law School*, 71 J. LEGAL ED. 387 (2022), https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4335905; *See also,* Chung Kwan, *What Is the Impact of ChatGPT on Education? A Rapid Review of the Literature*, 13 EDUC. SCI. 410 (2023); John Ney et al., *Large Language Models as Tax Attorneys: A Case Study in Legal Capabilities Emergence, Philosophical Transactions A 382(2270),* THE ROYAL SOCIETY, (Feb. 26, 2024), https://www.researchgate.net/publication/378489936_Large_language_models_as_tax_attorneys_a_case_study_in_legal_capabilities_emergence.

[6] Justin Henry, *We Asked Every Am Law 100 Law Firm How They're Using Gen AI. Here's What We Learned*, AM. LAW. (Jan. 29, 2024), https://www.law.com/americanlawyer/2024/01/29/we-asked-every-am-law-100-firm-how-theyre-using-gen-ai-heres-what-we-learned/?slreturn=20241013185149.

[7] Joseph J. Avery, Patricia Sánchez Abril & Alissa del Riego, *ChatGPT, Esq.: Recasting Unauthorized Practice of Law in the Era of Generative AI*, 26 YALE J. L. & TECH. 64 (2023), https://yjolt.org/sites/default/files/avery_abril_delriego_26yalejltech64.pdf; *see also, Amy B. Cyphert, A Human Being Wrote This Law Review Article: GPT-3 and the Practice of Law, 55 UC DAVIS L. REV. 401 (2021), https://lawreview.law.ucdavis.edu/sites/g/files/dgvnsk15026/files/media/documents/55-1_Cyphert.pdf; Ed Walters, The Model Rules of Autonomous Conduct: Ethical Responsibilities of Lawyers and Artificial Intelligence, 35 Ga. St. U. L. Rev. 1073 (2019), https://readingroom.law.gsu.edu/cgi/viewcontent.cgi?article=2974&context=gsulr; Nicole Yamane,*

Hallucination refers to false or deceptive outputs that AI models perpetuate for various reasons such as insufficient data training, incorrect assumptions, or biases in the dataset. AI models are trained on data and learn to make predictions by finding patterns in the data. The precision of outcomes is often subject to the quality and completeness of the training data. Where the training data is incomplete, biased, or otherwise flawed, the AI models may learn incorrect patterns, leading to inaccurate predictions or plausible fabricating links to webpages that never existed. While considering the efficiencies of AI solutions, new ethical challenges have been posed.[8]

A. Generative AI and its Tendency Towards Hallucinations

Given its functions, generative AI is a particular kind of AI that focuses on producing original content in response to users' questions. Generative AI is based on machine learning models, also known as deep learning models, which are algorithms that mimic the human brain's learning and decision-making process. These models learn patterns and structures from the training data and utilize them to comprehend users' natural language prompts and respond with new relevant content. The use of Generative AI became more active with the development of Large Language Models (LLMs), which can generate human-like text based on the features learned from the huge data on which these models are trained. By predicting the next element in a sequence, these models produce new content and host inherent challenges such as perpetuating misinformation.[9] Generative AI may produce erroneous output based on its probabilistic algorithms for making inferences. These models perpetuate the most probable response to a user's prompt without guaranteeing correctness, which may lead to a plausible but fabricated outcome.[10]

LLMs are advanced AI systems that fall under Natural Language Processing (NLP) and are designed to comprehend and produce human language. These models are trained on huge data to learn the intricacies of language by employing transformer architectures, which have revolutionized NLP and other AI tasks since their inception in 2017.[11] These models excel in tasks such as summarizing text, answering questions, and engaging in conversations by generating relevant and coherent text based on their input. For instance, ChatGPT-4 is an LLM developed by OpenAI. However, other generative AI tools such as Microsoft Copilot, Lexis +AI, and Westlaw Co-Counsel leverage the capabilities of LLMs to perform multiple tasks but are not LLMs themselves.

Artificial Intelligence in the Legal Field and the Indispensable Human Element Legal Ethics Demands, 33 GEO. J. LEGAL ETHICS 877 (2020), https://www.law.georgetown.edu/legal-ethics-journal/wp-content/uploads/sites/24/2020/09/GT-GJLE200038.pdf.

[8] Frances Green & Rebecca Porter, *The Legal Vision for the Future or an AI Hallucination? Navigating the Complexities of Attorney Ethics and Use of Artificial Intelligence*, NEW YORK L. J., (April 2, 2024), https://www.law.com/newyorklawjournal/2024/04/02/the-legal-vision-for-the-future-or-an-ai-hallucination-navigating-the-complexities-of-attorney-ethics-and-use-of-artificial-intelligence/?slreturn=20241010143515.

[9] IBM, Generative AI, (last visited Dec. 16, 2024), https://www.ibm.com/topics/generative-ai.

[10] Stefan Feuerriegel, Jochen Hartmann, Christian Janiesch & Patrick Zschech, Generative AI, 66 BUS. & INFO. SYS. ENG'G 111 (2024), https://link.springer.com/article/10.1007/s12599-023-00834-7.

[11] Ashish Vaswani et al., *Attention is All You Need*, 30 Advances in Neural Info. Processing Sys. 5998 (2017), https://scholar.google.com/scholar?hl=en&as_sdt=0%2C5&q=Attention+is+All+You+Need%22+by+Vaswani+et+al.+in+2017%2C&btnG=.

B. Modes of Legal Hallucinations

Undoubtedly, AI systems have the potential to address complex legal tasks but are limited by a notable issue: their tendency to produce incorrect or misleading outcomes.[12] Legal hallucinations can be referred to as the phenomenon where LLMs perpetuate fabricated legal responses, which could be problematic in the legal context where accuracy is paramount. Legal hallucinations are exhibited in many ways such as inventing fictitious precedents, nonexistent statutes, misinterpreting laws, offering inaccurate legal advice, and producing made-up legal content, which hosts various risks, including legal liability, malpractice, and miscarriage of justice.

Legal professionals are increasingly getting involved with AI chatbots without fully realizing how they work and their susceptibility to errors. Even if legal professionals are unwilling to deploy AI, they still need to learn and live with them. Legal professionals are expected to act as custodians of the legal system and should be capable of identifying errors in the outcomes of these models.[13]

Hallucinations occur when AI systems produce incorrect, misleading, or entirely fabricated content: Incorrect predictions, to predict the happening of an unlikely event such as the rain forecast when it does not rain. False positive, to identify something as a threat when it is not such as detecting a fraudulent activity when it is not. False negative, fails to identify something as a threat when it is a threat such as failing to identify a cancerous tumor. Hallucinations could be in any of the following forms: (1) Factual hallucinations, AI systems might produce information factually incorrect or nonexistent such as discovering scientific facts or historical events that are not true.[14] (2) Contextual hallucinations, where AI models misunderstand the context or misinterpret the user's intent. It comes to the fore where AI responses are contextually irrelevant or inappropriate to the given prompt.[15] (3) Logical Hallucinations, where AI responses are logically inconsistent or contradictory. For example, where AI-generated content lacks a coherent line of reasoning. (4) Visual hallucinations, where AI systems generate images containing elements other than input data or distorted unrealistically.[16] (5) Conversational hallucinations, where the AI system fabricates part of a conversation like contributing statements to the people or inventing quotes who never made them.

Like other fields, the recent adoption of LLMs into legal operations offers significant opportunities and considerable challenges.[17]

[12] Matthew Dahl, et al., *Large Legal Fictions: Profiling Legal Hallucinations in Large Language Models,* arXiv:2401.01301v2 [cs.CL], (Jun 21, 2024), https://arxiv.org/abs/2401.01301.

[13] David Rubenstein, *2024 Selected Topics and Miscellany CLE,* Washburn University School of Law, Presentation *(June 13, 2024), https://www.washburnlaw.edu/about/community/cle/_files/selected-topics-schedule.pdf.*

[14] Ankit, *What is AI Hallucination? Understanding and Mitigating AI Hallucination*, GeeksforGeeks (Jan. 27, 2025), https://www.geeksforgeeks.org/what-is-ai-hallucination/

[15] MIT Sloan Educational Technology Office, *When AI Gets It Wrong: Addressing AI Hallucinations and Bias*, https://mitsloanedtech.mit.edu/ai/basics/addressing-ai-hallucinations-and-bias.

[16] IBM, *What are AI Hallucinations?,* https://www.ibm.com/think/topics/ai-hallucinations.

[17] Darla Wynon Kite-Jackson, *2023 Artificial Intelligence (AI) TechReport*, ABA TECHREPORT 2023, (Jan. 15, 2024), https://www.americanbar.org/groups/law_practice/resources/tech-report/2023/2023-artificial-intelligence-ai-techreport/.

I. WHY DO AI MODELS HALLUCINATE?

With the widespread proliferation of AI systems, some critical challenges such as hallucinations have been confronted. Hallucinations result when AI models generate content that is not grounded or realistic. Consequently, AI models might fabricate responses that do not correspond to real-world data, potentially leading to dire consequences. Conventionally, AI hallucinations transpire the way AI models are trained. Most LLMs depend on the data available on the internet, which might contain both correct and incorrect content supplemented with inherent cultural and societal biases. The models mimic patterns from that data without recognizing their truthfulness and can perpetuate imprecision or biases.[18]

From the above conception, intriguing questions arise: Why do we expect AI to be 100% unbiased when humans themselves are not? Why is the burden of absolute accuracy placed on AI programs? It is worth considering why we hold AI to such high standards when, in human-to-human interactions, achieving complete impartiality and accuracy is impossible.

The LLMs are subject to limitations and work like advanced autocomplete tools – designed to foresee the next sequence or word based on the observed patterns – with the underlying objective of creating credible content and not verifying its truthiness. Inversely, any accuracy in their generated content is often inadvertent and might produce output that looks plausible but could be erroneous.[19]

As LLMs by design cannot distinguish between true and false even if these models are trained exclusively on accurate data, there is still a probability of producing new, potentially erroneous content by assimilation of patterns in an unexpected manner.[20] These LLMs are not infallible, and their most puzzling behavior is the production of hallucinations, either incorrect responses or entirely fabricated results that could create real-world challenges where accuracy is paramount. Considering these algorithms are not sentient and cannot distinguish between truth and lies, it is imperative to comprehend the nature of these hallucinations to harness the effectiveness and efficiency of these models responsibly. Though these models might appear sentient because they generate coherent and relevant text, it is notable that they cannot differentiate between truth and false, nor have intentions or beliefs. So, hallucinations are the byproduct of the models' probabilistic nature and limitations in the training data, rather than a deliberate act.[21]

Likewise, a deliberate act of the designer can cause the models to perpetuate inaccurate responses. For instance, data poisoning is an intentional cyberattack, which can degrade the model's performance or cause it to produce incorrect or biased

[18] Karen Weise & Cade Metz, *When A.I. Chatbots Hallucinate*, THE NEW YORK TIMES, (May 1, 2023), https://www.nytimes.com/2023/05/01/business/ai-chatbots-hallucination.html.

[19] Matt O'Brien, *Chatbots Sometimes Make Things Up. Is AI's Hallucination Problem Fixable?*, AP NEWS, (August 1, 2023), https://apnews.com/article/artificial-intelligence-hallucination-chatbots-chatgpt-falsehoods-ac4672c5b06e6f91050aa46ee731bcf4.

[20] *When AI Gets It Wrong: Addressing AI Hallucinations and Bias*, MIT SLOAN TEACHING & LEARNING TECHNOLOGIES, https://mitsloanedtech.mit.edu/ai/basics/addressing-ai-hallucinations-and-bias.

[21] Daniel A. Tysver, *AI Hallucinations (Why would I lie?)*, BITLAW, https://www.bitlaw.com/ai/hallucinations-and-AI.html.

outcomes. Data poisoning can occur by modifying existing data by injecting misleading information into the training dataset or deleting a portion to skew results.[22] These attacks aim to manipulate specific outcomes or to degrade the overall robustness of the model's performance.[23] To overcome the issue of data poisoning, it is critical to maintain the quality and integrity of the training data and employ robust security measures.

II. LEGAL HALLUCINATIONS: A COMPARATIVE CASE LAW STUDY

The following segment provides a detailed analysis of the judicial scholarship that evolved on legal hallucinations and its potential impacts:

A. Case Law Development in the USA

New York attorneys faced legal consequences for presenting a brief with fictitious case law precedents generated through ChatGPT.[24] Two lawyers were each sanctioned to pay a $ 5,000 fine for providing a legal brief that referred to six fictitious case citations produced by ChatGPT, which the court regarded to have acted in bad faith by declaring it as an act of conscious avoidance and false and misleading statements to the court. The lawyers used ChatGPT to prepare a personal injury case against Columbian airline Avianca and included references of false citations. The court dismissed the claim on the pretext of statutory limitation. While imposing the sanction, the court declared that using AI is not inherently improper, but the ethics rule requires the attorneys to ensure accuracy in their filings. The lawyers kept standing by their fake opinions despite the court and the airline having questioned the existence of the citations.[25]

Shortly after the New York case, *Ex parte Lee,* another fabricated case, was reported in a Texas appellate court.[26] Allen Michael Lee was charged with three sexual assaults for which the bail was set at $ 400,000, which Lee contested by filing a pre-trial application for the writ of habeas corpus for either his release or reduction of bail to $ 15,000, which the trial court refused. Hence, he challenged the court order at the Court of Appeals of Texas. The court denied review based on the deficient briefing, citing five cases, three of which did not exist in the Southwest Reporter. The court realized the cited cases did not exist and the two others were from the Missouri court, making them immaterial to the argument in the brief. Lee, however, did not address those issues through a reply or a supplemented brief. The court called the brief illogical

[22] Bart Lenaerts-Bergmans, *Data Poisoning: The Exploitation of Generative AI*, CROWDSTRIKE, Mar. 20, 2024, https://www.crowdstrike.com/en-us/cybersecurity-101/cyberattacks/data-poisoning/.

[23] Tom Krantz, *What is Data Poisoning?*, IBM, https://www.ibm.com/think/topics/data-poisoning.

[24] Benjamin Weiser, *Here's What Happens When Your Lawyer Uses ChatGPT*, THE NEW YORK TIMES, (May 27, 2023), https://www.nytimes.com/2023/05/27/nyregion/avianca-airline-lawsuit-chatgpt.html. *See for details*, Mata v. Avianca, Inc., No. 1:2022cv01461, Document 55 (S.D.N.Y. 2023).

[25] Hon. John G. Browning, *Robot Lawyers Don't Have Disciplinary Hearings—Real Lawyers Do: The Ethical Risks and Responses in Using Generative Artificial Intelligence*, 40 GA. ST. U. L. REV. 917, 925(2024), https://readingroom.law.gsu.edu/gsulr/vol40/iss4/9/; See also, Sara Merken, *New York lawyers sanctioned for using fake ChatGPT cases in legal brief*, Reuters (June, 26, 2023), https://www.reuters.com/legal/new-york-lawyers-sanctioned-using-fake-chatgpt-cases-legal-brief-2023-06-22/.

[26] Ex parte Lee, 673 S.W.3d 755, 756 (Tex. App. 2023).

and at least partly prepared with the help of AI.[27] Unlike the New York case, the court in the instant case did not issue a show cause order and report the authority for disciplinary action to the State Bar of Texas because it had addressed the issue raised in the appeal.

In *People v. Crabill*[28], Zachariah Crabill, a young attorney, filed a brief supported by dozens of cases prepared with the assistance of ChatGPT. On the hearing day, he realized the cases he submitted were not available on the LexisNexis and were 'garbage'. He compounded his mistake by not validating the citations or alerting the court and withdrawing the motion, Crabill blamed an intern when the court pointed out the made-up citations. While rejecting the motion, the court referred him to disciplinary action. After six days, Crabill filed an affidavit confessing the use of ChatGPT while drafting the motion. For his professional misconduct, he was terminated from his law firm and banned for one year and one day from practicing law.[29]

In April 2023, Lydia Nicholsen, a Los Angeles housing attorney, realized that the brief in an eviction case received from the opposing counsel, Dennis Block, was supported by fabricated citations. Nicholsen filed a motion and pointed out the fake cases. On confirmation, the judge declared the filings "rife with inaccurate and false statements" and imposed a fine of $ 999 on the firm, which was just under the threshold required for reporting to the state bar for further investigation.[30] The firm shifted liability onto a first-year lawyer, who had since left the firm, by blaming him for relying on an online search.[31]

In *United States v. Michel Cohen*[32], the defense attorney used AI while filing a motion for early release. Cohen, a former attorney for President Donald Trump, confessed to hush money to two women during the presidential campaign. Since November 2021, Cohen has been on supervised release after serving time in prison. His lawyer, Schwartz, filed a motion for his early release. Afterward, another attorney, Danya Perry, was added to the Choen's legal team who realized fabricated citations and alerted the court accordingly. The court issued a show cause notice to Schwartz to provide copies of the three cited cases or respond to why he should not be sanctioned. Based on attorney-client privilege, Schwartz requested to file a response under seal, which was unsealed on December 29, 2023. It was disclosed through a sworn declaration of Cohen that the citations were produced by Google Bard, which Schwartz incorporated with his submission without verification. Cohen, who was disbarred from

[27] Ibid.

[28] People v. Crabill, No. 23PDJ067, (Colo. O.P.D.J. Nov. 22, 2023).

[29] Hon. John G. Browning, *Robot Lawyers Don't Have Disciplinary Hearings—Real Lawy Do: The Ethical Risks and Responses in Using Generative Artificial Intelligence*, 40, GA. ST. U. L. REV., 917, 927 (2024), https://readingroom.law.gsu.edu/cgi/viewcontent.cgi?article=3275&context=gsulr.

[30] Pranshu Verma & Will Oremus, *These Lawyers Used ChatGPT to Save Time. They Got Fired and Fined.*, WASH. POST, https://www.washingtonpost.com/technology/2023/11/16/chatgpt-lawyer-fired-ai/ [https://perma.cc/TCU3-QLAW] (Nov. 16, 2023, 10:39 AM); see also, *Block v. Bramzon, No. B292129 (Cal. Ct. App. Jan. 22, 2021).*

[31] John G. Browning, *Robot Lawyers Don't Have Disciplinary Hearings—Real Lawyers Do: The Ethical Risks and Responses in Using Generative Artificial Intelligence*, 40 GA. ST. U. L. REV. 917, 928 (2023).

[32] United States v. Cohen, No. 18-cr-602 (S.D.N.Y. 2019).

practice years ago, admitted that he had not kept up with the trends in legal technology of these tools to produce citations that appeared real but were fake.[33]

In *United States v. Pras Michel*,[34] the defendant, a former Fugees rapper, was convicted on multiple charges including conspiracy and funds to influence US politics. The respondent filed a motion for a fresh trial on the pretext that his former attorney had spoiled the defense by employing AI to draft closing arguments. Part of the defense argument for a new trial was based on the ineffective assistance of the prior counsel due to his financial stake in the AI company whose tools he deployed in closing arguments. Michel's new lawyer asserted that the AI tools generated frivolous arguments, damaging the defense because these arguments were deficient and prejudiced against the defense. The court concluded that the error did not prejudice the result of the case, hence the conviction was upheld.[35] This case raised significant ethical questions: Was the client informed of and to what extent did he agree to the employment of generative AI? What are the obligations to notify the judge of using generative AI? The case constitutes a warning to the attorneys that improper employment of generative AI may result in a breach of care, leading to a legal malpractice claim or lawsuit.

In another case,[36] the attorney submitted AI-generated response papers that contained fictional and flawed citations. While underscoring the significance of accuracy in legal documents, the court underlined the risks associated with AI-produced content without proper verification. Consequently, the court denied the motion for summary judgment, permitting the case to continue to factual disputes.[37]

In a recent case[38], plaintiff Iovino sued her former employer, Michael Stapleton Associates (MSA) for alleged whistleblower retaliation under federal law. The plaintiff claimed she was fired for reporting the defendant's contract with the US Department of State. The MSA counterargued that the petitioner had shared confidential information with the media and violated a non-disclosure agreement. The court addressed the plaintiff's objections to the protective order granted in favor of the MSA, which restricted certain discovery requests. The court overruled the plaintiff's objections, affirmed a protective order, and the plaintiff's attorneys were given a show-cause notice for presenting fictitious cases and made-up quotations.[39]

The Chief Justice of the US Supreme Court, John Roberts, in the annual judicial report of 2023, regarded hallucinations as a substantial impediment to AI integration in legal operations. Legal determinations often navigate gray areas where the application of human judgment is essential, so key actors in court cannot be fully replaced with

[33] Andrew Zhang, *Michael Cohen's lawyer in hot water after citing court cases that don't exist*, POLITICO, (Dec. 12, 2023), https://www.politico.com/news/2023/12/12/michael-cohen-court-cases-00131435.

[34] United States v. Michel, No. 19-cr-148 (D.D.C. 2023), https://www.courtlistener.com/docket/15511282/united-states-v-michel/.

[35] *Ibid.*

[36] In re Estate of Samuel, No. 2016-2501/A&B, 2024 N.Y. Slip Op. 24014 (Sur. Ct. Jan. 11, 2024), https://caselaw.findlaw.com/court/sur-s-crt-new-yor-kin-cou/115735333.html.

[37] *Ibid.*

[38] Iovino v. Michael Stapleton Associates, LTD., No. 5:2021cv00064 - Document 177 (W.D. Va. 2024).

[39] *Ibid.*

machines.[40] Though the US courts are sufficiently sensitized about legal hallucinations and their potential impacts, considering AI integration in legal operations. The courts urged lawyers to counter-verify AI-assisted filings. Still, they barely spoke about how and to what extent the fictitious authorities could harm the reputation of the judges and courts.

B. Case Law Development in Pakistan

In Pakistan, the integration of AI in legal operations is in its infancy. Interestingly, a judge in a recent case used ChatGPT-4 while adjudicating a civil lawsuit.[41] In his judgment, the judge explained how AI is transforming the future of legal adjudications. The court queried the LLM, ChatGPT, *"What are the principles for granting an injunction in a civil case in Pakistan?"* and compared the generated principles, which corresponded to the civil law (irreparable loss, balance of convenience, and prima facie case).[42] Nevertheless, the LLM produced three extra conditions for granting an injunction (good faith, public interest, and equitable consideration). These excessive conditions are not stipulated in the statutes for granting injunctions and may amount to hallucinations. The judge seems oblivious to the legal hallucinations and considers the additional conditions within the purview of statutory laws and the byproduct of the judicial precedents that evolved over the years.[43] In the instant case, the court overlooked statutory requirements where precision was paramount at the cost of securing the infallible character of the LLM. Therefore, the court declared the AI-generated results different in form but identical in substance, ignoring their inherent character of confabulation.[44]

In another case,[45] while granting pre-arrest bail to a juvenile, the same judge employed ChatGPT-4 to demonstrate how AI-powered solutions can help adjudication. As provided in the Order, the judge reported 18 responses[46] assigned to the GPT-4, which provided an interesting phenomenon for conceptualizing legal hallucinations: the conversation with the chatbot started with "Whether in Pakistan, a juvenile of 13 years is entitled to post-arrest bail?". In response to question No. 2, "Discuss it concerning section 83 of Pakistan penal code." the chatbot provided outdated information, children under 12 years are considered incapable of committing crimes. It failed to produce post-amended details, which the judge identified in question no. 3 that the age of sufficient understanding is now amended as 16 years.[47] In response to question no. 4, "In the above situation, if the offense is an attempt to commit rape, then what do you suggest? Option for bail request." the GPT summarized that "if the offense is an attempt to commit rape, the juvenile would not be entitled to bail as a matter of

[40] Chief Justice John G. Roberts, Jr*., 2023 Year-End Report on the Federal Judiciary,*6, (Dec. 31, 2023), https://www.uscourts.gov/news/2023/12/31/chief-justice-roberts-issues-2023-year-end-report.

[41] Muhammad Iqbal v. Zayad, (2023), CA 11 of 2023.

[42] *For details see,* Order 39, Section 94 (c) and (e) of the Code of Civil Procedure 1908 & Section 37(1) of the Specific Relief Act 1963.

[43] see (2014) PLD Sindh 268 (pak.); see also (2011) CLC 1866 (pak.).

[44] Bakht Munir, *Artificial Intelligence and Legal Decision-Making in the USA and Pakistan: A Critical Appreciation of Regulatory Frameworks* (Oct. 25, 2024), https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4999590

[45] The State through Sameera Zulfiqar v. AM (a juvenile), FIR No.15/2023, dated 24.01.2023, Offence u/s 376(iii)/511PPC.

[46] For details, see ibid., pp. 6-15.

[47] *Ibid*., p. 7.

right under the Juvenile Justice System Act 2018". The judge highlighted the next question "Section 6 of the Juvenile Justice System Act 2018 deals with bail".[48]

While responding to the task in question no. 6 to cite case laws where bail is granted in the same subject matter, the GPT sought pardon for not having access to case laws and databases.[49] Likewise, the chatbot refused to provide legal advice sought in question no. 11 and preferred to provide general information when the judge asked whether to grant bail in such an eventuality. The chatbot replied, "As an AI language model, I cannot provide legal advice, but I can provide some general information."[50] While responding to question no. 12, the GPT quoted the wrong provision, "497 of the Code of Criminal Procedure (Cr.PC)" for granting pre-arrest bail, supplemented with the advice to consult some qualified legal professional or refer to the updated laws and precedents. The judge corrected the relevant provision in question no. 13, "Please note that pre-arrest bail is considered under section 498 of Cr.PC."[51]

In response to question no. 15 regarding its inability to provide precedents on the matter under discussion, the GPT responded that it did not have real-time access to the Internet, was incapable of browsing case laws to interpret or analyze cases, and could not provide legal advice.[52] In question no. 16, the chatbot was assigned to cite some research articles on juvenile pre-arrest bail in rape or other cases in Pakistan's context. The GPT responded to the inability to directly quote or provide references to specific research articles due to not having access to external databases or internet browsing capability. The GPT suggested legal research databases such as Westlaw, LexisNexis, Pakistan Legal Research Database, JSTOR, and legal experts for assistance in providing relevant research articles and precedents.[53]

Considering all the discussions, the court observed that AI has great potential for the judicial system of Pakistan. The court realized the sensitivity of the matter and the disclaimer clause of the GPT emphasized further testing to exploit the potential of AI fully. Moreover, the judge sent a copy of the order to the Lahore High Court and the Law and Justice Commission of Pakistan for their perusal and consideration as a law reform proposal.[54] Interestingly, the judge regarded the chatbot's responses as impressive based on the correct appreciation of the laws, emphasizing the judiciary to rely on the AI solutions, avoiding any reference to the legal hallucinations that he encountered throughout with the chatbot.[55] In Pakistan, the experience of both cases exhibits that the integration of AI in legal adjudication is at its beginning. In both instances, the judge confronted excessive, incorrect, and outdated responses, though he remains oblivious to the legal hallucinations and their consequences which may end up in a miscarriage of justice. The Federal Judicial Academy of Pakistan provides judges across Pakistan with "*Judge-GPT" – an* AI-powered solution – to assist the adjudication process. Neither the Pakistan Bar Council nor the Superior Judiciary of Pakistan have devised any conclusive ethics code to regulate the use of AI in legal

[48] *Ibid*., p. 8.
[49] *Ibid*., p. 9.
[50] *Ibid*., p. 12.
[51] *Ibid*., p. 13.
[52] *Ibid*., p. 14.
[53] *Ibid*.
[54] *Ibid*., pp. 17-18.
[55] *Ibid*., p. 16.

operations. Moreover, the courts have yet to identify cases where lawyers using AI-powered solutions have submitted drafts supported by hallucinated references.

C. Case Law Development in the UK

The first reported case[56] in the UK where the court confronted AI hallucination was found when the cases cited by a litigant were not genuine but rather generated through AI solutions. In the instant case, Mrs. Harber failed to notify His Majesty's Revenue & Customs (HMRC) of her liability to capital gain tax following the disposal of her property. Consequently, she was issued a failure to notify penalty. She filed an appeal with the First-tier Tax Tribunal (FTT) against HMRC on the pretext of a reasonable excuse for her mental health and resulting ignorance of the law. She presented nine cases in which FTT sided with the taxpayer. However, the HMRC's representative asserted that the cases she presented were not identifiable. After the verification, the FTT concluded that the cases were not genuine, rather they were fabricated and generated through an AI tool like ChatGPT, though these cases were plausible but inaccurate. She confirmed that the cases were provided by a friend in a solicitor's office and could be AI-generated. The court disregarded the fabricated cases, and the appellant lost the appeal. The court opined that in addition to wasting time and other resources, attributing fake opinions to the judges and courts can damage their reputation, and harm the repute of any party wrongfully associated with illusionary conduct.[57]

D. Case Law Development in the Australia

In a July 2024 hearing, a Melbourne lawyer was referred to investigation for presenting fabricated precedents in a family lawsuit. The attorney representing a husband provided the family court with a list of cases to support his plea. Neither the judge nor her associates could identify the enlisted cases. The lawyer confirmed that he used Leap, an AI-powered legal software specifically designed for legal use like Lexis+ AI, to prepare the list without verifying its accuracy and offered an unconditional apology. He paid the other party's solicitor for the costs of the thrown-away hearings. The court referred him to an investigation to appraise professional conduct issues based on the growing use of AI in legal operations. The family court has yet to issue guidelines on the use of AI. The Supreme Court of Victoria has already issued standards that the lawyers using AI should know the inherent limitations of these tools and how they work.[58]

Even though an AI model designed specifically for legal use can still create false or inaccurate information. AI solutions offer various means to validate their accuracy. For instance, 66,000 legal professionals are using Leap worldwide and it provides free verification through a human expert in the local laws, also known as human-in-the-loop. It is the lawyers' ethical obligation to verify the sources. Based on the request, the

[56] Harber v. Commissioners for His Majesty's Revenue and Customs, *[2023] UKFTT 1007 (TC),* https://www.casemine.com/judgement/uk/65720f72cd29093de5347804.

[57] Burges Salmon, A cautionary tale of using AI in law; UK case finds that AI generated fake case law citations, UK, (Dec. 18, 2023), https://www.lexology.com/library/detail.aspx?g=18d97112-59a2-4513-af0f-bedc4bb594cc.

[58] Josh Taylor, *Melbourne lawyer referred to the complaints body after AI generated made-up case citations in family court,* (Oct. 10, 2024), https://www.theguardian.com/law/2024/oct/10/melbourne-lawyer-referred-to-complaints-body-after-ai-generated-made-up-case-citations-in-family-court-ntwnfb.

lawyer was provided with the correct information within four hours which he didn't utilize in court.[59] Before this, a group of Australian academics in November 2023, sought an apology for submitting AI-generated reports through *Google Bard,* now *Gemini,* against Big Four consultancy firms in submission to a parliamentary inquiry.[60]

E. Case Law Development in the Canada

Likewise, In February 2024, a Canadian lawyer was referred to investigation for producing fictitious cases generated through ChatGPT in a child custody case in the Supreme Court of British Colombia. The attorney, Chong Ke, represented a father who wanted to take his children on a foreign trip. However, he was locked in a separation dispute with his wife. Chong Ke employed AI for precedents applicable to her client's circumstances. ChatGPT generated three responses and Key produced two of them in court. The opposing lawyer, however, could not trace those cases. Based on the confronted differences, Ke backtracked, maintaining the cases might be erroneous based on the AI-generated tool. She submitted an unconditional apology in the Court, having no intention to mislead the Court or the opposing counsel. The Court considered the submission of fake cases an abuse of process, which is equal to making false statements in the court and could lead to the miscarriage of justice. Her conduct is now under investigation by The Law Society of British Colombia, which issued guidelines on the appropriate use of AI in the delivery of legal services.[61]

F. Impacts of Hallucinations

Given the analysis of the cases, legal hallucinations may pose the following potential repercussions. In the first place, legal hallucinations impact lawyers by introducing inaccuracies to legal documents, damaging their integrity and credibility. It can breach ethical standards and professional responsibilities, leading to disciplinary actions and adding civil liabilities. In the second place, AI may augment legal services, but their hallucinations impact clients represented by the attorneys and may trigger distrust in the justice system. Inaccuracies in legal arguments can undermine their case, resulting in unfavorable judgments causing monetary losses or even wrongful convictions. In third place, legal hallucinations impact courts and judges, leading to miscarriage of justice. It diminishes the integrity of the judicial process, wasting time and resources to validate information and erode trust in the legal system.[62]

III. RESPONSE TO AI HALLUCINATIONS

Considering the amplifying tendency towards AI in the legal province and its susceptibility to hallucination, a regulatory response in the form of patchwork has been

[59] Ibid.

[60] Henry Belot, *Australian academics apologize for false AI-generated allegations against big four consultancy firms* (Nov. 2, 2023), https://www.theguardian.com/business/2023/nov/02/australian-academics-apologise-for-false-ai-generated-allegations-against-big-four-consultancy-firms; *See also,* AI Hallucinations & Legal Pitfalls, (Sept. 17, 2024), https://www.madisonmarcus.com.au/news-media/areas-of-law/artificial-intelligence-law-areas-of-law/ai-hallucinations-legal-pitfalls/?cn-reloaded=1.

[61] Leyland Cecco*, Canada lawyer under fire for submitting fake cases created by AI chatbot,* (Feb. 29, 2024), https://www.theguardian.com/world/2024/feb/29/canada-lawyer-chatgpt-fake-cases-ai.

[62] John Doe, *Trust But Verify: Avoiding the Perils of Over-Reliance on AI in Legal Practice*, JD Supra (Dec. 1, 2024), https://www.jdsupra.com/legalnews/trust-but-verify-avoiding-the-perils-of-8176236/;

evolving. The following segment examines responses to attorneys' irresponsible use of generative AI.

A. Judicial Responses

The increasing number of judges issuing AI orders varies in terms of breadth of coverage. Some judges prohibit the use of AI altogether, while some only prohibit it if lawyers do not verify accuracy; and some require submissions relating to the protection of confidential client information. We can categorize these responses into the following heads:

1. The Courts Requiring Confirmation on the Use of AI

After the New York federal court of show cause order in *Mata,* the first-ever reported case in which an attorney was caught using generative AI with fabricated outcomes, the Texas Court Judge Brantley Starr issued the first standing order governing the employment of generative AI.[63] Starr updated the individual practice rule by mandating a certificate about generative AI, which requires both the attorneys and the litigants to file a declaration in the court that no segment of the filing is drafted via generative AI, or if any segment is so drafted will be counter verified because these AI tools tend hallucinations and can provide biased information. In case of failure to file the required certificate, Starr's rule directed to strike such filing under Rule 11 of the Federal Rules of Civil Procedure irrespective of whether the draft or any portion thereof is AI-generated.[64]

Likewise, Judge Michael Baylson of the District Court of Pennsylvania issued a standing order regarding the disclosure of generative AI, requiring the attorneys to clarify where AI is used and to certify that each citation and reference has been verified.[65] Similarly, Magistrate Judge Gabriel Fuentes only mandated a certificate when a party actively uses generative AI, including disclosure about the filing and the specific tool employed.[66] Judge Scott Palk of Oklahoma issued the same standing

[63] Mata v. Avianca, Inc., No. 22-cv-1461 (PKC), 2023 WL 4114965, at *1 (S.D.N.Y. June 22, 2023); Sara Merken, Wary Courts Confront AI Pitfalls as 2024 Promises More Disruption, REUTERS, https://www.reuters.com/legal/transactional/wary-courts-confront-ai-pitfalls-2024-promises-2023-12-27/; Shannon Capone Kirk, Emily A. Cobb & Amy Jane Longo, *Judges Guide Attorneys on AI Pitfalls with Standing Orders*, ROPES & GRAY (Aug. 2, 2023), Shannon Capone Kirk, Emily A. Cobb & Amy Jane Longo, Judges Guide Attorneys on AI Pitfalls with Standing Orders, ROPES & GRAY (Aug. 2, 2023), https://www.ropesgray.com/en/insights/alerts/2023/08/judges-guide-attorneys-on-ai-pitfalls-with-standing-orders.

[64] *Judge Brantley Starr – Judge Specific Requirements: Mandatory Certification Regarding Generative Artificial Intelligence*, U.S. DIST. CT. N. DIST. TEX., https://www.cit.uscourts.gov/sites/cit/files/Order%20on%20Artificial%20Intelligence.pdf.

[65] J. Michael M. Baylson, Standing Order Re: Artificial Intelligence ("AI") in Cases Assigned to Judge Baylson, (E.D. Pa. June 6, 2023), https://www.paed.uscourts.gov/sites/paed/files/documents/procedures/Standing%20Order%20Re%20Artificial%20Intelligence%206.6.pdf.

[66] Mag. J. Gabriel A. Fuentes, Standing Order for Civil Cases Before Magistrate Judge Fuentes, at 2 (N.D. Ill. May 31, 2023), https://www.ilnd.uscourts.gov/_assets/_documents/_forms/_judges/Fuentes/Standing%20Order%20For%20Civil%20Cases%20Before%20Judge%20Fuentes%20rev%27d%205-31-23%20(002).pdf.

order requiring disclosure about the use of AI and specific tools employed, coupled with a declaration about the accuracy of the draft and its supported citations.[67]

A New York Judge, Arun Subramanian, did not necessitate a disclosure but stressed that the attorneys must personally confirm the accuracy of the content before being presented to the court. The court ruled that the use of ChatGPT or other tools is prohibited unless the accuracy of these tools is personally confirmed.[68] On the other hand, New Jersey federal judge, Evelyn Padin, mandates the disclosure of the use of generative AI and certification that the accuracy of AI-generated content is confirmed under human supervision.[69] A District Judge of Hawaii, Leslie Kobayashi, directed that any party employing AI must disclose the use of AI along with the specific tool used and certify the authenticity of the generated contents, including citations. In case of default, the party will be held accountable under Rule 11, which may lead to the imposition of sanctions.[70] The US Magistrate Judge Jeffrey Cole of Illinois while adopting the standing order for the use of generative AI, requiring both disclosure and certification. The court declared that generative AI, by producing fabricated and inaccurate citations, compromises the court's mission to ascertain truth.[71]

In addition to the trial courts, other US federal judges have followed Judge Starr's pattern for governing the use of AI. For instance, the Bankruptcy Court of Texas requires that if someone uses generative AI while preparing a filing, they must ensure the accuracy of the generated text through reliable means, including conventional legal databases and print reports.[72] In the appellate courts, the US Court of Appeals for the Fifth Circuit was the first to give notice of the proposed rule governing the employment of generative AI. The court proposed an amendment to Fifth Circuit Rule 32.3 to add language addressing AI use to its existing certificate of compliance, which includes a certificate of whether generative AI was used, its extent, and its accuracy approval by a human.[73] Likewise, Juge Roy Ferguson of the 394th District Court in Texas was the first state court to pass a standing order governing the employment of generative AI. The order mandated the filers to certify that all the generative content is substantiated

[67] J. Scott L. Palk, Chambers of United States District Judge, Disclosure and Certification Requirements – Generative Artificial Intelligence, https://perma.cc/VYZ8-XNGH.

[68] J. Arun Subramanian, United States District Court Southern District of New York, Individual Practices in Civil Cases, at 7 (2023), https://perma.cc/SNN5-N6HR.

[69] Judge Evelyn Padin's General Pretrial and Trial Procedures 2 (2023), https://perma.cc/M6RY-FVGP.

[70] J. Leslie E. Kobayashi, Chambers of United States District Judge, Disclosure and Certification Requirements – Generative Artificial Intelligence, https://perma.cc/Z63A-VSQX.

[71] Mag. J. Jeffrey Cole, United States District Court for the Northern District of Illinois, The Use of "Artificial Intelligence" in the Preparation of Documents Filed Before This Court, https://www.ilnd.uscourts.gov/_assets/_documents/_forms/_judges/Cole/Artificial%20Intelligence%20standing%20order.pdf.

[72] United States Bankruptcy Court for the Northern District of Texas, General Order 2023-03, Pleadings Using Generative Artificial Intelligence (June 21, 2023), https://perma.cc/JQ6Y-THKV.

[73] Jacqueline Thomsen, Lawyers Must Certify AI Review Under Fifth Circuit Proposal, BLOOMBERG L. (Nov. 21, 2023, 6:26 PM), https://news.bloomberglaw.com/us-law-week/lawyers-must-certify-ai-review-under-fifth-circuit-proposal; see also, https://www.ca5.uscourts.gov/docs/default-source/default-document-library/public-comment-local-rule-32-3-and-form-6.

as accurate via conventional legal methods by an attorney licensed to practice law in Texas.[74]

2. Non-Disclosure of Confidential Information

A federal Judge, Stephen Alexander Vaden of the United States Court of International Trade, issued an 'Order on AI' showing concerns about privacy invasion that these tools learn from users' interaction and cannot differentiate between confidential and non-confidential information. Hence, Judge Vaden mandated two things with AI-generated filings: A disclosure notice regarding the tool employed along with the segment generated and a declaration that the use of AI has not disclosed any confidential information to an unauthorized person.[75] Likewise, the Bankruptcy Court of Oklahoma, while quoting Judge Starr, mandated disclosure about the AI tool, the details of the specific portion for which generative AI was employed, a certificate of accuracy checking, and to confirm that generative AI has not caused the disclosure of any confidential information to any unauthorized party.[76]

3. The Courts Prohibiting the Use of AI Solutions

The US District Judge of Montana, Donald Molloy, prohibited the employment of generative AI software like ChatGPT.[77] Judge Michael Newman of Ohio prohibited the use of generative AI and warned of the sanctions that might be imposed for using AI, including monetary, contempt, and dismissal of the suit. However, the court allows information collection from legal search engines like LexisNexis and Westlaw and common search engines like Google.[78] Similarly, Judge Stephen Clark of Missouri banned the use of generative AI.[79]

The courts' responses to the use of AI vary across the US necessitating the attorneys to double-check each court's policy on the use of AI before filing any submission to avoid any potential complications. As discussed, some courts allow the employment of AI subject to the disclosure of its use, the tool so employed, the extent of its assistance, and the confirmation of its accuracy. In addition to these standards, some courts require the confirmation that the employment of AI has not disclosed clients' confidential information to any unauthorized person. In contrast, some courts prohibited the use of AI altogether.

B. Other Responses

[74] District Court for the 394th Judicial District of Texas, Standing Order Regarding Use of Artificial Intelligence (June 9, 2023), https://edrm.net/wp-content/uploads/2024/04/Judge-Roy-Ferguson.pdf.

[75] Hon. Stephen Alexander Vaden, Order on Artificial Intelligence, 1 (Ct. Int'l Trade June 8, 2023), https://www.cit.uscourts.gov/sites/cit/files/Order%20on%20Artificial%20Intelligence.pdf.

[76] United States Bankruptcy Court for the Western District of Oklahoma, General Order 23-01, Pleadings Using Generative Artificial Intelligence (July 25, 2023), https://www.okwb.uscourts.gov/sites/okwb/files/GenOrder23-01.pdf.

[77] *Belenzon v. Paws Up Ranch, LLC, No. CV 23-69-M-DWM, 2023 U.S. Dist. LEXIS 123020, at 1 (D. Mont. June 22, 2023), https://casetext.com/case/belenzon-v-paws-up-ranch-llc.*

[78] Hon. Michael J. Newnan, United States District Court for the Southern District of Ohio, Standing Order Governing Civil Cases, at 11 (Dec. 18, 2023), https://perma.cc/V6P6-BSRZ.

[79] Self-Represented Litigants (SRL), U.S. Dist. Ct. E. Dist. Mo.: Hon. Stephen R. Clark, C.J. & Nathan M. Graves, Clerk of Ct., https://www.moed.uscourts.gov/self-represented-litigants-srl.

Following the Mata case ruling, the legal community is now more aware of using generative AI tools, necessitating policy guidelines for governing AI in legal operations. The policy should sensitize legal professionals about the appropriate use of AI and its ethical concerns. AI should be employed only to assist and augment legal services, but not at the cost of lawyers' subjective judgment and expertise. Moreover, attorneys should be held responsible for validating the accuracy of the generated contents. While employing AI in legal services, clients should also be taken into confidence. Attorneys should remain current about the emerging trends in AI as their ethical duty of technological competence.

MIT convened the first task forces in response to Mata's case to ensure responsible use of generative AI.[80] The State Bar of Texas initiated a task force to explore the proper employment of AI in legal services. The task force aimed to ensure that technological advancement served the community without compromising values central to the legal community. The Texas Task Force made numerous recommendations to the state bar, including technological education and ethical use of AI.[81] The New York Bar Association also declared its own AI task force to appraise the impacts of evolving technology on the legal profession and society.[82] The American Bar Association (ABA) announced the formation of a national task force to assess the risks of AI on the legal profession, including data privacy, disinformation, and cybersecurity. Further to examine AI governance, AI in legal education, and AI in access to justice. To address the impacts and ethical concerns of AI and provide insights on the trustworthy and responsible use of AI.[83]

In addition to the task forces, two ethics bodies have responded to the issue of AI. The Board of Governors Review Committee of the Florida Bar considered an advisory opinion on the use of AI after an inquiry on AI tools. The committee issued a proposed advisory opinion to address issues that the attorneys employing AI must take reasonable steps to safeguard clients' privacy information, a reasonable oversight on the use of generative AI, and lawyers must not entrust their subjective judgment to generative AI. The proposed opinion also demands lawyers to charge only a reasonable fee and should not overly charge their clients for using AI. Lawyers may publicize the employment of generative AI but cannot claim its authority over others unless the same is objectively verifiable. Since generative AI is still in its beginning, the existing ethical concerns should not be treated as final.[84]

[80] Dazza Greenwood, Task Force on Responsible Use of Generative AI for Law, MIT Computational Law Report (Feb. 28, 2023), https://law.mit.edu/pub/generative-ai-responsible-use-for-law/release/9.

[81] State Bar of Tex., Taskforce for Responsible AI in the Law (Trail) 2–3 (2023), https://www.texasbar.com/AM/Template.cfm?Template=/CM/ContentDisplay.cfm&ContentID=61655.

[82] Richard Lewis, What the NYSBA AI Task Force Hopes to Achieve for Law Practice, BLOOMBERG L. (July 31, 2023), https://news.bloomberglaw.com/us-law-week/what-the-nysba-ai-task-force-hopes-to-achieve-for-law-practice.

[83] ABA Forms Task Force to Study Impact of Artificial Intelligence on the Legal Profession, AM. BAR ASS'N (Aug. 28, 2023), https://www.americanbar.org/news/abanews/aba-news-archives/2023/08/aba-task-force-impact-of-ai/.

[84] Proposed Advisory Opinion 24-1 Regarding Lawyers' Use of Generative Artificial Intelligence – Official Notice, FLA. BAR (Nov. 13, 2023), https://www.floridabar.org/the-florida-bar-news/proposed-advisory-opinion-24-1-regarding-lawyers-use-of-generative-artificial-intelligence-official-notice/.

The State Bar of California necessitated the governance of generative AI. Its Committee on Professional Responsibility and Conduct (COPRAC) provided recommendations on and stipulated practical guidance on the use of generative AI. It examines how generative AI impacts professional responsibility obligations, including confidentiality, competence, supervision, and charging only a reasonable fee.[85]

Additionally, the state of Michigan is accredited to be the first to issue a Judicial Ethics Opinion, addressing judges' ethical obligation concerning generative AI that judicial officers must keep up with technological advancements including AI. It further says that with the proliferation of AI, the judges must comprehend the legal, regulatory, and ethical challenges of AI and consistently appraise how they or parties before them are employing AI in their docket.[86]

In Pakistan, the National Artificial Intelligence Policy, 2022 is launched with the proposed establishment of an AI regulatory directorate to ensure the ethical and responsible use of AI.[87] However, there is no reference to dealing with the emerging issues of AI hallucinations in decision-making. Notably, the Federal Judicial Academy of Pakistan facilitated judges across Pakistan with the *Judge-GPT* AI tool to assist the decision-making process, without providing guidelines about its probabilistic nature that could lead to plausible but inaccurate responses. In critical areas like health, finance, and law, where accuracy is paramount, fabricated outcomes can cause irreparable loss. In legal services, fictitious precedents could cause a miscarriage of justice. Neither the government, bar, bench, nor law firms have established definite standards on the rapidly evolving issue of hallucinations. Necessitating the establishment of a task force, comprising experts from the academia, government, judiciary, and tech developers, to devise an exclusive policy to deal with AI in legal practices and its ethical challenges.

In the UK, Artificial Intelligence (AI) Guidance for Judicial Office Holders, 2023, offers comprehensive guidelines about the responsible employment of AI. It underscores the limitations and capabilities of AI and urges its conformity with the judiciary's overreaching obligation to protect the integrity of the administration of justice. It applies to all the courts and tribunals across the UK and provides the following guidelines for the responsible use of AI: the AI chatbot produces results based on the prompts they receive, the data they are trained, the information available on the internet, and may generate a plausible but inaccurate response. Confidentiality and privacy are another concern. The public chatbot retains every prompt and information, which may be utilized in responding to other users, invading data privacy. Likewise, the accuracy of the responses must be confirmed before being used or relied upon. The AI tools may cause fabricated citations, cases, and quotes, or may refer to legal text that doesn't exist. Hence, these tools cannot be left unaccountable. It further provides for biases, security, responsibility, and potential employment of AI by other users. The draft exemplifies the positive use of AI such as its capabilities to summarize large legal text, prepare presentations, and perform administrative tasks like drafting emails and memos. However, the draft discourages conducting legal research on AI

[85] Memorandum from the Comm. on Pro. Resp. & Conduct to Members, Bd. of Tr. Sitting as the Regul. and Discipline Comm. 1 (Nov. 16, 2023), https://aboutblaw.com/bbpZ.

[86] State Bar of Mich., Ethics Op. JI-155 (2023), https://perma.cc/C58T-GCLX.

[87] Ministry of Information Technology and Telecommunication, National AI Policy Consultation Draft V1
(2022), https://moitt.gov.pk/SiteImage/Misc/files/National%20AI%20Policy%20Consultation%20Draft%20V1.pdf.

tools that cannot be independently counter-verified and legal analysis because the current tools are incapable of producing convincing reasoning or analysis.[88]

The Law Society of England and Wales also provided an AI strategy focusing on the following three long-term outcomes: innovation, to benefit both firms and clients; impact, to have an effective regulatory landscape; and integrity, to ensure the responsible and ethical employment of AI to advance the rule of law and access to justice. It embraces endeavors to ensure that the legal system operates impartially, safeguards individual rights, and advances the cause of justice, including the protection of the rights of marginalized communities, addressing prejudices, and striving to ensure that the legal system upholds principles of justice for every member of society.[89]

In Australia, the legal profession regulators from across the three uniform law states have jointly issued a statement to guide legal professionals in their ethical and responsible use of AI: the Law Society of NSW, the Legal Practice Board of Western Australia, and the Victorian Legal Service Board and Commissioner have established common principles to protect the client from risk, technology is employed for their benefits, and uphold the principles of justice. The following are the key points of the statement: while enjoying the assistance of AI, lawyers are obliged to provide accurate legal information, and it is not the duty of the AI tool being employed. The practitioners must understand AI, its capabilities, and the limitations of LLMs. This statement helps the lawyers understand the regulators' expectations when they employ AI to assist them in providing legal services. The regulators will frequently review and update their guidance on AI as it continues to evolve. While using AI, legal professionals must maintain high ethical standards and rules of conduct, including upholding clients' confidentiality, advising their clients, competent and diligent provision of legal services, charging a reasonable, fair, and proportionate fee, transparency, and limiting the use of AI.[90]

The Canadian Judicial Council (CJC) issued Guidelines for the Use of AI in Canadian Courts, 2024, which provides a framework for the responsible use of AI in judicial processes. It underlines upholding judicial independence, adhering to the core values and ethical standards, and ensuring information security, transparency, and accountability in AI-generated content. It underscores the significance of regular impact assessments, sensitizing, and user support for judges. It aims to strike a balance between innovation and caution, ensuring that AI advances the efficiency of legal services without compromising the integrity of the judicial system. The guidelines are broadly categorized into the following seven heads. First, protection of judicial independence, restricting the parliament's authority to empower a state agency from the legislative and judicial branches to oversee the use of AI by and before courts. Where the government moves forward with legislation to regulate AI, the independence of the judiciary must be preserved. Second, judges' employment of AI must adhere to the core values and ethical rules, including integrity, competence, impartiality, transparency, fairness, and accessibility to justice. Third, regards aspects like privacy and intellectual

[88] Judicial Office, Artificial Intelligence (AI) Guidance for Judicial Office Holders (2023), https://www.judiciary.uk/wp-content/uploads/2023/12/AI-Judicial-Guidance.pdf.

[89] The Law Society, Artificial Intelligence (AI) Strategy (2023), https://www.lawsociety.org.uk/topics/ai-and-lawtech/artificial-intelligence-ai-strategy.

[90] Legal Services Board of Victoria, Statement on the Use of Artificial Intelligence in Australian Legal Practice (2023), https://www.lsbc.vic.gov.au/news-updates/news/statement-use-artificial-intelligence-australian-legal-practice.

property, creating an equilibrium between safety and accuracy. Fourth, strictly adhere to the information security standards through robust information and cyber security programs. Fifth, the AI tools must provide reasons and explanations for their decision-making and generative outcomes. Sixth, to keep regular track of the use of AI considering judicial independence, security, privacy, access to justice, and the court's reputation. Seventh, user support and education, including judges training which is a prerequisite for upholding and maintaining judicial independence and technical support for AI integration in the administration of justice. The seven points outlined by the CJC reaffirm that AI should not be employed without a comprehensive understanding of the best practices for integrating technology.[91]

To conclude the above responses, the judges, bar, and law firms contribute to developing AI rules, but their contribution is a patchwork. The courts' responses create confusion even in the patchwork: certain courts proscribed the employment of generative AI, few require disclosure and certification, while some do not. Some judges are concerned about data confidentiality. Hence, attorneys should stay vigilant of technological advancement considering the applicable and often diverging court rules. The growing tendency of AI in legal operations necessitates an exclusive national policy governing the use of generative AI and its ethical challenges in the legal province.

The most striking question is how to overcome hallucinations in legal operations. Legal hallucinations are the byproduct of many contingencies and could be addressed accordingly. The following segment explains how to curtail if could not completely remove, the issue of hallucinations from AI-powered solutions.

IV. WHY DO AI MODELS HALLUCINATE?

One of the main concerns in AI legal practice is dealing with AI hallucinations. Considering its probabilistic nature and its susceptibility towards fabricated responses, the AI hallucination mitigating techniques can be broadly divided into the following two heads:

A. Recommendations for the Developers

The use of high-quality training data helps diminish the prospects of hallucinations. The first stage that leads to the likelihood of hallucinations is the lack of accuracy and reliability of datasets. Hence, hallucinations are inversely proportional to the accuracy and consistency of datasets. Hallucinations tend to decline as the precision and trustworthiness of the training data enhance so using data templates or structured data formats is advisable. Improving the quality of the training sample and subsequent testing of the generative data can help reduce the possibility of hallucinations. Clearly outlining what AI is tasked to do can generate focused and appropriate outcomes. Putting restrictions on the AI's responses can help improve the performance and reliability of the LLMs.

By applying modern artificial neural network architectures such as Convolutional Neural Networks (CNNs) and Long Short-Term Memory (LSTMs), the

[91] Canadian Judicial Council, Guidelines for the Use of Artificial Intelligence in Canadian Courts (2024), https://cjc-ccm.ca/sites/default/files/documents/2024/AI%20Guidelines%20-%20FINAL%20-%202024-09%20-%20EN.pdf.

rate of hallucinations can be significantly controlled: CNNs are useful for comprehending the context and structure of legal documents because they are exceptional to identify spatial hierarchies in data. By deploying CNNs to the legal data, LLMs can help realize the complexities and intricacies of legal language, diminishing the prospects of fictional outcomes. Likewise, LSTMs can help improve sequence prediction because they are designed to retain long sequential data, making them best for dealing with extensive legal documents, preserving context for an extended period, and allowing the LLMs to generate precise and relevant responses. LSTMs address vanishing and gradient problems encountered in other networks.[92] A hybrid of both these architectures can help design more robust models: CNNs for extracting features from the legal data while LSTMs for handling the sequential nature of legal documents, leading to a comprehensive understanding and producing legal text, reducing the likelihood of hallucinations. Extensive training of these models on specific legal data helps advance the accuracy of these models when they are exposed to voluminous legal text to learn different terminologies and contexts in the legal province, cutting errors and hallucinations.

Continuous model improvement, consistent updates, and advancements in AI models can help reduce the prospects of hallucinations, so regular appraisal and improvements of these models are inevitable. Human oversight is a significant tool to control the prospects of hallucinations. LLMs can be trained enough to overcome potential hallucinations by monitoring and correcting AI's responses.[93] AI-generated content should also undergo regression analysis before being presented or relied upon. Further, an ethical supervisor is advisable for the algorithms to monitor and impose ethical restrictions on the use of AI based on the idea that the former must have higher standards than the latter.[94] Humans in the loop are also recommended to review and correct AI responses with a focus to train AI intelligently not to repeat a fictitious outcome, and the end user may get the generated content counter-verified before relying on it.

Fine-tuning models for specific legal tasks may also reduce the likelihood of producing inaccurate responses by making small adjustments to the model's parameters, based on the existing knowledge of a model to learn new tasks. Implementation of robust evolution of metrics to frequently assess and address hallucination rates is also recommended. Another way to moderate the rate of hallucinations is to design a self-explanatory AI model, which can provide explanations and reasons for its decision-making process. This proposed model can help legal professionals identify potential hallucinations and appraise the reliability of AI content, resulting in more transparent and accountable AI systems.

Though these suggestions can significantly contribute to curtailing the rate of hallucinations, they cannot be completely overcome. LLMs operating on probabilistic

[92] Sepp Hochreiter et al., *Long Short-Term Memory*, 9, Neural Computation, (8): 1735–1780 (1997), https://dl.acm.org/doi/10.1162/neco.1997.9.8.1735; *see also, What is LSTM? Introduction to Long Short-Term Memory, Analytics Vidhya*, (Oct. 1, 2024), https://www.analyticsvidhya.com/blog/2021/03/introduction-to-long-short-term-memory-lstm/.

[93] What are AI hallucinations? IBM, https://www.ibm.com/topics/ai-hallucinations.

[94] Vadim Perov and Nina Perova, *AI Hallucinations: Is "Artificial Evil" Possible?,* USBEREIT, (2024), https://ieeexplore.ieee.org/document/10584048; *See also,* Amitai Etzioni and Oren Etzioni, *AI assisted ethics*, 18, Ethics Inf. Tech., (2016), https://link.springer.com/article/10.1007/s10676-016-9400-6.

algorithms attempt to predict or foresee the next word in a sequence by considering the prospect of various possible words that may lead to a conceivable but inaccurate response.[95] These processes by their very nature can lead to inaccuracies and potential hallucinations. Despite the widespread and high-quality training data, no dataset can cover every eventuality due to the complexity and context-dependent nature of languages.[96] Unlike conventional search engines, these models are designed to be creative and capable of generating diverse and stimulating results. Nevertheless, precision can compromise creativity, creating a challenge in maintaining an equilibrium between the two. The inherent limitations of the current machine learning algorithm create prospects of hallucinations in LLMs when generalizing from training data to unanticipated data.[97]

B. Recommendations for Legal Professionals

Given the above discussion, AI models still have the potential to produce fabricated responses owing to their probabilistic nature. Precision is highly valued in legal operations, so it is highly recommended that legal professionals counter-verify AI-generated content and citations against reliable sources. AI solutions should be used as a supplement to augment legal services rather than a substitute. Legal professionals must stay abreast of the limitations, common kinds of hallucinations and errors specific to legal context, and pitfalls of AI tools.

Prompt skilling can substantially mitigate the prospects of hallucinations. By crafting precise and effective commands, the AI models recognize exactly what is being queried. Accuracy and comprehension of these models can be further improved by employing the following techniques: prompt chaining, which breaks down a long and complex proposition into simple and sequential inputs. Employing multimodal or a diversity of prompts can help enhance AI comprehension. Consistent appraisal and feedback significantly contribute to refining AI models. Another advisable solution is to pass the AI-generated content through robust quality control by establishing review protocols for AI responses, including multiple layers of review: cross-referencing AI information with conventional databases, consultation with colleagues, peer reviews, human oversight, other rounds of fact-checking, and maintaining a healthy skepticism towards AI content.[98]

Legal professionals should regularly update their knowledge of AI innovations and best practices by participating in seminars and training sessions based on the efficient and ethical employment of AI in legal services. A regular audit of the AI tools is needed to ensure compliance with the approved standards and relevant laws,

[95] Major research into 'hallucinating' generative models advances reliability of artificial intelligence, University of Oxford, (Jun 20, 2024), https://www.ox.ac.uk/news/2024-06-20-major-research-hallucinating-generative-models-advances-reliability-artificial.

[96] Changlong Wu at al., *No Free Lunch: Fundamental Limits of Learning Non-Hallucinating Generative Models*, CSoI, Preprint under review, (2024), https://www.cs.purdue.edu/homes/spa/papers/hallucination_preprint.pdf.

[97] Minhyeok Lee, *A Mathematical Investigation of Hallucination and Creativity in GPT Models*, 11 Mathematics 2320 (2023), https://doi.org/10.3390/math11102320.

[98] AI Hallucinations: Legal Information Risks, Attorneys Media, https://attorneys.media/ai-hallucinations-legal-information-risks/.

necessitating calls for a regulatory framework.[99] A loop among legal researchers, practitioners, and AI developers is highly recommended for designing more trustworthy AI models. The developers should provide guidelines and training to their users on the effective employment of AI in legal services. Legal practitioners should stay connected with the AI service providers to report flaws and propose improvements, which can significantly contribute to refining these models by not repeating a specific hallucination.

CONCLUSION

AI-driven models are revolutionizing legal operations, simultaneously creating inherent challenges in navigating legal landscapes. Despite high-quality training and data optimization, LLMs are susceptible to hallucinations. This intrinsic problem is the outcome of their functional modalities: considering their probabilistic nature, the LLMs calculate the possibility of a particular word following in a sequence. While training the data, these models learn patterns, structures, and correlations between the words. These models follow a sequence based on the assigned probability of each word. These models depend on the context provided by the preceding data and complete the sequence of words by generating the most probable content comprehended in their co-relationship. In the given scenario, these models cannot verify the truthfulness and relevance of the context, hence the required outcomes might be plausible but imprecise or fictitious. The models are only concerned with a high probability of words next in sequence. These models, however, cannot authenticate the trustworthiness of their generated content, thus adding a disclaimer that AI-generated content may be inaccurate is feasible to exonerate civil law liability. It shifts the onus to the end user to counter-verify the generated content otherwise face the music.

Imposing limitations on the training data can help narrow the likelihood of false or fabricated content at the cost of creativity. Conversely, the models trained on widespread data without such confines may produce more novel outcomes. In sensitive fields like finance, healthcare, and law where precision is paramount, utmost care to avoid hallucinations is required, though at the expense of novelty. An equilibrium between hallucinations and creativity can help design a more robust and versatile model, capable of addressing complex tasks with enhanced performance, leaving the end user with ultimate liability to corroborate the generated content before relying on it.

Legal professionals are swiftly integrating AI systems into their legal provinces without fully realizing their probabilistic nature which could lead to fabricated outcomes, affecting the cause of justice. Given their utility despite their unpredictable nature, these AI systems can be referred to as necessary evil. They are unavoidable owing to the unparalleled services they offer, but their irresponsible employment can transpire into malpractice, compromising the reputation of lawyers, and creating financial liability. Even if judges and lawyers are reluctant to use AI, they still need to

[99] Kiara Brunel Fink, *AI Hallucinations in Legal Practice: Risks, Real Cases, and Solutions,* Mondaq, https://www.mondaq.com/new-technology/1540712/ai-hallucinations-in-legal-practice-risks-real-cases-and-solutions.

learn AI. Since AI can go wrong, legal professionals are under obligation to act as guardians of the legal system to correct their abuses.[100]

Transparency and accountability could help moderate the probability of hallucinations. The AI enterprises must be transparent about conceivable errors, including accountability measures and setting up expectations for clients where the AI-produced content leads to issues. Bar Associations such as California, Florida, and New York have published guidelines for the trustworthy use of AI in legal operations.[101] More than 25 US Federal Judges passed standing orders requiring lawyers to reveal and circumscribe the use of IA in their courtrooms.[102] The judges' directions to the attorneys to certify the use and accuracy of AI in their briefs are motivated by the ethical challenges posed by AI and the significance of the precision of documents submitted in the court.

By incorporating a disclaimer about the accuracy of the generated content, the AI developers exculpate their liability for disseminating fictitious content. However, AI responses based on erroneous opinions could damage the reputation of judges, courts, or parties implicated in fictional conduct. The greatest risk lies on the part of the legal user who may not and arguably should not be able to escape liability for over-reliance on AI. Although these tools are still in their developmental stages and evolving towards maturity, the judiciary and legal community must determine the acceptable extent of their fabricated responses, necessitating the establishment of policy guidelines. Regardless of their legal liability, AI enterprises are responsible for providing trustworthy and reliable services. They must adhere to ethical and legal standards, confirming their models do not create harmful or misleading responses.

To have confidence in the AI solutions, a shared liability clause in user agreements should be incorporated, clearly outlining the responsibility of both the users and the service providers and demonstrating the extent of their liability in cases where AI hallucinations cause harm. For instance, the European Union (EU) has recently initiated a Product Liability Directive (PLD), placing obligations on AI tool developers, suppliers, and other entities for providing defective products, including AI software. So, the manufacturers can be held accountable for the harm caused by defects in their AI solutions, and the injured party is not even required to prove fault or negligence.[103] In

[100] David Rubenstein, *2024 Selected Topics and Miscellany CLE,* Washburn University School of Law*,* Presentation (June 13, 2024), https://www.washburnlaw.edu/about/community/cle/_files/selected-topics-schedule.pdf.

[101] David Alexander, New York State Bar Association Task Force To Address Emerging Policy Challenges Related to Artificial Intelligence, N.Y. St. Bar Ass'n (July 17, 2023), https://nysba.org/new-york-state-bar-association-task-force-to-address-emerging-policy-challenges-related-to-artificial-intelligence/; *See also*, Report and Recommendations of the New York State Bar Association Task Force on Artificial Intelligence, N.Y. St. Bar Ass'n (April 2024), https://nysba.org/app/uploads/2022/03/2024-April-Report-and-Recommendations-of-the-Task-Force-on-Artificial-Intelligence.pdf; The State Bar of California, Practical Guidance for the Use of Generative Artificial Intelligence in the Practice of Law, State Bar of Cal. (2023), https://www.calbar.ca.gov/Portals/0/documents/ethics/Generative-AI-Practical-Guidance.pdf; The Florida Bar, Florida Bar Ethics Opinion, Technical Report 24-1, Fla. Bar (2024), https://www.floridabar.org/etopinions/opinion-24-1/.

[102] Law360, Tracking Federal Judge Orders on Artificial Intelligence (2024), Pulse (law360), https://www.law360.com/pulse/ai-tracker.

[103] Kennedys Law, *A New Liability Framework for Products and AI*, https://kennedyslaw.com/en/thought-leadership/article/2024/a-new-liability-framework-for-products-and-ai/.

addition to PLD, the EU is considering the AI Liability Directive (AILD), to address risks posed by AI tools by introducing a fault-based civil liability regime, which would require proving the developer's fault or negligence where AI solutions cause harm. These directives are part of the EU's comprehensive efforts to regulate AI, offering users legal pathways to seek compensation for any damage caused by AI tools.[104]

[104] Giskard, *AI Liability in the EU: Business Guide to Product (PLD) and AI Liability Directives (AILD),* https://www.giskard.ai/knowledge/ai-liability-in-the-eu-business-guide-to-product-pld-and-ai-liability-directives-aild; Kennedys Law, A New Liability Framework for Products and AI, https://kennedyslaw.com/en/thought-leadership/article/2024/a-new-liability-framework-for-products-and-ai/.

DEEP FAKES, DEEPER CONSEQUENCES: COMBATING AI CHILD PORNOGRAPHY BY MANDATING SEX OFFENDER REGISTRATION

Allison Mitton[*]

Abstract: Recent advancements in artificial intelligence and machine learning have led to deepfakes and AI-generated images being created and distributed at an unprecedented rate. While deepfakes are used for many purposes, the overwhelming majority are used to create non-consensual deepfake pornographic content depicting women and minors. This raises a critical issue: If deepfake pornography is so prevalent, how can the law effectively intervene to prevent more individuals from becoming victims? Shockingly, little to no effective federal legislation has been enacted to combat deepfake pornography—even when the images depict minors. I suggest that the most effective legislation would both (1) deter individuals from publishing deepfake pornography involving minors and (2) raise awareness of those who exploit others' images in violating ways. To accomplish this, I argue that the TAKE IT DOWN Act should be amended to include a provision requiring mandatory sex offender registration for those who publish deepfake pornography of minors. By incorporating this simple addition into a proposed federal law already poised for success, states can help prevent more people from becoming victims of deepfake pornography by publicly identifying individuals who may pose a threat.

Keywords: Deepfakes; Artificial Intelligence; Pornography

[*] J. Reuben Clark Law School, Brigham Young University, US.

Table of Contents

INTRODUCTION

You are outraged. You just found out a classmate created a deepfake nude video of you and posted it online, where it's getting hundreds of views. You call the police, hoping for some help, but they tell you there is nothing they can do—to your horror, what happened to you is not considered a crime in the state you live in.

This nightmare scenario was a reality for Francesca Mani, a fourteen-year-old New Jersey high school student.[1] After being brought to her vice principal's office despite knowing she had done nothing wrong, she was told that she—and several other female peers—had been depicted in fake nude images created by male classmates, who then shared these images with many other students.[2]

Although, in theory, deepfake pornography could depict people of all genders, women are disproportionately victimized.[3] Many minors, including middle and high school-aged girls, have been put on display in forged pornography created by predators or even their own peers.[4]

Because of recent advancements in artificial intelligence (AI) and machine learning, more deepfake pornography is being made now than ever.[5] According to experts, "[t]here are 550% more deepfake videos online in 2023 than in 2019," with a 464% increase in deepfake pornography between 2022 and 2023 alone.[6]

Although the problem of deepfake pornography has become so rampant, little to no effective legislation has been passed to help resolve the issue.[7] Shockingly, less

[1] Jessica Le Masurier, '*A Global Problem': US Teen Fights Deepfake Porn Targeting Schoolgirls*, FRANCE 24 (Apr. 18, 2024, 1:31 PM), https://www.france24.com/en/tv-shows/focus/20240418-a-global-problem-us-teen-fights-deepfake-porn-targeting-schoolgirls.

[2] *Id.*

[3] 99% of deepfake pornography depicts women. *See, e.g., 2023 State of Deepfakes*, SECURITY HERO, https://www.securityhero.io/state-of-deepfakes/#:~:text=Between%202022%20%26%202023%2C%20the%20amount,year%20was%20a%20startling%20464%25 (last visited Oct. 12, 2024); Arwa Mahdawi, *Nonconsensual Deepfake Porn is an Emergency that Is Ruining Lives*, THE GUARDIAN (Apr. 1, 2023, 9:00 AM), https://www.theguardian.com/commentisfree/2023/apr/01/ai-deepfake-porn-fake-images.

[4] *See generally* Kat Tenbarge, *Beverly Hills Middle School Expels 5 Students After Deepfake Nude Photos Incident*, NBC NEWS (Mar. 8, 2024, 11:55 AM), https://www.nbcnews.com/tech/tech-news/beverly-hills-school-expels-students-deepfake-nude-photos-rcna142480 (explaining that five eighth grade boys were expelled from a school in Beverly Hills, California, after creating and sharing deepfake pornographic images of sixteen of their female classmates); Hyung-Jin Kim, *In South Korea, Rise of Explicit Deepfakes Wrecks Women's Lives and Deepens Gender Divide*, PBS NEWS (Oct. 3, 2024, 6:55 PM), https://www.pbs.org/newshour/world/in-south-korea-rise-of-explicit-deepfakes-wrecks-womens-lives-and-deepens-gender-divide (stating that "[m]ost suspected perpetrators [of creating deepfake pornography] in South Korea are teenage boys . . . [who] target female friends, relatives and acquaintances—also mostly minors—as a prank, out of curiosity or misogyny.").

[5] *See, e.g.,* Stu Sjouwerman, *Exponential Deepfake Porn is Out of Control and a Huge Security Risk*, KNOWBE4 (Oct. 16, 2024), https://blog.knowbe4.com/exponential-deepfake-porn-is-out-of-control-and-a-huge-security-risk; *2023 State of Deepfakes*, *supra* note 3.

[6] *2023 State of Deepfakes*, *supra* note 3.

[7] *See, e.g.,* Le Masurier, *supra* note 1; Andrew R. Chow, *Francesca Mani: Anti-Deepfake Activist*, TIME (Sept. 5, 2024, 7:17 AM), https://time.com/7012803/francesca-mani; Michelle M. Graham, *Deepfakes: Federal and State Regulation Aims to Curb a Growing Threat*, THOMSON REUTERS (June 26, 2024), https://www.thomsonreuters.com/en-us/posts/government/deepfakes-federal-state-regulation; *contra Most States Have Enacted Sexual Deepfake Laws*, MULTISTATE AI (June 28, 2024),

than half of states have passed legislation restricting "the distribution of nonconsensual sexual deepfakes,"[8] and, as of this writing, no federal legislation has been passed to regulate deepfake pornography yet—even when it depicts minors.[9]

This Note argues that the current available remedies for victims of deepfake pornography pose challenges due to their heavy reliance on civil law and the inconsistency of state criminal laws, so it is imperative for states to help prevent more people from becoming victims by publicly identifying individuals who may pose a threat through sex offender registries. To do so, a provision should be added to the proposed TAKE IT DOWN Act to mandate sex offender registration for those who publish deepfake pornography featuring minors.

Part I gives a brief background on the emergence of deepfakes and how they operate. Part II will discuss the current available civil and criminal remedies for victims of deepfake pornography. Part III describes constitutional First Amendment issues which pose an obstacle to passing legislation. Part IV describes solutions others have offered, including holding tech companies liable and federal proposed legislation. Finally, Part V argues that the TAKE IT DOWN Act should be amended to include mandatory sex offender registration for those who publish deepfake pornography featuring minors and outlines a simple approach for its implementation.

I. EMERGENCE OF DEEPFAKES

A. Deepfakes and How They Work

Deepfakes are forged videos or images "created via deep learning,[10] a form of artificial intelligence, where a person's likeness, including their face and voice, can be realistically swapped with someone else's."[11] The term "deepfake" was first coined in 2017, when a person under the username "deepfakes" started posting deepfake celebrity pornography on Reddit.[12] The term itself is a portmanteau of the words "deep" (to signal that it was created through deep-learning AI technology) and "fake" (to signal that the content was created using AI).[13]

https://www.multistate.ai/updates/vol-32 (stating that most states have passed enacted laws; however, these solutions are ineffective).

[8] *Most States Have Enacted Sexual Deepfake Laws*, *supra* note 7. *See also* Graham, *supra* note 7, *Can State Laws Actually Stop Political Deepfakes?*, MULTISTATE AI (Apr. 15, 2024), https://www.multistate.ai/updates/vol-22; *Dozens of AI Laws Go Into Effect*, MULTISTATE AI (July 12, 2024), https://www.multistate.ai/updates/vol-33.

[9] Graham, *supra* note 7.

[10] For more information on deep learning generally, see *What is Deep Learning?*, AMAZON WEB SERVICES, https://aws.amazon.com/what-is/deep-learning (last visited Oct. 16, 2024) ("Deep learning is a method in artificial intelligence (AI) that teaches computers to process data in a way that is inspired by the human brain. Deep learning models can recognize complex patterns in pictures, text, sounds, and other data to produce accurate insights and predictions.").

[11] *Deepfake Technology*, ORG. FOR SOC. MEDIA SAFETY, https://www.socialmediasafety.org/advocacy/deepfake-technology (last visited Oct. 5, 2024).

[12] *See, e.g., id.*; Laura Payne, *Deepfake*, BRITTANICA (Oct. 1, 2024), https://www.britannica.com/technology/deepfake; Meredith Somers, *Deepfakes, Explained*, MIT MGMT. SLOAN SCH. (July 21, 2020), https://mitsloan.mit.edu/ideas-made-to-matter/deepfakes-explained.

[13] Payne, *supra* note 12.

The typical process of creating a deepfake face-swapping video is done in several steps.[14] First, "thousands[15] of face shots of the two people [must be] run through an AI algorithm called an encoder. The encoder finds and learns similarities between the two faces, and reduces them to their shared common features, compressing the images in the process."[16] This process essentially creates a "lower dimensional representation" of each face, which at this point, "might not look like . . . face[s] at all."[17] Next, each image passes through a second algorithm called a decoder.[18] Normally during this step, each decoder reconstructs the original image it was given, making the images look more like real faces again.[19] Face swapping, on the other hand, takes place when the encoded images are fed into the opposite decoder (i.e., when "a compressed image of person A's face is fed into the decoder trained on person B . . . [so that] the face of person B [is reconstructed] with the expressions and orientation of [person] A").[20] The result is a forged but realistic-looking face.

Face-swapping is not the only way false images can be made, though—generative AI also enables users to create images with only a single photo or even from scratch.[21] By using the same AI systems as face-swapping programs, people can "animate one or several photos of people by first training an AI system on a dataset of videos including many celebrities, so it could learn about key points on the face."[22]

Even without a reference image, people can use generative AI to make new images.[23] To create images from scratch, "machine learning model[s] scan[] millions of images across the internet along with the text associated with them."[24] The algorithms are able to "spot trends in the images and text and eventually guess which

[14] *See* Ian Sample, *What Are Deepfakes – And How Can You Spot Them?*, THE GUARDIAN (Mar. 3, 2023), https://www.theguardian.com/technology/2020/jan/13/what-are-deepfakes-and-how-can-you-spot-them. The author explains that deepfakes can also be created through generative adversarial networks, or GANs. The author describes: "A G[AN] pits two artificial intelligence algorithms against each other. The first algorithm, known as the generator, is fed random noise and turns it into an image. This synthetic image is then added to a stream of real images—of celebrities, say—that are fed into the second algorithm, known as the discriminator." Although the images will not look like real faces at first, repeating the process eventually results in extremely realistic faces.

[15] In the past, thousands of images used to be required to create a deepfake, but now not as many are required. *See* Stuart A. Thompson, *Making Deepfakes Gets Cheaper and Easier Thanks to A.I.*, N.Y. TIMES (Mar. 12, 2023), https://www.nytimes.com/2023/03/12/technology/deepfakes-cheapfakes-videos-ai.html.

[16] Sample, *supra* note 14.

[17] Alan Zucconi, *Understanding the Technology Behind DeepFakes*, ALAN ZUCCONI (Mar. 14, 2018), https://www.alanzucconi.com/2018/03/14/understanding-the-technology-behind-deepfakes.

[18] *See, e.g.,* Sample, *supra* note 14; Zucconi, *supra* note 17.

[19] *Id.*

[20] Sample, *supra* note 14.

[21] *See, e.g.,* Sarah Morrison, *How Unbelievably Realistic Fake Images Could Take Over the Internet*, VOX (Mar. 30, 2023, 4:30 AM), https://www.vox.com/technology/2023/3/30/23662292/ai-image-dalle-openai-midjourney-pope-jacket; Rachel Metz, *Researchers Can Now Use AI and Make Fake Videos of Anyone*, CNN: BUSINESS (May 24, 2019, 7:40 PM), https://www.cnn.com/2019/05/24/tech/deepfake-ai-one-photo/index.html.

[22] Metz, *supra* note 21.

[23] *See Free Online Image Generator*, CANVA, https://www.canva.com/ai-image-generator (last visited Nov. 5, 2024).

[24] *Id.*

image and text fit together."[25] After the model learns what a prompt's given image should look like, it can generate a new image.[26]

Because images created using generative AI are so similar to forged images created by face-swapping, and because legislation tends to refer to deepfakes and AI-generated images interchangeably,[27] this Note refers to both types of images as deepfakes.

B. The Dangers of Deepfakes

Although deepfakes can be used for fun,[28] they can—and often are—used for much more nefarious purposes.[29] They are frequently used in scamming schemes (through replicating a person's voice or image to convince those they know or work with to transfer money)[30] or to otherwise distribute misinformation, especially

[25] *Id.*

[26] *Id.*

[27] *See* TAKE IT DOWN Act, S. 4569, 188th Congress (2024).

[28] *See generally* Rob Cover, *Celebrity Deepfakes Are All Over TikTok. Here's Why They're Becoming Common – And How You Can Spot Them*, THE CONVERSATION (July 18, 2022, 4:05 PM), https://theconversation.com/celebrity-deepfakes-are-all-over-tiktok-heres-why-theyre-becoming-common-and-how-you-can-spot-them-187079 (stating that deepfakes have been used in "silly videos featuring actors such as Robert Pattinson[,] . . . Keanu Reeves[, and Tom Cruise]"); Simon Ellery, *Fake Photos of Pope Francis in a Puffer Jacket Go Viral, Highlighting the Power and Peril of AI*, CBS NEWS (Mar. 28, 2023, 11:39 AM), https://www.cbsnews.com/news/pope-francis-puffer-jacket-fake-photos-deepfake-power-peril-of-ai (noting a deepfake image of Pope Francis wearing "a stylish white puffer jacket and silver bejewelled crucifix" that went viral on social media).

[29] *See generally* Sophie Compton & Reuben Hamlyn, *Opinion: The Rise of Deepfake Pornography Is Devastating for Women*, CNN (Oct. 29, 2023, 12:07 PM), https://www.cnn.com/2023/10/29/opinions/deepfake-pornography-thriving-business-compton-hamlyn/index.html (explaining that the majority of deepfake videos are pornographic); Kat Tenbarge, *Found Through Google, Bought With Visa and Mastercard: Inside the Deepfake Porn Economy*, NBC NEWS (Mar. 27, 2023, 9:56 AM), https://www.nbcnews.com/tech/internet/deepfake-porn-ai-mr-deep-fake-economy-google-visa-mastercard-download-rcna75071 (same); Stuart A. Thompson, *How 'Deepfake Elon Musk' Became the Internet's Biggest Scammer*, N.Y. Times (Aug. 14, 2024), https://www.nytimes.com/interactive/2024/08/14/technology/elon-musk-ai-deepfake-scam.html (explaining how a deepfake video of Elon Musk was used in a scam); Dylan Butts, *Deepfake Scams Have Robbed Companies of Millions. Experts Warn It Could Get Worse*, CNBC (May 27, 2024, 10:20 PM), https://www.cnbc.com/2024/05/28/deepfake-scams-have-looted-millions-experts-warn-it-could-get-worse.html (stating that "deepfake scans ha[ve] looted millions of dollars from companies worldwide"); Alexei Alexis, *Deepfake Scams Escalate, Hitting 53% of Businesses*, CFO DIVE (Sept. 3, 2024), https://www.cfodive.com/news/deepfake-scams-escalate-hitting-53-percent-of-businesses/725836 (explaining that most businesses have been targets of deepfake scams); Heather Chen & Kathleen Magramo, *Finance Worker Pays Out $25 Million After Call with Deepfake 'Chief Financial Officer'*, CNN: WORLD (Feb. 4, 2024, 2:31 AM), https://www.cnn.com/2024/02/04/asia/deepfake-cfo-scam-hong-kong-intl-hnk/index.html (describing how a realistic deepfake led an employee to pay millions to a scammer).

[30] *See generally* Butts, *supra* note 28 (mentioning that "the chief executive officer of a British energy provider reportedly transferred £220,000 ($238,000) to a scammer who had digitally mimicked the head of his parent company and asked for a wire to a supposed supplier on a phone call"); Thompson, *supra* note 28 (explaining that Elon Musk "was featured in nearly a quarter of all deepfake scams since late [2023]"); Chen & Magramo, *supra* note 28 (describing an incident where "[a] finance worker at a multinational firm was tricked into paying out $25 million to fraudsters using deepfake technology to pose as the company's chief financial officer in a video conference call"). *See also* Alexis, *supra* note 28 (explaining that 53% of U.S. and U.K. businesses "have been targets of a financial scam powered by 'deepfake' technology, with 43% falling victim to such attacks").

politically.[31] But the overwhelming majority of deepfakes are used to create pornographic videos—according to a company that monitors AI-developed content, a whopping "96% of deepfakes are sexually explicit and feature women who didn't consent to the videos."[32]

Deepfakes are especially dangerous because the videos can "cost as little as $10 to create" and can be made "in less than [ten] minutes."[33] Less realistic deepfake videos can even be created "for free in less than [thirty] seconds."[34] A quick Google search for deepfake video creation apps returns several popular options, with only some prohibiting users from creating pornographic images of others. People seeking to create deepfake pornography who feel it may be too risky to use these websites or apps on their own may also turn to more experienced deepfake creators advertising their services on platforms like Discord,[35] Reddit,[36] and Telegram[37] to create the videos for them.

[31] *See generally* Em Steck & Andrew Kaczynski, *Fake Joe Biden Robocall Urges New Hampshire Voters Not to Vote in Tuesday's Democratic Primary*, CNN: POLITICS (Jan. 22, 2024, 5:44 PM), https://www.cnn.com/2024/01/22/politics/fake-joe-biden-robocall/index.html (detailing an incident of a fake robocall using Joe Biden's voice to tell voters not to vote in a Democratic primary); Matt Brown & David Klepper, *Fake Images Made to Show Trump With Black Supporters Highlight Concerns Around AI and Elections*, AP NEWS (Mar. 7, 2024, 10:09 PM), https://apnews.com/article/deepfake-trump-ai-biden-tiktok-72194f59823037391b3888a1720ba7c2 (explaining how fabricated images of "Donald Trump surrounded by groups of Black people smiling and laughing" sought to influence the voting behaviors of Black voters). In fact, so many deepfakes have been used to distribute political misinformation that scholars at Purdue University have created a database, called the Political Deepfakes Incidents Database (PDID), to track them. Andrea Azzo, *Tracking Political Deepfakes: New Database Aims to Inform, Inspire Policy Solutions*, NW.: CTR. FOR ADVANCING SAFETY OF MACH. INTEL. (Jan. 26, 2024), https://casmi.northwestern.edu/news/articles/2024/tracking-political-deepfakes-new-database-aims-to-inform-inspire-policy-solutions.html.

[32] Sophie Compton & Reuben Hamlyn, *Opinion: The Rise of Deepfake Pornography Is Devastating for Women*, CNN (Oct. 29, 2023, 12:07 PM), https://www.cnn.com/2023/10/29/opinions/deepfake-pornography-thriving-business-compton-hamlyn/index.html; Kat Tenbarge, *Found Through Google, Bought With Visa and Mastercard: Inside the Deepfake Porn Economy*, NBC NEWS (Mar. 27, 2023, 9:56 AM), https://www.nbcnews.com/tech/internet/deepfake-porn-ai-mr-deep-fake-economy-google-visa-mastercard-download-rcna75071.

[33] Thompson, *supra* note 28.

[34] Lutz Finger, *Overview of How to Create Deepfakes – It's Scarily Simple*, FORBES (Sept. 8, 2022, 10:09 AM), https://www.forbes.com/sites/lutzfinger/2022/09/08/overview-of-how-to-create-deepfakesits-scarily-simple.

[35] *See generally* Kat Tenbarge, *Found Through Google, Bought With Visa and Mastercard: Inside the Deepfake Porn Economy*, NBC NEWS (Mar. 27, 2023, 9:56 AM), https://www.nbcnews.com/tech/internet/deepfake-porn-ai-mr-deep-fake-economy-google-visa-mastercard-download-rcna75071 (explaining that one Discord user advertised that he could create a five-minute long "deepfake of a 'personal girl,' meaning anyone with fewer than [two] million Instagram followers, for $65.").

[36] Deepfakes themselves originated from a Reddit community, as discussed in Part I(A). *See, e.g., Deepfake Technology, Org. for Soc. Media Safety, https://www.socialmediasafety.org/advocacy/deepfake-technology (last visited Oct. 5, 2024); Laura Payne, Deepfake, Brittanica (Oct. 1, 2024), https://www.britannica.com/technology/deepfake;* Meredith Somers, *Deepfakes, Explained*, MIT MGMT. SLOAN SCH. (July 21, 2020), https://mitsloan.mit.edu/ideas-made-to-matter/deepfakes-explained.

[37] *See generally* Bill Goodwin, *Deep Fake AI Services on Telegram Pose Risk for Elections*, COMPUTERWEEKLY.COM (Mar. 18, 2024, 12:00 PM), https://www.computerweekly.com/news/366574113/Deep-fake-AI-services-on-Telegram-pose-risk-for-elections (explaining that "[s]ecurity analysts have identified more than 400 channels promoting

Many deepfake videos feature celebrities, because they are well-known and because there are hundreds—if not, thousands—of readily available photos that can be used to generate the deepfake video.[38] However, because deepfake technology has become more advanced over the years and no longer requires such large datasets,[39] deepfake pornography also now commonly features non-celebrities.[40]

In fact, "[w]ith just a single good image of a person's face, it is now possible . . . to make a 60-second [pornographic] video of that person."[41] Unsurprisingly, because most people have social media accounts with plenty of high-quality images of themselves publicly available, many deepfake creators get the images they use from their victims' social media profiles or the social media profiles of others[42]—yet many people are likely unaware of the potential dangers associated with sharing their images (or images of their children) publicly online.[43]

C. How Victims of Deepfake Pornography Are Affected

Deepfake pornography creators are hard to pin down because most of the time, the images and videos are uploaded by anonymous users.[44] As a result, victims can "feel isolated, disconnected, and mistrustful of many people around them [and] are likely to experience poor mental health symptoms like depression and anxiety."[45] In any case, including when the victim knows who created the deepfake,[46] victims

deep fake services on the Telegram Messenger app, ranging from automated bots that help users create deep fake videos to individuals offering to create bespoke fake videos.").

[38] See Ben Dickson, *What Are Deepfakes?*, TECHTALKS (Sept. 4, 2020), https://bdtechtalks.com/2020/09/04/what-is-deepfake, explaining that "[t]he need for large datasets is why most deepfake videos you see target celebrities. You can't create a deepfake of your neighbor unless you have hours of videos of them in different settings."

[39] *See generally* Stuart A. Thompson, *Making Deepfakes Gets Cheaper and Easier Thanks to A.I.*, N.Y. TIMES (Mar. 12, 2023), https://www.nytimes.com/2023/03/12/technology/deepfakes-cheapfakes-videos-ai.html (explaining that the creation of deepfakes "once required elaborate software to put one person's face onto another's[, b]ut now, many of the tools to create them are available to everyday consumers—even on smartphone apps, and often for little to no money."); Andrea Hauser, *Deepfakes Analysis: Amount of Images, Lighting, and Angles*, SCIP, https://www.scip.ch/en/?labs.20181122 (last visited Oct. 6, 2024) (noting that "[the amount of faces plays less of a role than [she] expected. Much more important is the similarity of the [images] in terms of illumination and angles of the faces.").

[40] *See, e.g.,* Vittoria Elliott, *The US Needs Deepfake Porn Laws. These States Are Leading the Way*, WIRED (Sept. 5, 2024, 6:00 AM), https://www.wired.com/story/deepfake-ai-porn-laws.

[41] Nicholas Kristof, *The Online Degradation of Women and Girls That We Meet With a Shrug*, N.Y. TIMES (Mar. 23, 2024), https://www.nytimes.com/2024/03/23/opinion/deepfake-sex-videos.html.

[42] *See, e.g.,* Gina Silva, *Stolen Instagram Pics Used in Deepfake AI Porn: What to Know*, FOX 11 L.A. (Jan. 31, 2024, 8:38 AM), https://www.foxla.com/news/la-women-victims-of-deepfake-ai-porn; Jillian Krasusky, *Someone Might Be Using Your Instagram Stories to Make Deepfake Porn*, MEDIUM (Apr. 13, 2022), https://medium.com/art-of-the-argument/someone-might-be-using-your-instagram-stories-to-make-deepfake-porn-310ff06b6481.

[43] *See generally* Jim Axelrod, Teen victim of AI-generated "deepfake pornography" urges Congress to pass "Take It Down Act", CBS NEWS (Dec. 18, 2024, 7:47 PM), *https://www.cbsnews.com/news/deepfake-pornography-victim-congress* (describing a story of a teen who was shocked when a classmate used an image from her Instagram account to create a forged naked photo of her).

[44] *See* Krasusky, *supra* note 41.

[45] Halle Nelson, *Taylor Swift and the Dangers of Deepfake Pornography*, NAT. SEXUAL VIOLENCE RES. CTR. (Feb. 7, 2024), https://www.nsvrc.org/blogs/feminism/taylor-swift-and-dangers-deepfake-pornography.

[46] Many teenagers who create deepfake pornography do so as a "prank" toward their female friends, acquaintances, or classmates. In these scenarios, the victims usually know who created the deepfake,

experience extreme distress and humiliation.[47] Victims also are likely to experience reputational damage, with some even losing their jobs or missing out on future employment prospects because of the online permanency of the images.[48] Some victims are so humiliated that they attempt suicide or take their own lives because they no longer want to suffer from the horrible and unjust consequences they face.[49]

Because victims of deepfake pornography face such serious and horrific consequences, lawmakers and society at large must address this issue and, where possible, notify potential victims about individuals in their communities who previously created or possessed deepfake pornography.

II. CURRENT REMEDIES

While holding the tech companies liable who permit this content to be spread may seem like an obvious solution, they are currently protected from liability for third-party content posted to their sites under Section 230 of the Communications Decency Act, as discussed below in Part IV(A)(3). As a result, adult victims of deepfake pornography are currently only able to rely on general civil law.[50]

Lawsuit claims could be "filed for defamation, invasion of privacy, [or intentional infliction of] emotional distress."[51] For such a lawsuit to succeed, "victims must prove that the deepfake was false, was made with reckless disregard for the truth, and caused harm to their reputation or finances."[52] If they do so, victims can receive monetary damages or injunctions—ideally, the creator would take the deepfake images off the internet to prevent further harm from occurring.[53]

Beyond civil remedies, creators of deepfake pornography may also face criminal penalties. Victims, however, have no control over prosecution of these crimes, as this responsibility is one that lies with the government.[54]

A. Civil Remedies

since the creator does not try to hide it. *See* Jessica Le Masurier, *'A Global Problem': US Teen Fights Deepfake Porn Targeting Schoolgirls*, FRANCE 24 (Apr. 18, 2024, 1:31 PM), https://www.france24.com/en/tv-shows/focus/20240418-a-global-problem-us-teen-fights-deepfake-porn-targeting-schoolgirls; Hyung-Jin Kim, *In South Korea, Rise of Explicit Deepfakes Wrecks Women's Lives and Deepens Gender Divide*, PBS NEWS (Oct. 3, 2024, 6:55 PM), https://www.pbs.org/newshour/world/in-south-korea-rise-of-explicit-deepfakes-wrecks-womens-lives-and-deepens-gender-divide.

[47] *See* Nelson, *supra* note 45.

[48] Nelson, *supra* note 45. *See, e.g.,* Coralie Kraft, *Trolls Used Her Face to Make Fake Porn. There Was Nothing She Could Do*, N.Y. TIMES (July 31, 2024), https://www.nytimes.com/2024/07/31/magazine/sabrina-javellana-florida-politics-ai-porn.html (describing how "a rising star in local politics" feared her career would be over if she lost "the respect of her older colleagues").

[49] Kim, *supra* note 43.

[50] As discussed below in Part II(B)(1), possessing deepfake child pornography can be prosecuted under criminal law in some states.

[51] *Legal Remedies for Deepfake Victims: Guide*, SCOREDETECT: BLOG (June 29, 2024), https://www.scoredetect.com/blog/posts/legal-remedies-for-deepfake-victims-guide.

[52] *Id.*

[53] *See* Danielle Keats Citron, *Privacy Injunctions*, 71 EMORY L.J. 955, 970–72 (2022).

[54] *Criminal Cases*, U.S. COURTS, https://www.uscourts.gov/about-federal-courts/types-cases/criminal-cases (last visited Nov. 12, 2024).

Civil remedies may include claims for violation of privacy rights, intentional infliction of emotional distress, and defamation. However, as discussed below, none of these options offers an ideal solution, as their requirements are typically difficult to satisfy.[55]

1. Violation of Privacy Rights: False Light

First, victims may sue under violation of privacy rights, which are generally understood as "the right to be let alone."[56] Though most people consider deepfake pornography to be a violation of privacy[57] because it "annihilates victims' sexual privacy and inherently strips [victims] of their humanity,"[58] privacy-based torts may not offer the best solution for victims. Privacy can be invaded in several ways, including "intrusion on seclusion, wrongful appropriation, false light, and public disclosure of private fact."[59] However, given the elements for each, only false light is likely to meet the requirements.

False light torts aim to hold accountable those who portray others in a misleading way.[60] Under this approach, individuals are considered liable for invasion of privacy if (1) "the false light in which the other was placed would be highly offensive to a reasonable person," and (2) "the actor had knowledge of or acted in reckless disregard as to the falsity of the publicized matter and the false light in which the other would be placed."[61]

With deepfake pornography, the first element is likely satisfied, as most would consider such videos to be highly offensive.[62] The second element, however, is much more challenging to meet.[63] Many deepfake videos are low-quality and easily recognized as fake by viewers,[64] which would not place the victim in a false light. However, a creator of high-quality deepfake videos could attempt to avoid liability by simply labeling the video as fake, because with such a disclaimer, "the portrayal cannot be taken seriously as an accurate depiction."[65]

[55] Anne Pechenik Gieseke, *"The New Weapon of Choice": Law's Current Inability to Properly Address Deepfake Pornography*, 73 Vand. L. Rev. 1479 (2020).

[56] Restatement 2d of Torts, § 652A, comment a.

[57] A study indicates that 68% of individuals "would feel shocked and outraged by the violation of someone's privacy and consent in the creation of deepfake pornographic content." *2023 State of Deepfakes*, SECURITY HERO, https://www.securityhero.io/state-of-deepfakes/#:~:text=Between%202022%20%26%202023%2C%20the%20amount,year%20was%20a%20startling%20464%25 (last visited Nov. 9, 2024).

[58] Gieseke, *supra* note 54, at 1483.

[59] *Id.* at 1496; Restatement 2d of Torts, § 652A.

[60] *See* Restatement 2d of Torts, § 652E.

[61] Restatement 2d of Torts, § 652E.

[62] *See, e.g.,* Coralie Kraft, *Trolls Used Her Face to Make Fake Porn. There Was Nothing She Could Do*, N.Y. TIMES (July 31, 2024), https://www.nytimes.com/2024/07/31/magazine/sabrina-javellana-florida-politics-ai-porn.html; Halle Nelson, *Taylor Swift and the Dangers of Deepfake Pornography*, NAT. SEXUAL VIOLENCE RES. CTR. (Feb. 7, 2024), https://www.nsvrc.org/blogs/feminism/taylor-swift-and-dangers-deepfake-pornography.

[63] See Gieseke, *supra* note 54, at 1498.

[64] *See, e.g.,* Gieseke, *supra* note 54, at 1498; Ian Sample, *What Are Deepfakes – And How Can You Spot Them?*, THE GUARDIAN (Mar. 3, 2023), https://www.theguardian.com/technology/2020/jan/13/what-are-deepfakes-and-how-can-you-spot-them.

[65] Gieseke, *supra* note 54, at 1498.

2. Intentional Infliction of Emotional Distress

Victims may also sue under the tort claim of intentional infliction of emotional distress (IIED). An action for IIED can be brought when (1) "[t]he defendant intended to inflict emotional distress or . . . knew or should have known that emotional distress was likely to result from [their] conduct;" (2) "[t]he defendant's conduct was extreme and outrageous;" (3) "[t]he defendant's conduct was the cause of the plaintiff's emotional distress;" and (4) "[t]he emotional distress sustained by the plaintiff was severe."[66]

Although this may seem like a viable option, it is difficult to meet most of these requirements. First, it can be difficult to prove that the creator intended, knew, or should have known that emotional distress was likely to result from their conduct, as many deepfakes are created for private use[67] or are created to merely "prank" the victim.[68]

Next, the conduct must be extreme and outrageous. Conduct meets this requirement when it is "so outrageous in character, and so extreme in degree, as to go beyond all possible bounds of decency, and to be regarded as atrocious, and utterly intolerable in a civilized community."[69] This element is likely met because a reasonable person would consider deepfake pornography to go beyond the bounds of decency.[70]

However, it may be difficult to prove the final element: that the emotional distress the victim experiences is severe. The definition of severe depends on the jurisdiction, but many consider emotional distress to be severe when "the distress inflicted is so extreme that no reasonable [person] could be expected to endure it without undergoing unreasonable suffering."[71] Certainly many victims of deepfake pornography experience this unreasonable level of suffering,[72] but for victims who are merely embarrassed and want the deepfake taken down, this element would not be met.

3. Defamation

Defamation may be the most effective option for victims of deepfake pornography, but it still presents a challenging standard to meet.[73] For a victim to sue for defamation, they must prove that (1) "[t]he defendant made a false and defamatory statement concerning the plaintiff;" (2) "[t]he statement was published to a third party;"

[66] 1 Jury Instructions on Damages in Tort Actions § 7.09.

[67] The Yatterbog, *Deepfakes: Public vs Personal Use*, MEDIUM (Dec. 1, 2023), https://medium.com/@yatterbog/deepfakes-public-vs-personal-use-b48e55bff745.

[68] *See generally* Hyung-Jin Kim, *In South Korea, Rise of Explicit Deepfakes Wrecks Women's Lives and Deepens Gender Divide*, PBS NEWS (Oct. 3, 2024, 6:55 PM), https://www.pbs.org/newshour/world/in-south-korea-rise-of-explicit-deepfakes-wrecks-womens-lives-and-deepens-gender-divide (stating that "[m]ost suspected perpetrators [of creating deepfake pornography] in South Korea are teenage boys . . . [who] target female friends, relatives and acquaintances—also mostly minors—as a prank, out of curiosity or misogyny.").

[69] Restatement 2d of Torts, § 46, comment d.

[70] *See* Gieseke, *supra* note 54 at 1499.

[71] Tidelands Auto. Club v. Walters, 699 S.W.2d 939, 941 (Tex. App. 1985).

[72] *See* Halle Nelson, *Taylor Swift and the Dangers of Deepfake Pornography*, NAT. SEXUAL VIOLENCE RES. CTR. (Feb. 7, 2024), https://www.nsvrc.org/blogs/feminism/taylor-swift-and-dangers-deepfake-pornography.

[73] Emma Grey Ellis, *People Can Put Your Face on Porn—and the Law Can't Help You*, WIRED (Jan. 26, 2018, 7:00 AM), https://www.wired.com/story/face-swap-porn-legal-limbo.

(3) [t]here was . . . negligence, intent, or actual malice on the part of the defendant;" and (4) "[t]he plaintiff suffered harm to [their] reputation."[74]

For a statement to be considered defamatory, the statement must "harm the reputation of another as to lower him in the estimation of the community or to deter third persons from associating or dealing with him."[75] Embarrassing or frustrating the plaintiff is not enough—the statement "must provoke the kind of harm which has grievously fractured [one's] standing in the community of respectable society."[76]

When deepfake pornography is shared with others, the first two requirements are typically met: the content of the video is considered a false statement, and posting or otherwise sharing it fulfills the publication requirement.[77] Not every jurisdiction treats the publication requirement the same, but most require the statement to be seen or heard by a third party. For example, New Mexico considers publication to be "an intentional or negligent communication to one other than the person defamed,"[78] meaning that even if the content is not posted publicly but rather shared privately to someone else, it could still meet the publication requirement. Because deepfake pornography can be distributed via text messages or social media chat services like Snapchat and can easily be displayed to others in person,[79] the publication requirement would likely be met.

However, requiring the producer to have negligence, intent, or actual malice demonstrates a limitation of this approach. No uniform standard for each of these terms exists, and definitions may differ from one jurisdiction to another.[80] Generally, negligence is defined as "the failure to do what a reasonable and prudent person would ordinarily have done under the circumstances of the situation."[81] Intent refers to "the purpose formed in [one's] mind,"[82] while some jurisdictions define actual malice as acting with "spite and ill will . . . with a design willfully or wantonly to injure another."[83] While the standard for negligence may be achievable, proving intent or actual malice would be challenging to prove, because as one scholar notes, "many producers have no idea the victim will ever discover the video [and] neither intend emotional distress nor reasonably know that their deepfake winds up in the victim's hands."[84]

[74] 1 Jury Instructions on Damages in Tort Actions § 12.03A.

[75] Graboff v. Colleran Firm, 744 F.3d 128, 136 (2014) (quoting Tucker v. Fishchbein, 237 F.3d 275, 282 (2001)).

[76] *Id.* (quoting Tucker v. Phila. Daily News, 577 Pa. 598, 615 (2004)).

[77] Anne Pechenik Gieseke, *"The New Weapon of Choice": Law's Current Inability to Properly Address Deepfake Pornography*, 73 Vand. L. Rev. 1479, 1500 (2020).

[78] 13-1003 NMRA.

[79] Jessica Le Masurier, *'A Global Problem': US Teen Fights Deepfake Porn Targeting Schoolgirls*, FRANCE 24 (Apr. 18, 2024, 1:31 PM), https://www.france24.com/en/tv-shows/focus/20240418-a-global-problem-us-teen-fights-deepfake-porn-targeting-schoolgirls.

[80] 1 Jury Instructions on Damages in Tort Actions § 12.03A, comment 3.

[81] 13B M.J. Negligence § 2 (2024).

[82] 9B M.J. Homicide § 34 (2024).

[83] Heuer v. John R. Thompson Co., 251 S.W.2d 980, 986 (Mo. Ct. App. 1952).

[84] Anne Pechenik Gieseke, *"The New Weapon of Choice": Law's Current Inability to Properly Address Deepfake Pornography*, 73 Vand. L. Rev. 1479, 1500 (2020).

Last, plaintiffs must suffer harm to their reputation.[85] While many victims face reputational harm—such as when deepfake pornography appears in internet searches associated with their name—some victims might not.[86] For example, minors who are sexually promiscuous or who have taken consensual nude photographs of themselves may be denied relief because similar images of them already exist, perhaps leading to the perception that their reputation has not been harmed.[87]

4. Issues with Current Available Civil Remedies for Victims

Overall, litigation is an unfair solution for victims of deepfake pornography. Although there are countless victims of deepfake pornography, only two civil lawsuits have been filed to date over deepfake pornography as of this writing—and one was brought by San Francisco's City Attorney, not by a victim directly.[88] Part of this disparity can be attributed to the likelihood that many potential cases are filtered out for the reasons mentioned above. However, even when all elements of a given civil remedy can be established, civil lawsuits likely remain rare due to the many barriers to litigation.

One barrier is that litigation is notoriously time-consuming and expensive.[89] When online content is the subject of a legal dispute, attorneys or digital forensics experts must be consulted,[90] which only increases these costs. Many people, including minors who rely on their parents, have neither the resources nor the time to pursue litigation. Moreover, civil litigation typically compensates victims monetarily, as injunctions are rarely granted—but most victims just want the deepfake content removed from the internet and deleted permanently.

Apart from these issues, civil litigation does not provide the penalties criminal law does. Litigation is time-consuming and costly for defendants, too, but that alone does little to deter perpetrators from publishing deepfake pornography of others in the future, given how infrequently victims pursue litigation. Although involvement in civil lawsuits must be disclosed in situations like background checks and certain insurance or credit applications, this level of disclosure does not carry the same public awareness as a criminal charge.

[85] 1 Jury Instructions on Damages in Tort Actions § 12.03A.

[86] *See* Sophie Compton & Reuben Hamlyn, *Opinion: The Rise of Deepfake Pornography is Devastating for Women*, CNN (Oct. 29, 2023, 12:07 PM), https://www.cnn.com/2023/10/29/opinions/deepfake-pornography-thriving-business-compton-hamlyn/index.html.

[87] *See* Camille Sojit Pejcha, *Deepfake Porn Isn't Just a Consent Issue, It's a Labor Issue*, DOCUMENT (Feb. 2, 2023), https://www.documentjournal.com/2023/02/twitch-streamer-deepfake-controversy-ai-porn-sex-work-labor-technology (explaining that some may think "sex worker[s] [are] essentially considered to be 'asking for it' by participating in sex work").

[88] This lawsuit seeks to permanently shut down sixteen popular websites that create AI pornography of women. Isaiah Poritz, *San Francisco Files Nation's First Suit Over AI Pornography*, BLOOMBERG LAW (Aug. 15, 2024, 11:50 AM), https://news.bloomberglaw.com/litigation/san-francisco-files-nations-first-suit-over-ai-generated-porn. The only other lawsuit that has been filed is Francesca Mani's case mentioned in Part I.

[89] Sam Bock, *4 Barriers Blocking Access to Justice (and How to Help Break Them)*, RELATIVITY: BLOG (Mar. 25, 2021), https://www.relativity.com/blog/4-barriers-blocking-access-to-justice-and-how-to-break-them.

[90] *Legal Remedies for Deepfake Victims: Guide*, SCOREDETECT: BLOG (June 29, 2024), https://www.scoredetect.com/blog/posts/legal-remedies-for-deepfake-victims-guide.

While civil litigation can be a viable option for victims who manage to overcome these barriers, it is evident that the existing civil remedies available to victims do not fully address the problem.

B. Criminal Penalties

In addition to civil laws, criminal laws may penalize those who create or distribute deepfake pornography under certain circumstances. While there are no federal criminal laws specifically prohibiting deepfake pornography,[91] certain states have created their own laws—typically modeled after their legal treatment of similar crimes—or have "expanded existing crimes to cover these acts."[92]

1. State Deepfake Pornography Laws

Many states have recognized the dangers of deepfakes and have either initiated or enacted laws to address the problem, while others have attempted to enact legislation but have not succeeded.[93] These state laws differ widely in their definitions of key terms, overall regulations, and associated penalties.

First, key terms are defined differently. To illustrate, Mississippi defines "intimate part" as "the naked genitals, pubic area, anus[,] or female nipple of the person,"[94] while Illinois expands this definition to include parts that are "fully unclothed, partially unclothed, or transparently clothed" and "partially or fully exposed."[95] Illinois includes the same intimate parts as Mississippi, but it limits female nipples to only include "post-pubescent nipple[s]," which poses an issue for images depicting minors.[96]

Overall regulations also vary widely. For example, Alabama criminalizes creating an AI image if the person "knowingly creates, records, or alters a private image when the depicted individual has not consented to the creation, recording, or alteration and the depicted individual had a reasonable expectation of privacy."[97] New York similarly criminalizes "dissemination or publication of an intimate image," but adds more requirements—the perpetrator must "inten[d] to cause harm to the emotional, financial or physical welfare of another" and the image must depict "one or more intimate parts" of a person or someone "engaging in sexual conduct with another."[98]

Some states have also enacted laws specifically criminalizing deepfake child pornography. For instance, South Dakota recently enacted a law providing that "a person is guilty of possessing child pornography if the person knowingly possesses any

[91] Rebecca Pirius, *Is Deepfake Pornography Illegal?*, CRIM. DEF. LAW. (Sept. 26, 2024), https://www.criminaldefenselawyer.com/resources/is-deepfake-pornography-illegal.html.
[92] *Id.*
[93] *Deceptive Audio or Visual Media ('Deepfakes') 2024 Legislation*, NCSL (Oct. 10, 2024), https://www.ncsl.org/technology-and-communication/deceptive-audio-or-visual-media-deepfakes-2024-legislation#:~:text=These%20states%20are%20California%2C%20Connecticut,at%20the%20use%20of%20deepfakes.
[94] 2024 MS Senate Bill 2288.
[95] 2023 IL Senate Bill 382.
[96] *Id.*
[97] NCSL, *supra* note 92; Alabama Code Title 13A. Criminal Code § 13A-6-240.
[98] NCSL, *supra* note 92; 2023 NY Senate Bill 1042.

visual depiction of a minor engaging in a prohibited sexual act, or in a simulation of a prohibited sexual act, or any computer-generated child pornography."[99]

Each state also punishes these crimes differently. For example, Alabama's law described above classifies this offense as a Class A misdemeanor,[100] but South Dakota's law categorizes it as a Class 4 felony.[101]

States with laws penalizing deepfake pornography are examples of good progress, but addressing this issue solely at the state level presents several issues. As described above, state laws vary widely, with some states providing more protections to minors than others. States may define terms differently and impose different rules with different penalties. Furthermore, because online content can easily cross state lines, jurisdictional complications often arise.

It may seem unusual that more serious crimes, such as murder or rape, are typically handled only on the state level rather than under federal law.[102] However, these charges are "prosecuted as state crimes because the allegations within the charge violate state law," and only fall under federal jurisdiction in rare scenarios.[103] However, because online content can easily cross state lines, jurisdictional complications could easily arise with deepfake pornography, demonstrating a need for federal regulation.

2. Legal Treatment of and Issues with Similar Crimes

When states create their own laws, they are often modeled after their laws criminalizing revenge porn or child sexual abuse material.

a. Revenge Porn Laws

Revenge pornography—a tort commonly referred to as "revenge porn"—is "the intentional distribution of non-consensual porn [that] occurs when an ex-partner, hacker, or others post sexually explicit images of a person online without permission."[104] To date, forty-nine states, the District of Columbia, and Guam have statutes against revenge porn.[105]

Using revenge porn laws as a framework to create new laws about deepfakes is a good starting point but poses several issues. First, each state has different revenge

[99] NCSL, *supra* note 92; 2024 SD Senate Bill 79.

[100] NCSL, *supra* note 92; Alabama Code Title 13A. Criminal Code § 13A-6-240.

[101] NCSL, *supra* note 92; 2024 SD Senate Bill 79.

[102] Neil Shouse, *7 Situations Where "Murder" Is a Federal Crime*, SHOUSE CAL. LAW GRP. (Mar. 7, 2024), https://www.shouselaw.com/ca/blog/murder/is-murder-a-federal-crime-7-ways-it-can-be/#:~:text=Many%20murder%20charges%20are%20prosecuted,is%20handled%20by%20federal%20prosecutors.

[103] For example, murder is considered a federal crime when the person murdered is "a federal judge or . . . federal law enforcement official," "an immediate family member of a federal law enforcement official," or "an elected or appointed federal official." Murder can also be considered a federal crime when "the killing is committed during a bank robbery," "takes place on federal property" or "aboard a ship at sea," or if the murder "was designed to influence a court case." Shouse, *supra* note 101.

[104] 1 Punitive Damages § 9.28.

[105] *Nonconsensual Distribution of Intimate Images*, CYBER C.R. INITIATIVE, https://cybercivilrights.org/nonconsensual-distribution-of-intimate-images (last visited Nov. 16, 2024); Anne Pechenik Gieseke, *"The New Weapon of Choice": Law's Current Inability to Properly Address Deepfake Pornography*, 73 VAND. L. REV. 1479, 1501 (2020).

porn requirements and punishments. For example, some jurisdictions require "intent to cause emotional harm,"[106] others require "malicious intent when distributing the images,"[107] with others having no intent requirement at all.[108] The offense can also be classified as either a misdemeanor or a felony, depending on the jurisdiction.[109] If each state were to model their deepfake pornography laws after their existing revenge porn laws, the lack of consistency between states would cause problems, given the online nature of these crimes and their ability to transcend state lines.

Most importantly, revenge porn laws are based on principles of privacy law, but deepfakes may not be considered a violation of privacy. This is because the content may be easily recognizable as fake and the images may not be "taken seriously as an accurate depiction,"[110] as discussed above in Section II(A)(1). As one scholar notes, "[d]eepfakes are not fully 'real' in that they depict an act that never actually happened," and are rather based on photos that the victim has publicly posted online.[111] Thus, deepfakes are "not—legally speaking—a privacy violation."[112] Deepfakes, she notes, occupy a middle ground somewhere between real and fake, necessitating a new approach to account for this unique challenge.[113]

b. Child Sexual Abuse Material

For deepfake pornography depicting minors, some states have modeled their legislation after their existing laws for child sexual abuse material (CSAM). For example, Utah recently amended their CSAM laws to add that "[a]n actor is guilty of an offense if they 'commit the offense with the aid of a generative artificial intelligence' or 'intentionally promote[] or otherwise cause[] a generative artificial intelligence to commit the offense.'"[114]

Although deepfakes of minors can currently be prosecuted under CSAM statutes, this does not happen as frequently as it should due to various challenges. First, identifying perpetrators may be difficult. Although deepfake pornography creators sometimes reveal themselves to the victims—mostly in the case of teens creating deepfake porn of their friends—much of the time, the creators remain anonymous. Experts note that "[p]erpetrators can use various tools and techniques to mask their identities, making it challenging for law enforcement to track them down."[115]

Even when perpetrators can be identified, jurisdictional challenges may arise. As mentioned above, the online nature of these crimes makes it difficult for only one state's law to govern, as these cases may involve several states or even be distributed

[106] Katherine Gabriel, *Feminist Revenge: Seeking Justice for Victims of Nonconsensual Pornography Through "Revenge Porn" Reform*, 44 VT. L. REV. 849, 870 (2020).
[107] *Id.* at 869–70.
[108] *Id.* at 869.
[109] *Id.* at 868.
[110] Gieseke, *supra* note 104, at 1498.
[111] *Id.* at 1501–02.
[112] *Id.*
[113] *See id.*
[114] Lacey Johnson & John Feinauer, *Deepfakes, AI, and Intimate Images*, UTAH STATE LEG. 4 (Aug. 21, 2024), https://le.utah.gov/interim/2024/pdf/00003098.pdf.
[115] Samuel Dordulian, *Which States Have Passed Deepfake Laws?*, DORDULIAN LAW GRP. (Sept. 5, 2024), https://www.dlawgroup.com/states-have-passed-deepfake-porn-laws.

internationally.[116] Thus, it is essential that a federal law be enacted to create consistency across the board.

III. CONSTITUTIONAL FIRST AMENDMENT ISSUES

When creating laws, it is essential to understand constitutional barriers that must be complied with—most notably, First Amendment concerns—as well as their exceptions. The First Amendment states that "Congress shall make no law . . . abridging the freedom of speech."[117] Speech includes creating pictures and videos, so passing laws regulating deepfake pornography must fall under guidelines set forth by the Supreme Court to remain constitutional.

In order to pass a law that relates to deepfake pornography, the government can either "construct a narrowly tailored law that fits within the confines of the First Amendment," or they must "regulate deepfake pornography under existing categories of unprotected speech," such as obscenity or child pornography.[118]

A. Obscenity

Material is considered to be obscene, and thus unprotected by the First Amendment, when (1) "the average person . . . would find that the work, taken as a whole, appeals to the prurient interest"; (2) "the work depicts or describes, in a patently offensive way, sexual conduct specifically defined by the applicable state law;" and (3) "the work, taken as a whole, lacks serious literary, artistic, political, or scientific value."[119]

Obscenity could be applicable deepfake child pornography, but meeting its requirements can still be challenging. First, although many deepfakes would appeal to the prurient interest or be patently offensive, others may lack overtly sexual or offensive content yet still be embarrassing for the victim. For example, imposing someone's face onto a nude image from a medical textbook would not be sexual in nature but could still be damaging to the victim. Furthermore, even deepfake pornography—though likely not involving images of children—may be found to have artistic value, complicating obscenity classification.

B. Child Pornography

Child pornography is another unprotected category of speech. This category first began to be recognized in the late 1900s. In 1996, Congress passed the Child Pornography Prevention Act of 1996 (CPPA), banning "content that 'appears to be' child pornography but produced by means other than using real children, such as through the use of youthful-looking adult actors or computer-imaging technology."[120]

[116] *Id.*

[117] U.S. Const. amend. I.

[118] Emily Pascale, *Deeply Humanizing, Degrading, and Violating: Deepfake Pornography and the Path to Legal Recourse*, 73 SYRACUSE L. REV. 335, 342 (2023).

[119] Miller v. California, 413 U.S. 15, 24 (1973).

[120] Krista L. Baughman, *Will SCOTUS Reconsider Virtual Child Porn Laws in Light of Deepfake Culture?*, DAILY JOURNAL (Feb. 8, 2024), https://www.dailyjournal.com/articles/377089-will-scotus-reconsider-virtual-child-porn-laws-in-light-of-deepfake-culture.

In 2002, however, the Supreme Court introduced "limits on what can qualify as child pornography in *Ashcroft v. Free Speech Coalition*,"[121] and decided that the "appears to be" language used in the CPPA was overly broad, as it extended beyond obscenity and could ban content that had "redeeming artistic value," such as Shakespeare's *Romeo and Juliet*.[122] Thus, the Court "concluded that the statute was unconstitutional, explaining that virtual child pornography fell outside the constitutional category of child pornography."[123]

By doing so, the Court decided that "protecting *future* victims of child sex abuse [was not] a sufficient government interest," and clarified that "the harm of creation—that is, the sexual exploitation and abuse of children to produce child pornography—is the [basis] of . . . child pornography doctrine[,] . . . not what [the image] communicated."[124] However, the Court in *Ashcroft* left open the question of "whether images depicting real children, but created without sexual molestation or exploitation, are sufficiently similar to real child pornography to be exempt from First Amendment protection."[125]

Since then, "most courts have decided that images depicting real children, but created without sexual molestation or exploitation, still qualify as child pornography," and are thus included in this carve-out of the First Amendment.[126] Most notably, in 2009, the PROTECT Act was passed, defining "child pornography" to include "any computer-generated image that is, or is indistinguishable from, that of a minor engaging in sexually explicit conduct," or images that "ha[ve] been created, adapted, or modified to appear that an identifiable minor is engaging in sexually explicit conduct."[127] However, scholars argue that this Act's provisions may be unconstitutional, as it "criminalizes all materials that are indistinguishable from child pornography," and may "grant[] prosecutors too much power to determine whom to charge." [128]

IV. POTENTIAL SOLUTIONS

Recognizing the challenges posed by deepfake pornography, lawmakers and companies alike have taken steps to address the issue, with some social media companies updating their privacy policies, and lawmakers proposing federal legislation.

A. Holding Tech Companies Liable

When determining liability, responsibility could fall on one of two parties: developers or deployers.[129] Developers are those who create the software, and

[121] Carissa Byrne Hessick, *The Expansion of Child Pornography Law*, 21 NEW CRIM. L. REV. 1, 3 (2019).
[122] Baughman, *supra* note 119; Ashcroft v. Free Speech Coalition, 535 U.S. 234 (2002).
[123] Hessick, *supra* note 120, at 3.
[124] *Id.* at 3–4.
[125] *Id.* at 4.
[126] *Id.*
[127] 18 U.S.C.S. § 2256.
[128] Rosalind E. Bell, *Reconciling the PROTECT Act with the First Amendment*, 87 NYU L. REV. 1878, 1903–04 (2012).
[129] Emory Odom, *Developers vs. Deployers: AI Leaders Demand Role Distinction in U.S. Legislation*, THE NAT'L CIO REV. (Oct. 18, 2024), https://nationalcioreview.com/articles-insights/technology/artificial-intelligence/developers-vs-deployers-ai-leaders-demand-role-distinction-in-u-s-

deployers are those who "apply these systems in the real world," such as using the software in a negative way to create or share deepfake pornography.[130] To hold tech companies liable, it is important to target the right actors, "avoiding unfair burdens on developers while ensuring deployers take responsibility for their usage."[131]

Although AI generation websites, categorized as developers, cannot be regulated due to First Amendment protections, some argue that tech companies themselves should play a part in resolving the issue, whether by having social media sites prohibit deepfake content from being posted or holding search engines accountable for allowing deepfake pornography sites to come up in search results. Such actions could effectively limit deployers in their ability to share harmful content.

1. Social Media Restrictions

Social media platforms may be able to help mitigate the problem through updating their policies on permissible content. For instance, X (formerly Twitter) updated its policies to prevent the spread of misinformation and other false content.[132] However, this approach has limitations: policies may too vaguely worded and may be abused to silence the wrong voices.[133] Additionally, "rely[ing] on private companies to police our societies" may not be very effective and perhaps holds the wrong parties accountable.[134]

2. Amending Search Engine Algorithms

Some even argue that search engine websites—in addition to AI generation websites or the social media platforms deepfakes are shared on—should play a part in the solution.[135] One writer notes that though Google tailors its search results in a positive way in some scenarios—such as displaying a suicide hotline as the first result when someone searches for ways to take their life—it fails to use similar precautions when someone searches for deepfake pornography.[136] While this would be a step in the right direction, encouraging search engines to prevent this kind of content from appearing in search results would not be sufficient to address the issue on its own.

3. Section 230 Immunity

Though passing legislation to hold these websites liable may sound like a good solution, this is not currently possible. Under Section 230 of the Communications Decency Act of 1996, social media platforms or other websites that host this content

legislation/#:~:text=For%20example%2C%20developers%20who%20build,responsibility%20for%20real%2Dworld%20implementations.

[130] *Id.*

[131] *Id.*

[132] Arian Garshi, *Deepfakes in 2022: How Individual Non-Celebrities are Targeted*, MEDIUM (Oct. 17, 2022), https://ariangarshi.medium.com/deepfakes-in-2022-how-individual-non-celebrities-are-targeted-a7dab59cac3a; *Synthetic and Manipulated Media Policy*, X: HELP CTR. (Apr. 2023), https://help.x.com/en/rules-and-policies/manipulated-media.

[133] Garshi, *supra* note 131.

[134] *Id.*

[135] Nicholas Kristof, *The Online Degradation of Women and Girls That We Meet With a Shrug*, WIRED (Mar. 23, 2024), https://www.nytimes.com/2024/03/23/opinion/deepfake-sex-videos.html.

[136] *Id.*

cannot be held liable for allowing it to be spread.[137] This Act aims "to promote the continued development of the Internet," "to preserve the vibrant and competitive free market that presently exists for the Internet," and "to remove disincentives for the development and utilization of blocking and filtering technologies" that parents may use to limit children's access to harmful material.[138] There have been "dozens of [bipartisan] proposals to amend [or repeal] Section 230,"[139] so this obstacle may be eliminated in the future. However, for now, websites are effectively shielded from liability for third-party content its users post to its website.

Although many plaintiffs have challenged Section 230 immunity, only one has been successful thus far. This recent and notable case involved a ten-year-old girl who died doing the "Blackout Challenge," a TikTok trend that encouraged users "to asphyxiate themselves to the point of losing consciousness."[140] While the district court held that Section 230 shielded TikTok from liability, the court of appeals later reversed, clarifying that tech companies are "immunized only if they are sued for someone else's expressive activity or content (i.e., third-party speech), but they are not immunized if they are sued for their own expressive activity or content (i.e., first-party speech)."[141] Though the videos encouraging the challenge were created by third party users, the videos were pushed to the girl through TikTok's algorithm, which "decides on the third-party speech that will be included in or excluded from a compilation—and then organizes and presents the items" on each user's feed.[142] Thus, the appellate court concluded that the algorithm recommending the Blackout Challenge was TikTok's first-party speech and, therefore, was not protected under Section 230.[143]

However, this recent exception to Section 230 applies only to websites that use tailored algorithms to push content to users, meaning that websites that do not use such algorithms—like most sites that currently host deepfake pornography—remain protected.

B. Proposed Federal Criminal Legislation

No federal legislation has been passed yet to regulate deepfake pornography, but several acts have been proposed. Some federal proposed acts provide for civil remedies;[144] however, due to the challenges associated with relying on civil remedies

[137] 47 U.S.C.S. § 230.

[138] 47 U.S.C.S. § 230(b).

[139] *Section 230: A Brief Overview*, CONG. RSCH. SERV. (Feb. 2, 2024), https://crsreports.congress.gov/product/pdf/IF/IF12584.

[140] Gigen Mammoser, *Dangerous Social Media 'Blackout Challenge' Can Cause Brain Damage, Death in Less Than 5 Minutes*, HEALTHLINE (Sept. 9, 2024), https://www.healthline.com/health-news/tiktok-blackout-challenge.

[141] Anderson v. TikTok, Inc., 116 F.4th 180, 183 (3d Cir. 2024).

[142] *Id.* at 184.

[143] *Id.*

[144] The two federal proposed acts that provide victims with civil remedies are the DEFIANCE Act (short for Disrupt Explicit Forced Images and Non-Consensual Edits Act), which was introduced on January 30, 2024, and the No AI FRAUD Act (short for No Artificial Intelligence Fake Replicas and Unauthorized Duplications Act), which was introduced on January 10, 2024.

The DEFIANCE Act allows "victims to sue if those who created the deepfakes knew, or 'recklessly disregarded' that the victim did not consent to its making." Solcyré Burga, *How a New Bill Could Protect Against Deepfakes*, TIME (Jan. 31, 2024, 4:34 PM), https://time.com/6590711/deepfake-protection-federal-bill. To learn more about the DEFIANCE Act, see also Press Release, Rep. Ocasio-Cortez Leads Bipartisan, Bicameral Introduction of DEFIANCE Act to Combat Use of Nonas discussed in Section II(A)(4), these proposals will not be discussed. Rather, this section will focus on the two federal proposed acts that provide victims with criminal remedies: the DEEP FAKES Accountability Act and the TAKE IT DOWN Act.

1. DEEP FAKES Accountability Act

The DEEP FAKES Accountability Act—short for Defending Each and Every Person from False Appearances by Keeping Exploitation Subject to Accountability—was introduced on September 20, 2023, by Representative Yvette Clark (D-NY).[145] This Act seeks "[t]o protect national security against . . . deepfake technology and to provide legal recourse to victims of harmful deepfakes" by requiring a watermark or other textual descriptions on deepfake content.[146] It also requires verbal statements disclosing that the content has been altered if the content contains audio.[147]

This bill criminalizes "the production of [deepfakes] which do not comply with related watermark or disclosure requirements" as well as "the alteration of [deepfakes] to remove or meaningfully obscure such required disclosures," with penalties including "a fine, up to five years in prison, or both."[148]

However, this bill is unlikely to pass, as there are several issues with it. Although the Act would create a taskforce at the Department of Homeland Security to help combat deepfakes, it "would serve more of a research and reporting function" rather than provide a real enforcement mechanism to assist victims.[149] Others also note that the legislation would "burden legitimate users of [deepfake] technology and encumber courts with litigation due to its overbroad definition of 'deep fake.'"[150]

2. TAKE IT DOWN Act

The TAKE IT DOWN Act—an acronym for Tools to Address Known Exploitation by Immobilizing Technological Deepfakes on Websites and Networks—was introduced on June 18, 2024, by Senator Ted Cruz (R-TX).[151] As of this writing, this Act has passed the Senate.[152] This act "criminalize[s] the publication of non-

Consensual, Sexually-Explicit "Deepfake" Media, ALEXANDRIA OCASIO-CORTEZ (Mar. 7, 2024), https://ocasio-cortez.house.gov/media/press-releases/rep-ocasio-cortez-leads-bipartisan-bicameral-introduction-defiance-act-combat; S. 3696, CONGRESS.GOV, https://www.congress.gov/bill/118th-congress/senate-bill/3696.
The No AI Fraud Act "provides for individual property rights in likeness and voice." Evan Harris, *Deepfake Laws: A Comprehensive Overview*, PLURAL POLICY (May 23, 2024), https://pluralpolicy.com/blog/deepfake-laws. The Act seeks to protect "artists in the music industry[] who are encountering direct attacks to their intellectual property from abuse of [AI]," but this Act may also apply to deepfake pornography. Press Release, Salazar Introduces the No AI FRAUD Act, CONGRESSWOMAN MARIA ELVIRA SALAZAR (Jan. 10, 2024), https://salazar.house.gov/media/press-releases/salazar-introduces-no-ai-fraud-act.

[145] DEEP FAKES Accountability Act, H.R. 3230, 116th Congress (2020).

[146] H.R. 3230.

[147] H.R. 3230; Lindsey Joost, *The Place for Illusions: Deepfake Technology and the Challenges of Regulating Unreality*, 33 U. FLA. J.L. & PUB. POL'Y 309, 330 (2023).

[148] H.R. 3230.

[149] Joost, *supra* note, at 330–31.

[150] Zachary Schapiro, *Deep Fakes Accountability Act: Overbroad and Ineffective*, B.C. INTELL. PROP. & TECH. FORUM 1, 1 (2020).

[151] TAKE IT DOWN Act, S. 4569, 188th Congress (2024).

[152] *Id.*

consensual intimate imagery (NCII), including AI-generated NCII (or 'deepfake pornography'), and require[s] social media and similar websites to have in place procedures to remove such content upon notification from a victim."[153] Although the Act only addresses NCIIs (including deepfake pornographic images) that are published—which may arguably be too narrow, as it may leave out unpublished deepfake pornography creations that may deserve penalty—the Act nonetheless is a step in the right direction toward effectively regulating deepfake pornography.

The Act permits victims to "submit a request for the . . . platform to remove [the] intimate visual depiction [of them]."[154] Victims must include in their request their signature (or a signature of their authorized representative), an "identification of the intimate visual depiction" of themselves, and "a brief statement that [they] ha[ve] a good faith belief that [the] intimate visual depiction . . . is not consensual, including any relevant information for the . . . platform to determine the intimate visual depiction was published without [their] consent."[155]

Upon receiving a removal request that meets the elements outlined above, a platform must "remove the intimate visual depiction and make reasonable efforts to remove any identical copies of such depiction as soon as possible, but not later than [forty-eight] hours after receiving [the] request."[156] By requiring websites to act on victim's requests within forty-eight hours, the bill seeks to ensure that "victims are protected from being retraumatized again and again" when this harmful content is posted.[157] If websites do not take the content down within the required time frame, their inaction will be "treated as a violation of a rule defining an unfair or deceptive act or practice under section 18(a)(1)(B) of the Federal Trade Commission Act,"[158] meaning that the website owners would be required to participate in a time-consuming process including an informal hearing.[159]

This bill also "criminalize[s] the publication of such content without the victim's consent"[160] and imposes fines, prison time (up to two years in prison for offenses involving adults and up to three years of imprisonment for offenses involving minors), or both for violations.[161]

Surprisingly, this bill has received no notable criticism, and many hope that it will be quickly enacted into law.[162] This bill addresses key issues in an effective way,

[153] Press Release, Sen. Cruz Leads Colleagues in Unveiling Landmark Bill to Protect Victims of Deepfake Revenge Porn, U.S. SENATE COMM. ON COM., SCI., & TRANSP. (June 18, 2024), https://www.commerce.senate.gov/2024/6/sen-cruz-leads-colleagues-in-unveiling-landmark-bill-to-protect-victims-of-deepfake-revenge-porn.
[154] S. 4569, Section 3(a)(1)(A).
[155] S. 4569, Section 3(a)(1)(B).
[156] S. 4569, Section 3(a)(3).
[157] S. 4569.
[158] S. 4569, Section 3(b)(1).
[159] 15 U.S.C. 57(a).
[160] *Considerations for Federal Right of Publicity as AI Advances*, LAW360: A LEXISNEXIS COMPANY (July 31, 2024).
[161] S. 4569.
[162] *Congress Must Pass TAKE IT DOWN Act*, AOL (July 6, 2024, 9:59 PM), https://www.aol.com/editorial-congress-must-pass-down-035900039.html.

"protect[ing] and empower[ing] victims of real and deepfake NCII, while protecting lawful speech."[163]

This legislation touches on several important points. First, it criminalizes not only the publication of NCII, including deepfake NCII, but also the threat to publish such content in interstate commerce when the intent is to "intimidat[e], coerc[e], extort[], or . . . distress" the victim.[164] Additionally, it "[p]rotects good faith efforts to assist victims" by including exceptions for law enforcement, legal and medical professionals, and others who intend to assist the victim.[165] Moreover, the bill is narrowly tailored to align with First Amendment guidelines, and criminalizes NCII without infringing on lawful speech, such as images used for "legitimate medical, scientific, or education purpose[s]."[166]

V. AMENDING THE TAKE IT DOWN ACT TO MANDATE SEX OFFENDER REGISTRATION

While the TAKE IT DOWN Act thoroughly addresses most issues, it could be improved with simple addition: clarifying that violators who publish deepfake pornography of children should also be categorized as Tier I sex offenders, which carries with it mandatory sex offender registration for ten to fifteen years.

A. Sex Offender Registries

Sex offender registries are meant to help law enforcement "monitor[] and track[] sex offenders following their release into the community" and to alert the public of their existence.[167] Registered sex offenders must appear in person at regular intervals "to take a current photograph and verify . . . where they live, work[,] and go to school."[168] Sex offender registries are managed at the state level, as sex offenders must report to state authorities to update their registration, but there is still a large federal component to them. While there is no federal sex offender registry, the Sex Offender Registration And Notification Act (SORNA)—a federal law—requires "every state to participate in the Dru Sjodin National Sex Offender Public Website, which pulls information from all the states into one searchable database,"[169] allowing people from every state to easily identify potential predators.

[163] *The TAKE IT DOWN Act: Tools to Address Known Exploitation by Immobilizing Technological Deepfakes on Websites and Network*, TODD YOUNG: U.S. SEN. FOR IND., https://www.young.senate.gov/wp-content/uploads/1-pager_TAKE-IT-DOWN-Act_6.18.2024-FINAL.pdf (last visited Nov. 17, 2024).

[164] S. 4569; TODD YOUNG, *supra* note 158.

[165] TODD YOUNG, *supra* note 158.

[166] *Id.*

[167] *See Sex Offender Registration And Notification Act (SORNA)*, U.S. DEP'T OF JUST. (Aug. 11, 2023), https://www.justice.gov/criminal/criminal-ceos/sex-offender-registration-and-notification-act-sorna#:~:text=Sex%20Offender%20Registration%20And%20Notification%20Act%20(SORNA),-MENU%20Subject%20Areas.

[168] *Frequently Asked Questions*, SMART, https://smart.ojp.gov/faqs#2-0 (last visited Nov. 17, 2024).

[169] Rebecca Pirius, *State and Federal Sex Offender Registration Laws*, CRIM. DEF. LAW. (Oct. 18, 2024), https://www.criminaldefenselawyer.com/resources/state-sex-offender-registration.htm.

There are three tiers of sex offender classifications: Tier I, Tier II, and Tier III.[170] Each category is associated with different crimes and comes with different penalties and registration requirements.[171]

Tier I is reserved for sex offenders who do not meet the requirements for Tier II or III,[172] and is generally reserved for misdemeanor sex crimes,[173] which may include "[h]aving or receiving child pornography[,] . . . [v]ideo voyeurism of a minor[,] . . . [or] [t]ransmitting information about a minor to further criminal sexual misconduct."[174] These offenders are only required to verify their registration information annually and typically must remain on the sex offender registry for a minimum of fifteen years.[175] However, if these offenders maintain a clean record, their registration period may be reduced to ten years.[176]

Tier II sex offenders are those who have committed "less serious felony sex crimes,"[177] including crimes that are "punishable by imprisonment for more than one year."[178] These crimes may include sex trafficking, coercion and enticement, transportation with intent to engage in criminal sexual activity, or abusive sexual contact.[179] Tier II offenders must verify their registration information every six months and remain on the registry for twenty-five years.[180]

Tier III is reserved for sex offenders who have committed serious felony sex crimes, such as kidnapping by a non-parent, aggravated sexual abuse, sexual abuse, or abusive sexual contact against a minor younger than thirteen years old.[181] These offenders must verify their registration information every three months and are required to remain on the sex offender registry for life.[182]

Because the TAKE IT DOWN Act already criminalizes deepfake pornography to be treated similarly to genuine pornography, offenders should be placed in the Tier I category, alongside those convicted of pornography-related offenses. Classifying those who create deepfake pornography of minors as Tier I offenders is also most appropriate because deepfake pornography, while serious, is not as egregious as crimes like sex trafficking (a Tier II offense) or aggravated sexual abuse (a Tier III offense). This would impose a lower burden on the perpetrators in comparison to other

[170] 34 U.S.C. § 20911.

[171] *See* 34 U.S.C. § 20911; SMART, *supra* note 163.

[172] 34 U.S.C. § 20911(2).

[173] John Devendorf, *An Overview of Tiers for Convicted Sex Offenders*, LAWINFO (Mar. 12, 2024), https://www.lawinfo.com/resources/sex-crime/sex-offender-tiers.html.

[174] Tammy Cohen, *What Violations Can Land You on a Sex Offender Registry?*, INFOMART (Apr. 29, 2016), https://www.infomart-usa.com/blog/violations-can-land-sex-offender-registry/#:~:text=Offenses%20that%20Can%20Lead%20to%20Sex%20Offender%20Registration&text=Offenses%20often%20fall%20into%20categories,Attempted%20offenses%20are%20also%20prosecuted.

[175] SMART, *supra* note 163.

[176] *Id.*

[177] Devendorf, *supra* note 167.

[178] 34 U.S.C. § 20911(3).

[179] *Id.*

[180] SMART, *supra* note 163.

[181] 34 U.S.C. § 20911(4).

[182] SMART, *supra* note 163.

classifications, as they would only need to verify their information annually and can be on the registry for as little as ten years.

To accomplish this, lawmakers would only need to amend the existing TAKE IT DOWN Act rather than draft an entirely new bill. While this may seem like a large undertaking, only Section (3)(B), which contains penalties for offenses involving minors, would need to be revised. Because statutory language outlining the registration requirements for sex offenders already exists,[183] and because the Act defers to other sections of the United States Code in its language,[184] lawmakers could follow that same framework and simply refer to that section of the U.S. Code rather than redefining the terms themselves.

Thus, the revised section could state something as simple as the following (modifications italicized): "Any person who violates paragraph (2)(B) shall be fined under title 18, United States Code, imprisoned not more than 3 years, or both, *and shall comply with registration requirements for Tier I sex offenders pursuant to title 34, United States Code, Sections 20911–20931*."

B. Policy Considerations

Requiring sex offender registration would not only hold perpetrators accountable but also inform communities of dangerous individuals around them. For social media users wanting to protect their images, this awareness could encourage them to make their accounts private or, for already private users, to be more cautious about accepting new follow or friend requests. Additionally, it would allow parents identify members of their communities who may be dangerous, whether that be physically or online.

This approach would hold adults responsible for their actions while offering grace to minors who may not have thought through the consequences of creating or sharing a deepfake of a peer. Under SORNA, minors are only required to register as juvenile sex offenders if they have committed "particularly serious sexual assault crimes," typically crimes that are "comparable to or more serious than aggravated sexual abuse," [185] meaning that minors who create or distribute deepfake pornography—assuming they have not been convicted of aggravated sexual abuse in the past—would be able to learn from their mistakes without having to endure the challenges that come with carrying a long-lasting mark on their record.

Mandating sex offender registration may seem like a harsh approach, as most criminals would rather serve more time than have a mark on their record that may put restrictions on where they can live or work.[186] However, even offenses as inconsequential as urinating in public can land someone on the sex offender registry, so deepfake child pornography offenses should warrant the same consequence.[187] If

[183] 34 U.S.C. § 20913.

[184] S. 4569.

[185] SMART, *supra* note 163.

[186] Rachel Marshall, *I'm a Public Defender. My Clients Would Rather Go to Jail Than Register as Sex Offenders*, VOX (July 5, 2016, 6:00 AM), https://www.vox.com/2016/7/5/12059448/sex-offender-registry.

[187] Tammy Cohen, *What Violations Can Land You on a Sex Offender Registry?*, INFOMART (Apr. 29, 2016), https://www.infomart-usa.com/blog/violations-can-land-sex-offender-

the public can be made aware of those who may indecently expose themselves—which arguably poses little to no threat to others—they should also be informed of predators who may exploit images of minors. Additionally, given that many offenders would prefer prison time over registration,[188] stating clearly that mandatory registration would be a consequence could potentially deter people from creating or publishing deepfake pornography in the first place.

Finally, because deepfake pornography of minors is treated the same as authentic child pornography under the TAKE IT DOWN Act, some states may already require sex offender registration for this offense. Amending the Act to include mandatory registration would ensure consistency across the board.

CONCLUSION

Deepfake pornography—especially that depicting minors—is a serious issue that demands a serious solution. Mandating sex offender registration for those who possess or publish deepfake pornography of minors would alert communities of people who are dangerous around them, notifying parents to guard their children and images of them, while also protecting minors who create deepfake pornography of their peers, as they would not be required to register. Adding this brief provision to the TAKE IT DOWN Act—a bill already poised for success—rather than drafting an entirely new bill would increase its likelihood of becoming enacted law while avoiding logistical issues. By raising awareness of and penalizing individuals who may use others' images in a violating way, lawmakers can deter individuals from creating deepfake pornography, hold offenders accountable for their actions, and most importantly, empower individuals to protect themselves.

registry/#:~:text=Offenses%20that%20Can%20Lead%20to%20Sex%20Offender%20Registration&text=Offenses%20often%20fall%20into%20categories,Attempted%20offenses%20are%20also%20prosecuted.

[188] Marshall, *supra* note 181.

This page intentionally left blank.

www.ijlet.org

La Nouvelle Jeunesse

ISSN 2769-7142

9 772769 714009