Look Before You Leap: Reflections on Ill-Informed Introductions of AI and Other Digital Technologies

*By Roland Trope and Candace Jones**

I. Introduction

Business enterprises, including law firms, frequently adopt and sometimes develop new digital technologies. A new technology is usually a mixed blessing. Its introduction may improve or add new capabilities to the enterprise. That's the upside. But most new technologies are also honeycombed with flaws, limitations, deficiencies, and vulnerabilities ("faults"). Unless identified and remediated, faults may increase the chances the enterprise's personnel will misuse the new technology with severe consequences. Unmanaged faults may also increase threats posed by bad actors (external or internal). Those are the downsides.

Managing the adoption or development of new technologies so that advantages outweigh risks is not new; it is a recurring imperative. Enterprises should, therefore, direct and train their executives and personnel to follow a set of practical, repeatable strategies when deciding whether, and to what extent, to adopt or develop a new technology and how to introduce the new technology in a manner that manages foreseeable risks. Those strategies should be summoned into practice routinely to improve an enterprise's chances of using new technologies (and changes to incumbent technologies) securely, reliably, and responsibly.

In this essay, we review judicial and administrative agency enforcements and government initiatives in cybersecurity and artificial intelligence ("AI") during the year May 2022–May 2023. Each enforcement illustrates a failure by an enterprise and its personnel to undertake a rigorous review of a new technology, to learn its capabilities and faults, and to assess how the enterprise can safely and securely avert or mitigate faults before deciding whether to introduce the

* Roland Trope is a partner in the New York City offices of Trope and Schramm LLP and an adjunct professor in the Department of Electrical Engineering and Computer Science at the U.S. Military Academy at West Point. Candace Jones is retired from in-house practice at a New York financial institution. Mr. Trope may be contacted at rltrope@tropelaw.com and Ms. Jones at candace.jones.aba@gmail.com.

Disclaimer: The views expressed by the authors are solely their own and have not been reviewed or approved by, and should not be attributed to, the U.S. Military Academy, U.S. Army, U.S. Department of Defense, the U.S. Government, or any institution to which they are, or have been, affiliated.

The authors thank Professor Sarah Jane Hughes for editing, and Carol Elizabeth Bunting for editing, cite checking, and bluebooking, this essay.

new digital tool. In each case, the enterprise could have followed some basic, systematic steps to manage its adoption or development of a new technology: *first* by seeking to understand and plan for the *known inherent faults* of the new digital tool; and *second* by implementing policies and procedures for using the digital tool or its output in ways that would avert or mitigate the faults and reduce the risks of personnel misusing the tool or of threat actors maliciously exploiting it.

Part II examines the sanctions decision in *Mata v. Avianca, Inc.*, where lawyers failed to investigate the generative AI application ChatGPT-4 (commonly referred to as "ChatGPT") before adopting it for legal research and citing in a court filing the "cases" ChatGPT "produced" that existed only as ChatGPT's factitious outputs.

Part III discusses settlements in cybersecurity and privacy cases, one brought by the Federal Trade Commission against Ring LLC and the other by the New York Department of Financial Services against bitFlyer USA, Inc.

Part IV reviews government initiatives in the United States, the European Union ("EU"), and the People's Republic of China ("PRC").

Part V provides concluding observations.

II. *Mata v. Avianca, Inc.*

In 2019, Roberto Mata ("Mata") suffered a knee injury when struck by a metal serving cart controlled by an Avianca Airlines employee on an international flight.[1] Avianca, Inc. ("Avianca") owned and operated the eponymous airline. Mata sued Avianca for damages caused by the airline's alleged negligence.

On February 22, 2022, Avianca moved for dismissal on affirmative defenses, including (i) the claim was time-barred by the applicable two-year statute of limitations under the Montreal Convention and (ii) Mata's claims were discharged by Avianca's subsequent bankruptcy case.[2]

Mata was represented by Levidow, Levidow & Oberman, P.C. Two of its attorneys, Peter LoDuca ("LoDuca") and Steven Schwartz ("Schwartz"), handled the response to Avianca's motion to dismiss. On March 1, 2023, LoDuca signed and submitted an Affirmation in Opposition ("Affirmation") contending the Montreal Convention did not preempt state law remedies, including the three-year statute of limitations applicable to Mata's New York law claim, and the bankruptcy action tolled the Montreal Convention's statute of limitations. In support, the Affirmation cited and quoted several federal circuit court decisions.[3]

1. *See* Mata v. Avianca, Inc., No. 22-cv-1461 (PKC), 2023 U.S. Dist. LEXIS 108263, at *4 (S.D.N.Y. June 22, 2023).

2. Answer of Avianca, Inc. at 3, 5, Mata v. Avianca, Inc., No. 22-cv-1461 (PKC), 2023 U.S. Dist. LEXIS 108263 (S.D.N.Y. June 22, 2023), https://storage.courtlistener.com/recap/gov.uscourts.nysd.575368/gov.uscourts.nysd.575368.4.0.pdf.

3. Affirmation in Opp'n at 1, 4–8, Mata v. Avianca, Inc., No. 22-cv-1461 (PKC), 2023 U.S. Dist. LEXIS 108263 (S.D.N.Y. June 22, 2023), https://storage.courtlistener.com/recap/gov.uscourts.nysd.575368/gov.uscourts.nysd.575368.21.0.pdf.

On March 3, 2023, Avianca's counsel submitted a Reply Memorandum in Further Support of Defendant's Motion to Dismiss Plaintiff's Verified Complaint ("Reply Memo"), representing that defense counsel was "unable to locate most of the case law cited" in plaintiff's Affirmation and the few cases that could be found "do not stand for the propositions for which they are cited."[4] The court searched for, but could not locate, multiple authorities cited in the Affirmation.[5]

On April 11 and 12, 2023, the court issued orders ("Case Orders") requiring LoDuca to file, by April 18, 2023, an affidavit annexing copies of nine cases, adding that "[f]ailure to comply will result in dismissal of the action."[6] The requested cases were those the Reply represented to be unlocatable.

On April 25, 2023, LoDuca submitted a sworn affidavit ("Affidavit"), annexing copies of eight of the nine court-ordered cases. The Affidavit adds that counsel was "unable to locate" the *Zicherman* case, six of the cases "may not be inclusive of the entire opinions but only what is made available by online database," and one opinion is "unpublished."[7] The Affidavit did not address counsel's reliance on those cases for arguments made in the Affirmation or seek to correct the Affirmation.

On May 4, 2023, the court issued an Order that LoDuca show cause on June 8, 2023, "why he ought not to be sanctioned pursuant to: (1) Rule 11(b)(2) & (c), Fed. R. Civ. P., (2) 28 U.S.C. § 1927, and (3) the inherent power of the Court" for citing non-existent cases in his Affirmation and submitting to the court his Affidavit with "copies of non-existent judicial opinions."[8] The Show Cause Order observed that "[s]ix of the submitted cases appear to be bogus judicial decisions with bogus quotes and bogus internal citations."[9] On May 26, 2023, the court issued a supplemental Order to Show Cause, incorporating its Order of May 4, 2023, and requiring LoDuca, Schwartz, and the Levidow firm to show cause why each should not be sanctioned pursuant to Rule 11.[10]

During the hearing on the Order to Show Cause ("Sanctions Hearing"), plaintiff's attorneys insisted the citations were a mistake and denied any willful attempt to mislead the court. On the other hand, statements made by LoDuca and Schwartz at the Sanctions Hearing disclosed that they misunderstood how ChatGPT produced its outputs and had not verified the accuracy of each

4. Reply Mem. at 2, Mata v. Avianca, Inc., No. 22-cv-1461 (PKC), 2023 U.S. Dist. LEXIS 108263 (S.D.N.Y. June 22, 2023), https://storage.courtlistener.com/recap/gov.uscourts.nysd.575368/gov.uscourts.nysd.575368.24.0_1.pdf.

5. *Mata*, 2023 U.S. Dist. LEXIS 108263, at *8.

6. *Id.* at *10.

7. LoDuca Aff. at 1–2, Mata v. Avianca, Inc., No. 22-cv-1461 (PKC), 2023 U.S. Dist. LEXIS 108263 (S.D.N.Y. June 22, 2023), https://storage.courtlistener.com/recap/gov.uscourts.nysd.575368/gov.uscourts.nysd.575368.29.0_1.pdf.

8. Order to Show Cause at 3, Mata v. Avianca, Inc., No. 22-cv-1461 (PKC), 2023 U.S. Dist. LEXIS 108263 (S.D.N.Y. June 22, 2023), https://storage.courtlistener.com/recap/gov.uscourts.nysd.575368/gov.uscourts.nysd.575368.31.0_1.pdf.

9. *Id.* at 1.

10. *See* Order to Show Cause at 1–2, Mata v. Avianca, Inc., No. 22-cv-1461 (PKC), 2023 U.S. Dist. LEXIS 108263 (S.D.N.Y. June 22, 2023), https://storage.courtlistener.com/recap/gov.uscourts.nysd.575368/gov.uscourts.nysd.575368.33.0_1.pdf.

ChaptGPT output before using it as their own work. They also didn't take the opportunity given them by the Reply Memo or Case Orders to correct their misplaced reliance on the ChatGPT output.

- Asked how he satisfied himself that the research reflected in the Affirmation was "accurate and truthful," LoDuca admitted he didn't. He relied on his colleague, Schwartz, to research and prepare it.[11]
- LoDuca didn't read any of the cases cited in the Affirmation.[12]
- He did nothing to "ensure that those cases existed."[13]
- He didn't read the five-page Reply Memo, which flagged cases that could not be located or did not stand for the propositions for which they were cited in the Affirmation.[14]
- LoDuca "didn't do anything other than" turn the Case Orders over to Schwartz,[15] read the draft Affidavit, and note "the cases that were attached to it."[16]
- LoDuca didn't notice the cases' conspicuous textual incongruities, but acknowledged them when the court pointed them out.[17]

The court's examination of Schwartz revealed he didn't really understand what the technology was doing; Schwartz assumed it was a "super search engine."[18] He did not appreciate that ChatGPT, a generative AI application, composes responses based on its predictive models and training.

- Schwartz's first prompt directed ChatGPT not to "search, find, and retrieve" (as one might with a natural language search engine), but to "argue that the statute of limitations is tolled by bankruptcy of defendant pursuant to montreal convention."[19]
- His second prompt directed ChatGPT to "provide" cases that supported its argument: "**User**: provide case law in support that statute of limitations is tolled by bankruptcy of defendant under montreal convention."[20]

11. Transcript of Hearing on Order to Show Cause at 6, Mata v. Avianca, Inc., No. 22-cv-1461 (PKC), 2023 U.S. Dist. LEXIS 108263 (S.D.N.Y. June 22, 2023), https://drive.google.com/file/d/1ilH_J8pzdF8huV57bx_Mop7vJNV4EkiF/view.
12. *Id.* at 9.
13. *Id.*
14. *Id.* at 10.
15. *Id.* at 13.
16. *Id.* at 14.
17. *Id.* at 15–16.
18. *Id.* at 24.
19. Decl. of Steven Schwartz, Exh. A at 1, Mata v. Avianca, Inc., No. 22-cv-1461 (PKC), 2023 U.S. Dist. LEXIS 108263 (S.D.N.Y. June 22, 2023), https://storage.courtlistener.com/recap/gov.uscourts.nysd.575368/gov.uscourts.nysd.575368.46.1.pdf.
20. *Id.* at 2.

In response, ChatGPT "produced" case names, citations, and a sentence ostensibly describing the holding.

- None of Schwartz's case prompts directed ChatGPT to shepherdize or check if the "produced" cases had been distinguished, reversed, or overturned.[21] The history of his prompts (included in Exhibit A to his Declaration) shows no curiosity about the existence of cases to the contrary.[22]
- None of Schwartz's prompts to ChatGPT reflected an awareness that the Montreal Convention's limit on the time to bring an action had been interpreted by the Second Circuit to be a "condition precedent," which could not be tolled by a bankruptcy filing, rather than a "statute of limitations," which could be tolled by a bankruptcy filing.[23]
- To comply with the court's Case Orders, Schwartz "went back to the only place" that he could find the cases, ChatGPT, and prompted it to "produce" them.[24]

On June 22, 2023, the court issued its opinion and order on sanctions, holding that LoDuca, Schwartz, and the Levidow firm "abandoned their responsibilities [to ensure the accuracy of their findings] when they submitted non-existent judicial opinions with fake quotes and citations created by . . . ChatGPT, then continued to stand by the fake opinions after judicial orders called their existence into question."[25] The court found LoDuca and Schwartz "acted with subjective bad faith in violating Rule 11" and found the Levidow Firm "jointly and severally liable" for their Rule 11 violations.[26]

The court sanctioned LoDuca, Schwartz, and the Levidow Firm under Rule 11 and, alternatively, under the court's inherent power.[27] The court imposed a $5,000 penalty and ordered the lawyers and law firm to send to their client, Mata, and to "each judge falsely identified as the author of the fake" opinion a copy of the court's Sanctions Opinion and Order, a transcript of the Sanctions Hearing, and a copy of the April 25 Affirmation, including the "fake 'opinion' attributed to the recipient judge."[28]

ChatGPT and other generative AI applications might, with improvements, be determined by counsel to be appropriate for client work, provided counsel bases that determination on a careful examination of its capabilities and *known* faults

21. *Id*. at 1–15.
22. *Id*.
23. Mata v. Avianca, Inc., No. 22-cv-1461 (PKC), 2023 U.S. Dist. LEXIS 108261, at *7–8 (S.D.N.Y. June 22, 2023) (explaining the Second Circuit's interpretation in its opinion dismissing the plaintiff's case).
24. Decl. of Steven Schwartz at paras. 24–25, Mata v. Avianca, Inc., No. 22-cv-1461 (PKC), 2023 U.S. Dist. LEXIS 108263 (S.D.N.Y. June 22, 2023), https://storage.courtlistener.com/recap/gov.uscourts.nysd.575368/gov.uscourts.nysd.575368.46.0.pdf.
25. *Mata*, 2023 U.S. Dist. LEXIS 108263, at *2.
26. *Id.* at *39–41.
27. *Id*. at *42.
28. *Id*. at *46.

and takes measures to avert or mitigate the adverse effects of those faults. Responsible introduction of any new technology by a lawyer or law firm for client work would include, at a minimum, adherence to procedures that determine the suitability of the tool for its intended purpose and manage limitations and faults. For generative AI, those basic steps would include:

1. Undertaking a thorough due diligence check on the generative AI application to determine how the application could advance work done for a client consistent with counsel's professional obligations;

2. Reading key documentation on the developer's website that describes the application, how the application works, what it was designed to do, its current limitations and deficiencies, and any tendency to put at risk a client's interests, such as by generating inaccurate outputs. Such review should include, at a minimum, the developer's Terms of Use, Frequently or Commonly Asked Questions ("CAQs"), and any System Cards, which explain, for example, a generative AI's performance, capabilities, and limitations;

3. Seeking out reviews and undertaking tests to identify inaccurate or biased results and training users about known indicia of inaccurate and biased outputs to help users recognize and remediate those deficiencies;

4. Examining how use of the technology could pose security risks, including collection and processing of data that could compromise client confidentiality, inadvertently waive attorney-client or attorney work product privileges, and compromise the enforceability of client trade secrets and opportunity to file timely patent applications. If risks to client confidentiality and interests cannot be averted, the AI application has no place in a lawyer's tool kit for legal work; and

5. Considering carefully whether introduction of the technology might cause counsel to violate ethical obligations to (i) provide competent representation to clients, (ii) inform clients and obtain their consent to the use, (iii) protect client confidential information, and (iv) avoid unauthorized practice.

Terms of Use are routinely bypassed as users click to get to their online objectives. Enterprises assume the risk of missing valuable information and product warnings when their personnel do that. Consider what the documentation posted on OpenAI's website reveals, for example, about the capabilities and faults in ChatGPT:

- Terms of Use, Section 3(d) ("Accuracy"):

  "Given the probabilistic nature of machine learning, use of our Services [ChatGPT] may in some situations result in incorrect Output that does not accurately reflect real people, places, or facts. You should evaluate

the accuracy of any Output as appropriate for your use case, including by using human review of the Output."[29]
Takeaway: Users must determine whether Output is accurate. Lawyers will need time to validate the Output (or risk Rule 11 sanctions).[30]

- Terms of Use, Section 5(b):

  "You must implement reasonable and appropriate measures designed to help secure your access to and use of the Services. If you discover any vulnerabilities or breaches related to your use of the Services, you must promptly contact OpenAI and provide details of the vulnerability or breach."[31]

  Takeaway: OpenAI gives no assurance that ChatGPT will be cyber secure to use, and instead puts users on notice that "they are responsible for cyber risks and may have incomplete information about risks to their data and systems."[32]

- CAQs:

  "5. **Who can view my conversations?**

  - . . . [W]e review conversations to improve our systems and to ensure the content complies with our policies and safety requirements.

  6. **Will you use my conversations for training?**

  - Yes. Your conversations may be reviewed by our AI trainers to improve our systems."[33]

  Takeaway: Client confidential information entered into a ChatGPT prompt should, upon transmission, be presumed to no longer be confidential.

- GPT-4 System Card:

  ChatGPT "maintains a tendency to make up facts, to double-down on incorrect information, and to perform tasks incorrectly. Further, it often exhibits these tendencies in ways that are more convincing and believable than earlier GPT models (e.g., due to authoritative tone . . .), increasing the risk of overreliance. Overreliance occurs when users excessively trust and depend on the model, potentially leading to unnoticed mistakes and inadequate oversight. This can happen in various ways: users may not be

29. *Terms of Use*, OpenAI § 3(d), https://openai.com/policies/terms-of-use (last visited Mar. 14, 2023).
30. *See* Roland Trope, Candace Jones & Claudia Ray, *If You Think It "Thinks," Think Again*, N.Y.L.J. (July 3, 2023), https://www.law.com/newyorklawjournal/2023/07/03/if-you-think-it-thinks-think-again/.
31. *Terms of Use*, *supra* note 29, § 5(b).
32. Trope, Jones & Ray, *supra* note 30.
33. *Commonly Asked Questions*, OpenAI, https://help.openai.com/en/articles/6783457-what-is-chatgpt (last visited June 18, 2023).

> vigilant for errors due to trust in the model; . . . or they may utilize the model in domains where they lack expertise, making it difficult to identify mistakes. . . . As mistakes become harder for the average human user to detect and general trust in the model grows, users are less likely to challenge or verify the model's responses."[34]
>
> *Takeaway:* The diligence required to scrutinize, identify, and correct every inaccuracy in ChatGPT's outputs suggests it will seldom reduce counsel's work or time.[35]

Beyond good practices for introducing new technology, lawyers must also watch for new rules from courts about use of AI. Judges in several jurisdictions have issued rules or practice directions prohibiting and requiring certification of non-use of generative AI or requiring a detailed disclosure of the use of generative AI applications in counsel's court submissions.[36]

III. Ill-Informed Technology Introduction Leading to Regulatory Enforcement Actions

Two settlements in cybersecurity and privacy cases this year illustrate how inadequate attention to foreseeable faults with new technology introduction led to consumer harm and regulatory enforcement action. These cases also foreshadow regulatory consequences for enterprises that rush AI applications to market leaving security, privacy, and other foreseeable dangers by the wayside. While the *Ring* and *bitFlyer* cases noted below are not about AI, they involve handling sensitive data and implementing effective risk management practices. Those activities are highly relevant to AI companies, which handle huge stores of data and are innovating at a furious pace in environments where risks are not completely known or understood.

---

34. OpenAI, ChatGPT-4 System Card 19–20 (Mar. 23, 2023), https://cdn.openai.com/papers/gpt-4-system-card.pdf.

35. Note the emergence of the "deep fake" defense where defendants challenge a party's proposed use of a video recording of an individual or officer making public statements, and insist the opposing party prove the video's authenticity and that it has not been created or manipulated with generative AI. Some courts have rejected the defense. *See* preliminary ruling in *Sz Huang v. Tesla, Inc.*, No. 19CV346663, slip op. at 29 (Santa Clara Cnty. Super. Ct. Apr. 27, 2023), where the court rejected the defense:

> "[W]hat Tesla is contending is deeply troubling to the Court. Their position is that because Mr. Musk is famous and might be more of a target for deep fakes, his public statements are immune. In other words, Mr. Musk, and others in his position, can simply say whatever they like in the public domain, then hide behind the potential for their recorded statements being a deep fake to avoid taking ownership of what they did actually say and do. The Court is unwilling to set such a precedent by condoning Tesla's approach here." https://cdn.arstechnica.net/wp-content/uploads/2023/04/musk-deepfake-ruling.pdf.

36. *See* Chief Justice Glenn D. Joyal, *Practice Direction re Use of Artificial Intelligence in Court Submissions*, Manitoba Cts. (June 23, 2023), https://www.manitobacourts.mb.ca/site/assets/files/2045/practice_direction_-_use_of_artificial_intelligence_in_court_submissions.pdf; Judge Brantley Starr, *Mandatory Certification Regarding Generative Artificial Intelligence*, U.S. Dist. Ct. (June 12, 2023), https://www.txnd.uscourts.gov/judge/judge-brantley-starr.

A. *FTC v. Ring LLC*[37]

The Federal Trade Commission ("FTC") and Ring LLC ("Ring"), maker of security cameras for indoor and outdoor household use, agreed to a Proposed Stipulated Order settling FTC allegations of unfair and deceptive trade practices (the "Proposed Order"). In a complaint filed with the Proposed Order (the "Complaint"), the FTC describes a young technology company launching its indoor security camera business without establishing technical controls, procedures, or monitoring designed to limit employee and contractor access to video captured on installed devices. Customers installed the cameras throughout their homes, monitoring private spaces including bedrooms and bathrooms.

In the absence of security controls and training, some employees accessed and viewed highly sensitive video for illegitimate purposes. According to the Complaint, Ring responded to incidents incrementally, taking almost two years to narrow video data access by employees and overseas contractors on a "need-to-know" basis and even longer to institute privacy and security training for its employees.[38] Furthermore, Ring's early terms of use did not give Ring adequate customer consent to use sensitive video data for product development and improvement.[39]

The FTC also alleged Ring's pursuit of commercial success prioritized easy data access for its developers over security controls. "Despite promising greater security as its products' core feature, Ring ignored information security considerations when management believed they would interfere with growth. In pursuit of rapid product development, before September 2017, Ring did not limit access to customers' video data to employees who needed the access to perform their job function (e.g., customer support, improvement of that product, etc.)."[40] What motivated Ring to change? Apparently, a different phase in its pursuit of commercial success; the FTC says Ring started "improving security practices to make Ring more appealing to potential acquirers."[41]

In the Proposed Order, Ring agreed to pay a money judgment of $5,800,000. In addition, Ring must delete video recordings collected without customer consents adequate to support Ring's continued use. Going beyond the video data, Ring must delete work product—its models and algorithms—"developed in whole or in part from review and annotation of [the early recordings]."[42] For

37. FTC v. Ring LLC, No. 1:23-cv-1549 (D.D.C. May 31, 2023).

38. Compl. at para. 48, FTC v. Ring LLC, No. 1:23-cv-1549 (D.D.C. May 31, 2023), https://www.ftc.gov/system/files/ftc_gov/pdf/complaint_ring.pdf, (summarizing in a timeline Ring's allegedly "unreasonable data security and privacy practices").

39. *Id.* Generative AI developers use large datasets to train and refine their applications, and the terms of use for some generative AI applications authorize use of user-entered data for product development and improvement. *See, e.g.*, *Terms of Use*, *supra* note 29, § 3(c) (permitting OpenAI to use content it collects outside its API to improve its services).

40. Compl. at para. 13, FTC v. Ring LLC, No. 1:23-cv-1549 (D.D.C. May 31, 2023), https://www.ftc.gov/system/files/ftc_gov/pdf/complaint_ring.pdf.

41. *Id*. at para. 21.

42. Proposed Stipulated Order § II(A)(3), Definitions para. A, FTC v. Ring LLC, No. 1:23-cv-1549 (D.D.C. May 31, 2023), https://www.ftc.gov/system/files/ftc_gov/pdf/proposed_stipulated_order_ring.pdf.

twenty years, Ring must maintain a comprehensive privacy and data security program ("Program"), train its employees to safeguard the privacy, security, confidentiality, and integrity of customer sensitive data, and periodically test, monitor, reinforce, and certify the effectiveness of the Program.[43]

The *Ring* case illustrates adverse consequences that may result from adopting a business strategy that pursues short-term profit at the expense of safeguarding customers' sensitive data. Having failed to implement sound practices for cybersecurity and handling sensitive information, Ring had "no idea how many instances of inappropriate access to customers' sensitive video data actually occurred."[44] Enterprises sometimes defer cybersecurity and privacy measures in the belief that it's better to err and ask forgiveness than to invest time and funds in such measures from the outset. That choice may cost more than the amount of damages and fines, as illustrated by recent FTC cases that include forfeiture of digital assets created with misused data.[45] That penalty can be a significant setback, depriving an enterprise of its early investments and, perhaps, its competitive advantages.

B. *In re bitFlyer USA, Inc.*[46]

bitFlyer USA is a cryptocurrency trading platform licensed by the New York Department of Financial Services ("NYDFS") under New York's Virtual Currency Regulation and subject to New York's Cybersecurity Regulation.[47] bitFlyer USA had been a subsidiary of bitFlyer, Inc. (Japan) until they and other affiliates were made wholly owned subsidiaries of bitFlyer Holdings, Inc.[48] bitFlyer, Inc. (Japan) supports bitFlyer USA with its internal operations including, among others, product and application development, information security and incident management, and information system development and maintenance.[49]

Examinations by NYDFS staff identified deficiencies in bitFlyer USA's cybersecurity program which led to an enforcement investigation. The Cybersecurity Regulation required bitFlyer USA to conduct periodic cybersecurity risk assessments according to practice standards that include written policies and procedures.[50] A proper risk assessment is essential for a properly designed and managed cybersecurity risk program.

NYDFS found bitFlyer USA's cybersecurity program lacking in its fundamentals. For example, bitFlyer USA adopted security policies copied from its Japanese

43. *Id.* § III(G), https://www.ftc.gov/system/files/ftc_gov/pdf/proposed_stipulated_order_ring.pdf.

44. Complaint at para. 24, FTC v. Ring LLC, No. 1:23-cv-1549 (D.D.C. May 31, 2023), https://www.ftc.gov/system/files/ftc_gov/pdf/complaint_ring.pdf.

45. For an earlier example, see *In re* Everalbum, Inc., Decision and Order (FTC May 5, 2021), https://www.ftc.gov/system/files/documents/cases/1923172_-_everalbum_decision_final.pdf. Similar to *Ring*, the FTC alleged Everalbum used customer photos and video for product development without customer consent. The FTC ordered Everalbum to delete photos and video as well as models and algorithms developed in whole or in part using them. *Id.*

46. *In re* bitFlyer USA, Inc., Consent Order (N.Y. State Dep't of Fin. Servs. May 1, 2023).

47. *Id.* at 1.

48. *Id.* at para. 5.

49. *Id.* at para. 6.

50. *Id.* at paras. 8–9.

affiliate without accurately or completely translating the policies to English, and it used template forms from various sources without even changing generic identifiers like "ABC Company."[51] bitFlyer USA's policies and procedures were not customized to its needs and risks. bitFlyer USA also had not conducted a comprehensive cybersecurity assessment,[52] and it did not have a written security policy approved by its board of directors.[53]

The bitFlyer case cautions AI developers and adopters against cutting corners on security risk assessment, particularly in regulated industries where the law requires attention to security. Like other digital technologies, AI can have security vulnerabilities throughout its lifecycle. The AI life cycle is complex and multi-layered. It includes software, model, and algorithm design and development, collection of training data, incorporation in applications and business processes, execution of automated functions, and utilization of output for decision-making with varying degrees of human intervention. AI also poses risks its developers and the public do not fully understand or have not yet discovered.[54] Knowing that, AI developers and adopters should conduct comprehensive risk assessments of AI applications from the outset and seek to implement safeguards, continuously monitor their effectiveness, and act promptly to respond to emergent risks and signs that safeguards might be insufficient. Failure to build effective security and risk programs alongside effective AI applications compounds potential adverse consequences for regulated activities like those in bitFlyer.

C. Nvidia's Vulnerability

AI adopters must also be attuned to the allocation of risk for faults in AI applications. Consider this illustration. Nvidia computer chips are used by AI developers in building generative AI systems. When the AI research firm Robust Intelligence, Inc. breached safeguards in the Nvidia chips, Nvidia responded that it had fixed one of the root causes of the reported vulnerabilities.[55] Nvidia's VP of applied research was also quoted as saying Robust Intelligence "identified additional steps that would be needed to deploy a production application."[56] Lawyers: note in that statement the business strategy of shifting liability downstream; AI application developers will have to consider in their designs the vulnerabilities that may be introduced through the component chips they use. That

51. *Id.* at para. 16.
52. *Id.* at paras. 13–14.
53. *Id.* at para. 15.
54. *See, e.g.*, OpenAI, *supra* note 34. The System Card is a sixty-page report analyzing GPT-4, highlighting its safety challenges, safety processes adopted by OpenAI, and demonstrating that mitigations are "limited and remain brittle in some cases. This points to the need for anticipatory planning and governance." *Id.* at 1.
55. Mehul Srivastava & Cristina Criddle, *Nvidia's AI Software Tricked into Leaking Data*, Fin. Times (June 9, 2023), https://www.ft.com/content/5aceb7a6-9d5a-4f1f-af3d-1ef0129b0934.
56. *Id.*

type of risk-shifting is accomplished by contract. In some cases, contracts could be Terms of Use accepted by a developer or tool user with a click.

D. Government AI Initiatives

AI-specific regulation emerged in multiple forms ranging from guidance for self-regulation to legislation and judicial orders.[57]

United States

- In October 2022, the White House released a Blueprint for AI Bill of Rights (the "Blueprint"). It sets out "five principles that should guide the design, use, and deployment of automated systems to protect the American public in the age of artificial intelligence":[58]
  (1) AI systems should be safe and effective;
  (2) algorithms should not discriminate and AI systems should be designed and used in equitable ways;
  (3) data practices should protect privacy;
  (4) design and use of AI should be disclosed and explained in plain language; and
  (5) there should be a human alternative and human backstop to AI applications.[59]

A Technical Companion offers practical steps for implementing the vision of the Blueprint.[60]

- In January 2023, the National Institute of Standards and Technology issued a first edition AI Risk Management Framework.[61] NIST intends the Framework to be used voluntarily "to improve the ability to incorporate trustworthiness considerations into the design, development, use, and evaluation of AI products, services, and systems."[62]
- In February 2023, President Biden signed an Executive Order requiring AI systems designed, developed, acquired, and used by the federal government to operate "in a manner that advances equity."[63]

57. *See, e.g.*, Joyal, *supra* note 36; Starr, *supra* note 36.
58. White House Off. of Sci. & Tech. Pol'y, Blueprint for an AI Bill of Rights: Making Automated Systems Work for the American People 3 (Oct. 2022), https://www.whitehouse.gov/wp-content/uploads/2022/10/Blueprint-for-an-AI-Bill-of-Rights.pdf.
59. *Id.* at 5–7.
60. *Id.* at 12–73.
61. Artificial Intelligence Risk Management Framework (AI RMF 1.0), NIST AI 100-1 (Jan. 2023), https://nvlpubs.nist.gov/nistpubs/ai/NIST.AI.100-1.pdf.
62. *AI Risk Management Framework*, NIST, https://www.nist.gov/itl/ai-risk-management-framework (last visited June 11, 2023).
63. Exec. Order No. 14,091, 88 Fed. Reg. 10825 (Feb. 22, 2023).

People's Republic of China ("PRC")

- In April 2023, the PRC released its draft Measures for the Management of Generative Artificial Intelligence Services[64] ("draft Measures") to be added to its existing regulatory regime for data security and privacy. The draft Measures would apply to "research, development, and use of products with generative AI functions and to the provision of services to the public within [PRC territory].[65] China's regulations place responsibility for generative AI—from training to output—on AI providers. The regulations require providers to submit security assessments and algorithm information (an existing mandate) to the PRC government and to generate output that "reflect[s] the Socialist Core Values."[66]

European Union ("EU")

- In May 2023, the European Parliament advanced the EU's A.I. Act to regulate AI systems used in Europe. "The rules follow a risk-based approach and establish obligations for providers and users depending on the level of risk the AI can generate. AI systems with an unacceptable level of risk to people's safety would be strictly prohibited."[67]

IV. Conclusion

As AI entered widespread use in generative AI applications during the year in review, the pace of enforcement and government initiatives to regulate AI also quickened. Judges in multiple jurisdictions responded to inaccurate court documents with new rules for counsel's use of AI in court submissions. Governments advanced various initiatives to regulate AI, which diverge in their approaches and focus. Enforcement of existing privacy and security regulations identify deficiencies foreseeable in the AI space. These developments suggest enterprises and lawyers will need to do more than ever before to understand each new technology's benefits and faults before relying on it to advance material business and professional objectives.

64. *See* Seaton Huang, Helen Toner, Zac Haluza, Rogier Creemers & Graham Webster, *Translation: Measures for the Management of Generative Artificial Intelligence Services (Draft for Comment)—April 2023*, Stan. Univ. DigiChina F. (Apr. 12, 2023), https://digichina.stanford.edu/work/translation-measures-for-the-management-of-generative-artificial-intelligence-services-draft-for-comment-april-2023/.

65. *Id*.

66. *Id*. The Stanford DigiChina Project also published expert views on the proposed Measures. Helen Toner et al., *How Will China's Generative AI Regulations Shape the Future? A DigiChina Forum*, Stan. Univ. DigiChina F. (Apr. 19, 2023), https://digichina.stanford.edu/work/how-will-chinas-generative-ai-regulations-shape-the-future-a-digichina-forum/. *See also* Yan Luo, Xuezi Dan, Vicky Liu & Nicholas Shepherd, *China Proposes Draft Measures to Regulate Generative AI*, Inside Privacy (Apr. 12, 2023), https://www.insideprivacy.com/artificial-intelligence/china-proposes-draft-measures-to-regulate-generative-ai/.

67. *AI Act: A Step Closer to the First Rules on Artificial Intelligence*, Eur. Parliament Press Room (May 5, 2023), https://www.europarl.europa.eu/news/en/press-room/20230505IPR84904/ai-act-a-step-closer-to-the-first-rules-on-artificial-intelligence.