Debevoise & Plimpton

Debevoise In Depth

Preparing for AI Whistleblowers – 2026 Update

March 30, 2026

As artificial intelligence ("AI") use and capabilities surged in 2024, we wrote about a new emerging risk for companies: AI whistleblowers. Two years later, in 2026, the current administration has generally reduced enforcement and de-emphasized whistleblower awards. Nonetheless, the Trump Securities Exchange Commission ("SEC") and Department of Justice ("DOJ") have signaled a focus on policing AI-related conduct. Moreover, during the same period, AI usage has accelerated substantially and whistleblower liability remains a significant compliance and litigation risk. Internal disputes concerning cybersecurity and AI governance have persisted and, in some sectors, intensified amid evolving but still fragmented regulatory guidance. Public and investor skepticism has also grown regarding companies' ability to safeguard data, mitigate AI-driven risks, and accurately represent AI capabilities—all of which dramatically increase the risk of AI whistleblower activity.[1] These risks are particularly acute with respect to agentic AI, which presents a new risk frontier.

AI Whistleblower Risks Have Sharply Increased Since 2024. Enterprise AI tool development and deployment have accelerated exponentially since 2024.[2] In particular, agentic AI—artificial intelligence systems that can complete tasks with little to no supervision—has exploded in development and usage over the past year, and poses multiple new compliance and operational risks. For example, agentic AI tools may undertake tasks beyond the scope of authorization; access data or systems beyond the scope of authorization; reinforce biased or erroneous outcomes; generate strategies to meet goals that developers did not program and cannot easily follow; and behave unpredictably when facing novel situations.[3] Malicious agents may also exploit trust

[1] *2025 AI Index Report*, STAN. U. HUM.-CENTERED AI (Apr. 2025), https://hai.stanford.edu/ai-index/2025-ai-index-report.

[2] *The State of AI*, MCKINSEY & CO. (Nov. 5, 2025), https://www.mckinsey.com/capabilities/quantumblack/our-insights/the-state-of-ai.

[3] Sanjay Bhakta, *Safeguarding Agentic AI: Why Autonomy Demands Governance and Security*, THOMSON REUTERS (Nov. 13, 2025), https://www.thomsonreuters.com/en-us/posts/technology/safeguarding-agentic-ai/; Cristina Catania & Ida Kristensen, *Deploying Agentic AI with Safety and Security: A Playbook for Technology Leaders*, MCKINSEY & CO. (Oct. 16, 2025), https://www.mckinsey.com/capabilities/risk-and-resilience/our-insights/deploying-agentic-ai-with-safety-and-security-a-playbook-for-technology-leaders; *Rise and Risks of*

mechanisms to trick agentic AI into granting unauthorized privileges, leading to inadvertent but potentially catastrophic exposure of systems and data.[4]

These risks along with recent AI data breaches have led to widespread public skepticism and distrust in AI.[5] According to a recent study conducted by Pew Research Center, "only 17% of the general public believes that AI will have a positive impact on the U.S. over the next two decades."[6] A December 2025 YouGov survey similarly found that only 5% of Americans say they trust AI "a lot" and only 26% trust it "somewhat".[7] The pervasive skepticism of AI—coupled with the rapid deployment of AI across enterprises—increases the likelihood of AI-related whistleblower activity.

The SEC's Current Approach to AI and Its Whistleblower Program

AI in Enforcement and Examinations. In February 2025, the SEC created the Cyber and Emerging Technologies Unit (CETU), which expressly targets fraud involving emerging technologies, including artificial intelligence and machine learning.[8] In parallel, the Division of Examinations' FY2026 priorities specifically identify registrants' use of AI technologies and trading algorithms—focusing on the accuracy of AI-related representations and the adequacy of policies, procedures, and supervision. Although the Atkins-led SEC has charged only one AI-related fraud case,[9] the inclusion of AI in both SEC examination and enforcement priorities signals that the SEC will scrutinize fraud risks—particularly in the retail investor context—in connection with AI-related claims.

Developments in The SEC's Whistleblower Program. Under federal law, whistleblowers who provide the SEC with original, timely, and credible information that leads to a successful enforcement action may be eligible for awards ranging from 10 to 30 percent of the money collected, when monetary sanctions exceed $1 million.

Agentic AI, PRICEWATERHOUSECOOPERS, (July 17, 2025), https://www.pwc.com/us/en/industries/tmt/library/trust-and-safety-outlook/rise-and-risks-of-agentic-ai.html.

[4] Bhakta, *supra* note 3; Catania & Kristensen, *supra* note 3; PRICEWATERHOUSECOOPERS, *supra* note 3.

[5] Aisha Down, *Meta AI Agents Instruction Causes Large Sensitive Data Leak to Employees*, THE GUARDIAN (Mar. 20, 2026), https://www.theguardian.com/technology/2026/mar/20/meta-ai-agents-instruction-causes-large-sensitive-data-leak-to-employees.

[6] Michelle Faverio & Emma Kikuchi, *Key Findings About How Americans View Artificial Intelligence*, PEW RSCH. CTR. (Mar. 12, 2026), https://www.pewresearch.org/short-reads/2026/03/12/key-findings-about-how-americans-view-artificial-intelligence.

[7] Clifton Mark, *Most Americans Use AI but Still Don't Trust It*, YOUGOV, https://yougov.com/en-us/articles/53701-most-americans-use-ai-but-still-dont-trust-it (last visited Mar. 25, 2026).

[8] *SEC Announces Cyber and Emerging Technologies Unit to Protect Retail Investors*, U.S. Sec. & Exch. Comm'n (Feb. 20, 2025), https://www.sec.gov/newsroom/press-releases/2025-42.

[9] *Securities and Exchange Commission v. Morocoin Tech Corp., et al.*, No. 25-cv-04102 (D. Colo. filed Dec. 22, 2025).

In FY 2025, the SEC awarded only $60 million to whistleblowers, a steep decline from previous years.[10] One explanation for the decline is increased scrutiny of whistleblower claims. Indeed, the SEC's Office of the Whistleblower FY 2025 report to Congress highlighted frivolous whistleblower award claim submissions and reasons for award denials.[11]

The SEC has similarly de-emphasized enforcement actions against companies for alleged violations of Rule 21F-17 of the Securities Exchange Act of 1934, which prohibits any person from taking any action to impede individuals from contacting the SEC to report a possible securities law violation, including enforcing or threatening to enforce a confidentiality agreement.

However, notwithstanding a moderation in enforcement activity during the current administration, the statute of limitations for fraud under the federal securities laws will outlast this administration, and companies could face continued risk from AI whistleblowers under future administrations—particularly as AI use continues to accelerate across enterprises.

The DOJ's Whistleblower Program—and Attention to AI

In 2024, the DOJ announced its Corporate Whistleblower Awards Pilot Program, which provides eligibility for a presumption of declination to companies that voluntarily self-report as soon as reasonably practicable, but no later than 120 days of receiving an internal whistleblower complaint, and that meet other requirements for voluntary self-disclosure and a declination under the policy—even if the whistleblower submits to DOJ before the company self-discloses.[12] Furthermore, individuals who report truthful information concerning misconduct not known to the government are eligible to receive a portion of the resulting forfeiture, so long as they were not involved in the underlying criminal activity and have no other relevant financial disclosure incentives.

The DOJ's program presents significant reporting issues for companies to consider. For example, because whistleblowers only get credit for reporting conduct that is not already known to the DOJ, employees are more likely to report misconduct to the DOJ without first notifying their companies. Similarly, the credit that companies get for self-reporting misconduct is dependent on reporting wrongdoing that is not already known

[10] Securities and Exchange Commission Office of the Whistleblower Annual Report to Congress for Fiscal Year 2025 (Feb. 11, 2025), https://www.sec.gov/files/fy25-annual-whistleblower-report.pdf.

[11] *Id.*

[12] U.S. Dep't of Justice, Off. of the Deputy Att'y Gen., *Department-Wide Corporate Enforcement and Voluntary Self-Disclosure Policy* (Mar. 10, 2026), https://www.justice.gov/dag/media/1430731/dl.

to the DOJ. Accordingly, the program can create incentives for both companies and whistleblowers to promptly report misconduct to the DOJ before the other does. Additionally, the visibility of the program, as well as the significant awards it offers, may create challenges for companies' efforts to encourage employees to report misconduct via internal channels. The DOJ's pilot program also increases the risk of potential criminal consequences for alleged interference with whistleblower activity.

Since the program's launch, the DOJ's Criminal Division updated its Evaluation of Corporate Compliance Programs (ECCP) guidance to address AI.[13] Prosecutors are now directed to ask how a company is using AI, whether the company has conducted a risk assessment around that use, and what controls exist to test, validate, monitor, and constrain AI systems.[14] Against this evolving regulatory backdrop, AI-related internal complaints can trigger overlapping disclosure, internal controls, and criminal risk considerations.

Civil Claims Risks

Even without SEC or DOJ involvement, individuals who are terminated after raising concerns about insufficient safeguards against AI risks may pursue civil action for retaliation or wrongful termination under state laws, such as California Labor Code §§ 1102.5 and 98.6, or Section 740 of the New York Labor Law.[15] Accordingly, despite the current administration's de-emphasis on whistleblower awards and enforcement, companies should ensure that they maintain sufficient internal whistleblower protections and train management at all levels on how to respond to concerns about potentially unsafe AI practices to minimize retaliation risks.

Practical Tips for Updating Whistleblower Policies and Procedures to Manage AI Risk

Even though the Trump SEC and DOJ have arguably lightened enforcement and whistleblower protection, AI is plainly a priority for both agencies. For this reason, companies deploying AI should proactively review and strengthen their governance, disclosure controls, and internal reporting mechanisms to mitigate AI-related whistleblower exposure, particularly given the rise of agentic AI. Companies should consider adopting the following measures:

[13] U.S. Dep't of Justice, Crim. Div., Fraud Section, *Evaluation of Corporate Compliance Programs* (Sept. 2024), https://www.justice.gov/criminal/criminal-fraud/page/file/937501/dl.

[14] *Id.*

[15] *Gruendel v. Figure AI, Inc.*, 5:25-cv-10094, (N.D. Cal. Nov 21, 2025) ECF No. 1.

- **Substantiating AI Capability Claims:** Assess substantiation, documentation, and review controls for AI-related disclosures (including marketing, fundraising, and investor materials) to mitigate "AI-washing" risk.

- **Accelerating Internal Response Timelines:** Consider whether internal investigation and escalation timelines appropriately account for the incentives created by DOJ's program and related self-reporting considerations.

- **Training**: Train managers involved in AI on relevant whistleblower protections and escalation procedures to mitigate whistleblower risks.

- **Employee or Contractor Agreements**: Review all confidentiality agreements, including severance agreements, releases, codes of conduct, ethics manuals, training materials, and investor materials, for compliance with the Rule 21F-17 requirement not to impede individuals from contacting the SEC to report a possible securities law violation.

- **Addressing Complaints Promptly**: Avoid delays in responding to whistleblowers where practicable so as not to increase the likelihood that whistleblowers will become frustrated and escalate their complaints externally.

- **Taking Concerns Seriously**: Take all whistleblower complaints seriously, including ones that are vague or inflammatory. Even one legitimate concern in an otherwise baseless complaint that is not properly investigated can trigger investigative and enforcement risk.

- **Protecting Whistleblower Anonymity**: If the whistleblower is anonymous, take reasonable measures to protect that anonymity throughout an investigation. If the identity of the whistleblower is known to investigators, it is best practice not to share this identity with others in order to limit the risk of retaliation or investigative taint.

- **Providing Context for Decisions**: Whistleblowers may have valid concerns but lack the broader context for the priorities and competing considerations of their companies. When addressing a whistleblower's concerns, consider providing them with the additional context, when appropriate, on the costs, risks, and business impacts of alternative proposed courses of action, and why those may not be achievable.

- **Consulting Counsel**: Consider involving counsel when faced with complaints regarding alleged violations of law, including those related to AI, especially if any adverse action (including cutting off access to company systems and denying access to company facilities) is being considered against an employee or independent contractor who has raised the concern. Involving outside counsel may also help strengthen privilege claims over the investigation and provide a level of independence.

- **Expert Investigation Team**: Ensure that the investigation team has the necessary AI expertise to evaluate the whistleblower’s allegations or has access to consultants who can assist in that evaluation.

To subscribe to the Data Blog, please click here.

The cover art used in this blog post was generated by Nano Banana Pro 2

* * *

Please do not hesitate to contact us with any questions.

Charu A. Chandrasekhar
Partner, New York
Tel: +1 212 909 6774
cchandrasekhar@debevoise.com

Avi Gesser
Partner, New York
Tel: +1 212 909 6577
agesser@debevoise.com

Arian M. June
Partner, Washington, D.C.
Tel: +1 202 383 8053
ajune@debevoise.com

Cooper Yoo
Associate, Washington, D.C.
Tel: +1 202 383 8039
chyoo@debevoise.com

Sharon Shaji
Law Clerk, New York
Tel: +1 212 909 7458
sshaji@debevoise.com

This publication is for general information purposes only. It is not intended to provide, nor is it to be used as, a substitute for legal advice. In some jurisdictions it may be considered attorney advertising.