Despite Regulation Lag, AI Whistleblowers Have Protections

By **Alexis Ronickher and Matthew LaGarde** (September 11, 2023)

Alexis Ronickher

Matthew LaGarde

The surge of attention on artificial intelligence technology this past year has sparked serious concerns about its impact on society. As AI technologies evolve, discussions surrounding ethical implications, job displacement, bias in algorithms and privacy breaches have intensified, underscoring the need for thoughtful regulation and responsible deployment to ensure society can benefit from the technology while limiting potential harms.

In July, seven leading AI companies based in the U.S. — Amazon.com Inc., Anthropic PBC, Google LLC, Inflection AI Inc., Meta Platforms Inc., Microsoft Corp. and OpenAI — agreed to abide by a set of voluntary commitments, collectively known as the AI commitments, to guide progress within the rapidly developing field.[1]

These voluntary commitments are meant to serve as a stopgap while governments within and outside the U.S. develop regulatory frameworks to govern AI. Given the pace of government action compared to the rate of AI advancement, these AI commitments may be the only safeguards in place during this critical time in the technology's development.

Without binding governmental regulation, however, it will be primarily left to the companies to self-police their compliance with these voluntary commitments, while at the same time competing for market and technological dominance. As history has shown, voluntary safeguards are often ignored in the pursuit of wealth and influence.[2]

One effective tool for ensuring that these companies live up to their word is internal whistleblowers. Employees of these companies will know when the companies are not meeting the AI commitments and the truthfulness of any additional public statements regarding the AI commitments. To the extent the companies are failing to meet their obligations, their employees will know where to find the data and documents that reveal those failures.

When employees report such misconduct — whether internally or to the government — they risk retaliation. How do we empower potential whistleblowers? Given that these AI commitments are voluntary, are there legal protections for these whistleblowers? Are there financial incentives?

What laws might protect whistleblowers at these companies?

In the U.S., there are few laws specifically designed to protect against retaliation in the cybersecurity, data privacy and technology industries. Workers in those industries are instead left with a patchwork of state and federal laws that can be retrofitted to provide protections and incentives for whistleblowers.

Nevertheless, there may be avenues to protection for whistleblowers who report noncompliance with the AI commitments or any related public statements.

On the federal level, the Sarbanes-Oxley Act, or SOX, protects employees of publicly traded companies and certain of their subsidiaries and affiliates from retaliation if they engage in protected activity by reporting specific categories of misconduct, including securities fraud, shareholder fraud or a violation of any U.S. Securities and Exchange Commission rule or regulation.[3]

Since four of the seven companies — Amazon, Google, Meta and Microsoft — that agreed to the AI commitments are publicly traded, SOX would protect employees of those companies who engage in protected activity. Relatedly, if a whistleblower reports potential securities violations to the SEC, the whistleblower would also have engaged in protected activity under the Dodd-Frank Act.[4]

Given the prominence and importance of the companies' voluntary agreement to these AI commitments, there are strong arguments that an employee who reports the companies' knowing failure to meet those commitments has engaged in protected activity under SOX.[5]

Securities law and SEC rules and regulations prohibit fraudulent practices in connection with the purchase or sale of a security, including the knowing misrepresentation or omission of material facts. The AI commitments — as well as any subsequent public statements relating to or arising from the AI commitments — may constitute material facts.

Accordingly, a whistleblower who raised concerns about those misrepresentations — either internally to a supervisor or compliance personnel, or externally to Congress or a regulator — likely engaged in protected activity under SOX, particularly if the whistleblower clearly states that they are concerned about public misrepresentations or fraud.

If a company retaliates against a whistleblower because they engaged in protected activity, the whistleblower would have a claim under the SOX or Dodd-Frank Act anti-retaliation provisions. SOX provides for recovery for economic and emotional distress damages, as well as attorney fees and costs, and the Dodd-Frank Act additionally provides for liquidated damages in the amount of twice the economic losses the whistleblower has suffered, i.e., back pay.

Additionally, a whistleblower with a SOX claim has the right to bring their case in court and the right to a jury trial regardless of whether the whistleblower signed any agreement that would force them into a nonpublic arbitration proceeding or otherwise would have waived a right to a jury trial. Importantly, these rights mean that a company defending itself from a SOX whistleblower claim cannot hide any asserted misconduct from public scrutiny.

While whistleblowers at the three private companies that have agreed to the AI commitments — Anthropic, Inflection and OpenAI — would not be eligible for SOX protections, they may be protected by state law. The employees at the publicly traded companies would also have the same state law protections.

Many states — most importantly California, where five of the companies are headquartered, including all three of the private companies — have general whistleblower statutes that protect employees from retaliation for reporting illegal activity. Many states that do not have such statutes still have common law claims for wrongful discharge in violation of public policy, which can protect employees who report or attempt to stop their employer from violating an existing statutory or constitutional provision.

For example, in states like California that protect employees who report violations of federal law, an employee's report of their employer's failure to comply with the AI commitments could constitute reports of potential violations of the Federal Trade Commission Act. Section 5 of the FTCA prohibits unfair or deceptive acts or practices in commerce.[6]

Under the FTCA, a company's representation is deceptive where it misleads or is likely to mislead a reasonable consumer as to a material fact. The AI commitments, while voluntary, still constitute public representations by the seven companies.

Given the national attention given to the risks associated with AI technology and the high-profile nature of the AI Commitments and their role in mitigating those risks, an employee could reasonably believe that the AI commitments are material facts under the FTCA. Under California law, that reasonable belief would be enough for protection. States like California also have consumer and data protection laws that could serve as potential bases for protection.

State laws generally provide for recovery of economic and emotional harm and attorney fees and costs. Many states, like California, also provide for punitive damages targeted at punishing misconduct and deterring further misconduct. Punitive damages awards for whistleblower retaliation can be significant.[7]

Unfortunately, a whistleblower who has signed a forced arbitration provision would have to arbitrate their state law claims, meaning that the company's misconduct would remain nonpublic, and it would be an arbitrator, not a judge or jury, determining the award.

What laws might reward whistleblowers at these companies?

While protection from retaliation is critical to empowering employees to report misconduct, for many, those protections are not enough to overcome jeopardizing their career and economic security. For this reason, the government has created whistleblower reward programs that can result in financial rewards for whistleblowers who report information that results in the government recovering money.

The SEC whistleblower reward program could be an avenue to whistleblowers receiving compensation for reporting misconduct related to the AI commitments. This program, subject to certain requirements, entitles individuals who provide the SEC with information leading to an enforcement action to an award of 10% to 30% of the amount collected by the commission.

Four of the seven companies that adopted the AI commitments are publicly traded, meaning that misrepresentations related to their AI commitments may constitute securities violations. If an employee reports such misconduct to the SEC and the commission successfully pursues an enforcement action against the company, that employee may be entitled to a share of any proceeds the SEC recovers through that enforcement action.

Looking Forward

While the AI commitments are currently voluntary, many are seeking specific government regulation and oversight of AI technology. As the regulatory environment evolves, so too will the protections and rewards available for AI whistleblowers. That being said, federal law and many state laws currently fail to provide adequate protections for technology whistleblowers.

Given how much we rely on technology whistleblowers to reveal threats to our safety, our democracy and our way of life, we need to do better by them. We need comprehensive whistleblower protections that expressly protect these brave individuals, so that they continue to speak up for the good of others.

Alexis Ronickher is a partner and Matthew LaGarde is a senior associate at Katz Banks Kumin LLP.

[1] See The White House, FACT SHEET: 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/.

[2] See, e.g., PBS, Companies' New Pledges to Boost Recycling Face Old Pitfalls (Mar. 31, 2020), https://www.pbs.org/wgbh/frontline/article/companies-new-pledges-to-boost-recycling-face-old-pitfalls/ (describing Coca-Cola's failure to fulfill promises to increase its use of recycled materials); N.Y. Times, What's Really Behind Corporate Promises on Climate Change? (May 12, 2021), https://www.nytimes.com/2021/02/22/business/energy-environment/corporations-climate-change.html (describing the failure of companies like Levi and Cargill to achieve voluntary carbon reduction promises); Wired, Old Promises Broken, Musk Offers New Pledges on Self-Driving (Apr. 24, 2019), https://www.wired.com/story/promises-broken-musk-offers-new-pledges-self-driving/ (describing Tesla's failure to meet self-imposed self-driving capability goals); Wash. Post, Google promised to delete sensitive data. It logged my abortion clinic visit (May 9, 2023), https://www.washingtonpost.com/technology/2023/05/09/google-privacy-abortion-data/ (finding that "Google still retains location data about users who visit clinics, hospitals and other 'particularly personal' locations, despite Google's commitment to delete it").

[3] 18 U.S.C. §1514A(a)(1).

[4] 15 U.S.C. §78u-6(h).

[5] See supra n.1.

[6] 15 U.S.C. §45.

[7] See, e.g., Wadler v. Bio-Rad Lab'ys, Inc., 916 F.3d 1176, 1191 (9th Cir. 2019) (affirming $5 million punitive damages award in California wrongful discharge claim).