FACT SHEET – FAQ ON ARTIFICIAL INTELLIGENCE NDAs

What law, rule or regulation do OpenAI’s nondisclosure agreements (NDAs) violate or potentially violate?

The Dodd-Frank Act, including SEC Rule 21F-17(a), the Defend Trade Secrets Act, 18 U.S.C. § 1833(b)(1)-(3), and the Sarbanes-Oxley Act’s whistleblower obstruction of justice provision, 18 U.S.C. § 1514A. The SEC Rule states that “[n]o person may take any action to impede an individual from communicating directly with the Commission staff about a possible securities law violation, including enforcing, or threatening to enforce, a confidentiality agreement . . . with respect to such communications.” The SEC regulations cover a wide range of entities from companies traded on the New York Stock Exchange to privately held LLCs.

What provisions in OpenAI’s pre-employment, severance, non-disparagement, and general release agreements violate the rules governing employment contracts?

- NDAs that broadly prohibit the release of “confidential” information to any person – which would include government regulators and law enforcement.

- NDAs that violate the Defend Trade Secrets Act by failing to inform employees/former employees of their statutory right to provide this information to law enforcement and government regulators.

- NDAs that illegally threaten employees/former employees with potential civil or criminal liability for disclosing company-defined “trade secrets” despite a specific federal law that permits trade secret information that may demonstrate a violation of law to federal or state law enforcement or regulatory authorities.

- NDAs that have broad non-disparagement clauses that strictly prohibit derogatory statements with exception only for compulsory disclosures to government agencies, without any exception for voluntary communications with the SEC or government agencies concerning potential violations of law. SEC precedent dictates that non-disparagement agreements without such exceptions for voluntary whistleblower disclosures constitute violations of Rule 21F-17(a).

- Confidentiality clauses that require prior consent from the company to make disclosures to government agencies.

- General release provisions that require employees to waive their rights to whistleblower rewards intended by Congress to incentivize reporting of securities violations.

- Non-disclosure clauses that threaten civil and criminal action for breaches of trade secret confidentiality requirements in plain violation of the Defend Trade Secrets Act of 2016 which explicitly carves out an exemption for whistleblowers making reports to the government, immunizes them from civil and criminal liability, and requires employers to notify all employees to this immunity.
What relief can the SEC order to correct these NDA-violations committed by OpenAI?

- Requiring companies to stop using illegal NDAs;
- Requiring companies to inform all current and past employees that the improper NDAs are void and will not be enforced;
- Requiring companies to pay significant fines and penalties for violating the NDA regulations, even if companies have remedied past violations;
- Take specific enforcement actions against companies that have enforced such agreements or otherwise retaliated against whistleblowers;
- Compensate whistleblowers who report illegal NDAs or other regulatory violations under the award provisions of the Dodd-Frank Act.

Do employees working on AI technologies, such as ChatGPT, AGI, among others, have the right to “blow the whistle” or report illegal NDAs to the proper authorities?

The safest method to report safety concern or regulatory violations within AI is to file an anonymous and confidential complaint with either the SEC or the Commodity Futures Trading Commission pursuant to the procedures set forth in the Dodd-Frank Act. Other laws that may also apply. See FAQ on Dodd-Frank Whistleblower Protections and Rewards.

Why are whistleblower rights within the AI space so important?

The potential threats to the public interest are well documented. As explained by the various government agencies, AI companies, and experts below:

- **US government**: “irresponsible use could exacerbate societal harms such as fraud . . . and pose risks to national security.”

- **UK government**: “[AI systems] could also further concentrate unaccountable power into the hands of a few, or be maliciously used to undermine societal trust, erode public safety, or threaten international security . . . [AI could be misused] to generate disinformation, conduct sophisticated cyberattacks or help develop chemical weapons.”

- **OpenAI**: “AGI would also come with serious risk of misuse, drastic accidents, and societal disruption.”

- **Anthropic**: “[R]apid AI progress would be very disruptive, changing employment, macroeconomics, and power structures . . . [we have already encountered] toxicity, bias, unreliability, dishonesty”

- **Google DeepMind**: “it is plausible that future AI systems could conduct offensive cyber operations, deceive people through dialogue, manipulate people into carrying out harmful actions, develop weapons (e.g. biological, chemical), . . . due to failures of alignment, these AI models might take harmful actions even without anyone intending so.”
- **Bletchley Declaration**: “There is potential for serious, even catastrophic, harm.”

- **Encode Justice and the Future of Life Institute**: “we find ourselves face-to-face with tangible, wide-reaching challenges from AI like algorithmic bias, disinformation, democratic erosion, and labor displacement.”

- **Statement on AI Risk (CAIS)**: “Mitigating the risk of extinction from AI should be a global priority alongside other societal-scale risks such as pandemics and nuclear war.”