

Related Regulations and Standards	California SB 53 § 1107.1. (a); Illinois SB 315 § 20(a) Anti-retaliation protection for a covered employee to report to authorities if the information discloses that (1) the frontier developer's activities pose a specific and substantial danger to the public health or safety resulting from a catastrophic risk; or (2) the frontier developer has violated this Act. California SB 53 § 1107.1(e)(1); Illinois SB 315 § 20(e)(1) Requirement for an internal anonymous-disclosure channel for covered employee to raise good-faith concern that the developer's activities either (a) present a specific and substantial danger to public health or safety from a catastrophic risk, or (b) violate the relevant frontier-AI statute (here, Chapter 25.1 of Division 8 of the Business and Professions Code). The developer must give the discloser monthly status updates on the investigation and on what actions the company has taken in response.
Anthropic	Anthropic has a public-facing whistleblowing policy “Responsible Scaling Policy (RSP) Noncompliance and Anti-Retaliation Policy” (Last updated: Feb 2026). [Anthropic, 2026] It includes aspects of covered violations, reporting mechanism, report content and documentation, investigation mechanism for solutions, as well as confidentiality and no retaliation protection. The quality of the whistleblower policy will be addressed in the indicator below.
OpenAI	OpenAI has a public-facing whistleblowing policy (“OpenAI Raising Concerns Policy”). (Last updated: Jan, 2026) [OpenAI, 2026] It includes aspects of covered violations, reporting mechanism (including Integrity Line), investigation mechanism for solutions, as well as confidentiality and no retaliation protection. The quality of the whistleblower policy will be addressed in the indicator below.
Google DeepMind	Google DeepMind doesn't have a public-facing policy. Google Code of Conduct delineates channels through which employees can raise their concerns towards different parties and the scope covered by such reporting. These concerns include a “no retaliation” clause. These measures apply to both employees and the extended workforce. [Google] Google shared more details about their whistleblowing policy in the company survey. The quality of the public information of the whistleblower policy will be addressed in the indicator below.
Meta	Meta has a public harassment policy publicly available, a Whistleblower and Complaint Policy available to covered individuals, and is currently developing further protocols on AI-related risks, according to the company survey response. The protocols under development aim to report any instances of non-compliance with its Advanced AI Scaling Framework, and any specific and substantial danger to the public health or safety arising from catastrophic risk. Under this protocol, employees will be able to confidentially, and, if they choose, anonymously issue reports through internal channels, and all reports will be ultimately submitted to the internal governance function, the Chief AI Officer, and the Director of Alignment and Risk. Its Code of Conduct referenced a Whistleblower and Complaint Policy, but it is not linked and not publicly retrievable. The Code delineates channels through which employees can raise their concerns, the mechanisms of investigation that follows, and “no retaliation” protections. Integrity line and harassment policy are available and linked. Meta shared more details about their whistleblowing policy in the company survey. The quality of the public information of the whistleblower policy will be addressed in the indicator below.
xAI	xAI doesn't have a public-facing policy. However, xAI has stated that its employees have “whistleblower protections enabling them to raise concerns to relevant government agencies regarding imminent threats to public safety.” Moreover, it has shared in AI Safety Index's Winter 2025 company survey more details, including the role designated to oversee the whistleblowing function, the investigative independence, the scope of policy, “no retaliation” and “confidentiality” protections towards employees, the reporting mechanisms etc. The quality of the whistleblower policy will be addressed in the indicator below.
DeepSeek	No public-facing whistleblower policy found.
Z.ai	No public-facing whistleblower policy found. Z.ai skipped the whistleblower policy section in the company survey.
Alibaba Cloud	No public-facing whistleblower policy found. Its Code of Ethics states that employees have established whistleblower rules and procedures that are subject to update from time to time. The covered topics include violations of applicable laws or regulations, the Code, or Alibaba Group's related policies. Employees should report relevant information to the Compliance Officer. “No-retaliation” protection applies here.
Mistral	No public-facing whistleblower policy found. The company states that “it has established a formalized whistleblower policy, and an anonymous communication channel is in place for users to report potential issues or fraud concerns.” [Mistral]