OpenAI

# OpenAI Raising Concerns Policy

I. Overview and Purpose 1
II. Scope 2
III. Definitions 2
IV. Policy 3
A. Our Responsibility to Speak Up 3
B. Kinds of Concerns 4
1. Potential Misconduct 4
2. AI Safety Concerns 5
C. No Retaliation for Raising Concerns 5
D. External Reporting 5
E. How to Raise Concerns 6
F. Confidentiality 7
G. How We Address Reports 7
1. Misconduct 7
2. AI Safety Concerns 8
V. Related Policies and Resources 8
VI. Questions? 9
VII. Revision History 9
Country Appendix 10

## I. Overview and Purpose

Open communication and constructive debate are essential to OpenAI's culture, integrity, and ability to fulfill our mission.  We expect and encourage concerns to be raised through normal, good-faith dialogue whenever possible.

However, if you see or experience something that feels off —something that might violate a law, our policies, or that could create a serious health or safety risk to our customers, society, or OpenAI—you should speak up. Raising concerns helps safeguard the integrity of the organization and supports the responsible development of our technology.

OpenAI does not tolerate Retaliation for raising concerns in good faith or participating in an investigation. There are methods for raising concerns anonymously, and we handle reports with care and discretion, involving only those with a legitimate need to know.

This Policy reinforces our commitment to protecting those who speak up, and provides clear structure and escalation paths for raising concerns. See the **Country Appendix** for country-specific terms.

## II. Scope

This Policy applies to all OpenAI Personnel, and applies equally to concerns about workplace conduct, business practices, and the safety or societal impact of OpenAI's technology.

## III. Definitions

**OpenAI (or "Company" or "we"):** OpenAI and its subsidiaries.

**OpenAI Personnel (or "you"):** OpenAI employees, agents, interns, members of OpenAI's extended workforce, and anyone else providing services to OpenAI.

**Misconduct:** Conduct that violates law, regulation, or OpenAI policy, or that presents health and safety risk in the workplace, including but not limited to fraud, harassment, discrimination, Retaliation, or misuse of company property, or intentional violations of AI policies. Misconduct does not include purely interpersonal issues unless they raise policy or legal violations.

**AI Safety Concerns:** Any good-faith concern that OpenAI's AI technology products, models, or services could pose a risk of harm to our customers, third parties or society. Examples include unsafe AI model behavior, misuse potential, or inadequate mitigations.

**Retaliation.** Any adverse action taken against you because you exercised your rights under this policy or intended to discourage you or others from raising concerns in good faith. Retaliation can be direct and obvious, such as attempting to terminate your employment, demoting you, reducing your responsibilities, or engaging in harassment. It can also take more subtle forms, such as a pattern of intentionally isolating you or singling you out for negative treatment because you raised a concern. Retaliation does not include honest feedback or performance management.

## IV. Policy

### A. Our Responsibility to Speak Up

We all play a role in protecting OpenAI's integrity and mission. If you notice, experience, or have good faith concerns about possible Misconduct—by OpenAI Personnel or by third parties we work with, such as suppliers or partners—we ask you to raise it.

Additionally, if you have good faith AI Safety Concerns that are not being addressed through discussion and collaboration and need further attention or escalation, we ask you to raise them.

OpenAI works at the frontier of fast-evolving science and technology. We expect, value, and protect robust and good faith internal discussion, disagreement, and debate about what the right decisions are. Many important issues come to light because someone shared a concern early.

At times, decisions will feel complex or hard to resolve, and you may not fully agree with the outcome. In those moments, we ask you to pause and reflect: was the decision made with good intent, grounded in sound reasoning, and informed by meaningful debate? If so, disagreement alone does not mean something has gone wrong.

At the same time, if a decision or process raises concerns about safety, integrity, compliance, or our ability to carry out our mission, it's important to surface those concerns. Raising concerns can help teams surface risk, test assumptions, and make better decisions together. We expect concerns to be raised thoughtfully, constructively, and in a timely way, using the channels outlined below. OpenAI will protect you from Retaliation for raising concerns in good faith.

OpenAI managers have additional responsibilities as leaders. They are expected to listen thoughtfully, take concerns seriously, promptly escalate issues through the appropriate reporting pathways (see Section E), and foster an environment where people feel safe and supported in speaking up—especially when they're unsure or uncomfortable.

## B. Kinds of Concerns

### 1. Potential Misconduct

The Company's Behavior Spec, Employee Handbooks, and company policies explain the behavior we expect from OpenAI Personnel, and the Supplier Code of Conduct does the same for our suppliers.[1] We expect you to tell us about conduct that might violate those policies.[2] Common categories of Misconduct with non-exhaustive examples include:

- **Financial Misconduct:** circumventing accounting procedures or internal controls, gross mismanagement or negligence.
- **Business Records:** failing to keep records subject to legal holds, using non-approved communication channels.
- **Security:** violating Security policies, including improper system access, bypassing access controls, unauthorized creation/use of alternative communication channels, mishandling restricted data, or exfiltration of internal information.
- **Confidential Information:** unauthorized disclosure to third parties, including media leaks; using a former employer's confidential documents at OpenAI.
- **Theft:** taking OpenAI's documents outside of our IT system, unless protected by law; using OpenAI's information, resources, money, or other property for personal purposes.
- **Privacy:** misuse or improper handling of OpenAI personnel, customer, or user data.
- **Corruption:** fraud, embezzlement, money laundering, bribery.
- **Conflicts of Interest:** self-dealing, undisclosed activities, or relationships.
- **Employment:** bullying, harassment, or discrimination.
- **Environment, Health, and Safety:** unsafe working conditions, workplace violence, violation of applicable environmental regulations.
- **AI Safety Policy Violations:** violation of laws or regulation, or intentional violations of OpenAI policies, related to the safe development, testing, or use of AI.
- **Retaliation** (see Section III. Definitions)**.**

We welcome reports of types of potential Misconduct not listed above. If you are not sure if the behavior you observed rises to the level of Misconduct, tell us anyway. Early reporting helps us address issues before they become a bigger problem.

[1] Our extended workforce may be governed by their employer's Code of Conduct and related policies of their employer rather than or in addition to OpenAI's policies.

[2] If you are a member of our extended workforce, report concerns to your primary employer first if they don't involve OpenAI employees and are not an emergency. Your employer will let us know as appropriate (for example, if the alleged conduct happened at our offices or threatened our business or employees).

### 2. AI Safety Concerns

Raising AI Safety concerns is appropriate and encouraged, and is a core part of OpenAI's mission and culture. This Policy supports that culture by helping people raise concerns thoughtfully and in good faith—whether through discussion, collaboration, or appropriate escalation—so risks can be understood and addressed early. Examples of such concerns may include:

- Noncompliance with the Preparedness Framework
- A model producing unsafe, biased, or misleading outputs that could cause harm.
- Features that could enable misinformation, privacy violations, or fraud.
- Gaps in alignment, red-teaming, or launch processes.
- Insufficient mitigations for known misuse or safety risks.
- Emerging risk from new model capabilities.
- Weaknesses in data governance, monitoring, or rollout safety protocols.

## C. No Retaliation for Raising Concerns

OpenAI strictly prohibits Retaliation against anyone who raises concerns in good faith. Retaliation for speaking up, assisting someone else in raising a concern, participating in an investigation, or refusing to participate in misconduct is not tolerated and may result in discipline, up to and including termination of employment or engagement.

What does "good faith" mean? It means you have a reasonable belief that (a) Misconduct may have occurred, or (b) that an AI Safety Concern needs to be further addressed. Raising concerns that you know to be false or misleading is considered bad faith and is not protected.

## D. External Reporting

Our priority is to ensure that OpenAI Personnel have accessible, safe, and low-friction ways to raise concerns internally. We strongly encourage raising concerns through our internal channels so we can understand issues early, learn from them, and address them thoughtfully.

That said, you are always free to report concerns to an outside authority. You do not need to notify OpenAI in advance, and you are protected from Retaliation for doing so in good faith.

Outside authorities include regulatory and law enforcement agencies such as the Equal Employment Opportunity Commission ("EEOC"), US National Labor Relations Board ("NLRB"), US Securities and Exchange Commission ("SEC"); legislative authorities such as Congress; or any other national, federal, state or local agency or governmental body charged with the enforcement of any laws or regulations or authorized to receive reports under applicable law. For example, you can use the California Attorney General's hotline to provide information regarding possible violations of state or federal statutes, rules, or regulations, or violations of fiduciary responsibility, including reports regarding OpenAI's failure to comply with the California Transparency in Frontier Artificial Intelligence Act or reports of potentially catastrophic safety issues.

### E. How to Raise Concerns

**Misconduct Concerns:** You can raise your Misconduct concerns in the manner most comfortable to you. If the person or team you reach out to can't resolve it, they'll pass it along to someone who can. For example, you may report a concern about Misconduct to:

- Your manager (or if you are an extended worker, your OpenAI contact)
- Compliance
- Employee Relations or HR
- Legal
- OpenAI's Integrity Line (allows for oral and written anonymous reporting)
- Global Security Investigations
- Grievance Committee (for some international locations)

**AI Safety Concerns**: AI Safety Concerns are often best raised and discussed through normal business channels, including during research, product development, launches, and safety or policy discussions. This is because evaluating, discussing and resolving such concerns can be a high context endeavor. We encourage those conversations and rely on them as a core part of how we build safe AI.

However, we want to provide additional channels if you're unsure how to raise a concern, feel uncomfortable doing so through normal channels, or believe an issue needs additional attention. At any time, you can escalate your concern to Compliance or through the Integrity Line, which the Compliance team manages. The Integrity Line allows you to raise concerns anonymously and allows us to communicate with you while maintaining your anonymity.

Why Compliance? The Compliance team manages the Integrity Line and oversees the review process for reported concerns. This includes tracking each concern, coordinating with relevant subject matter experts, helping ensure timely consideration, and communicating outcomes to the extent appropriate.

### F. Confidentiality

Any information you provide (including the names of those involved) is shared only with a limited number of people on a strict need-to-know basis. If you report anonymously through our Integrity Line, we have no way to access the metadata to determine your identity, although we may ask for more information or if you would meet with us if it would help fact-finding, which would be your choice. Those requests would be made through the Integrity Line in a way that maintains your anonymity.

### G. How We Address Reports

#### 1. Misconduct

We take every report seriously, and are committed to handling them independently, timely, and in a manner that complies with all relevant laws and policies and is fair to complainants and subjects. To help you understand our process, here is a quick overview:

**Acknowledge:** We aim to acknowledge receipt of your report within 2 business days.

**Assign**: Next, we route your report to the right team, after confirming no conflicts of interest exist. If your report involves a credible allegation of a potential legal or policy violation, one of our investigation teams — typically, Compliance, Employee Relations, Security Legal, or Security — will take the lead, often working with HR. If a report is too vague, we might not be able to investigate further. We may also notify OpenAI's Audit Committee depending on the seriousness of the allegations.

**Fact-Finding**: After assignment, the investigator gathers documents and interviews those involved to understand the facts.

**Analyze:** Once we have all of the facts, the investigator or the appropriate team will try to determine if a legal or policy violation occurred.

**Align:** The investigator will discuss their findings with need-to-know partners like HR and then report to the appropriate manager for alignment on outcome and any disciplinary action. We also consider improvements to related policies and processes.

**Close**: We will let you know once the issue has been resolved, and, absent extenuating circumstances, will provide an update no later than 30 days after the initial report. We will follow up with you about the outcome as transparently as we can while still protecting privacy and legal obligations.

**Check**: For certain types of concerns, including Retaliation, discrimination, harassment, and hostile workplace, we may reach out to you 30-45 days after the investigation closes to see if you need any additional support or have been the subject of Retaliation. You should also let us know if you feel that someone is retaliating against you for raising a concern, assisting someone else in raising a concern, or participating in an investigation.

### 2. AI Safety Concerns

When you escalate an AI Safety Concern, the goal is to determine whether the right decision makers are making or have made an AI Safety decision based on sufficient information, and to further escalate or inform as necessary.

The Compliance team will review reports of AI Safety Concerns with relevant subject matter experts to understand the issue and determine the appropriate next steps. Depending on the nature of the concern, the Compliance team working with subject matter experts may conclude that the concern raises a business or policy decision that has been or is being made with appropriate stakeholder involvement, or may escalate the issue to senior leadership, the Safety Advisory Group (SAG) or the Safety and Security Committee (SSC) of the Company's Board of Directors as appropriate. To be clear, the aforementioned process is not considered a formal investigation in nature, as the goal is to enable OpenAI to make the best safety decisions by providing additional channels for OpenAI Personnel to raise safety concerns for input and consideration.

If you raise a concern through the Integrity Line anonymously, we will respect your request for anonymity. Compliance will follow up with you and let you know when the review process has concluded. While we provide employees with this opportunity to raise concerns, because AI Safety Concerns can be subject to good faith disagreements, not every concern may be resolved to the employee's complete satisfaction.

## V. Related Policies and Resources

- Integrity Line
- Behavior Spec
- Employee Handbooks

- Supplier Code of Conduct

## VI. Questions?

Please reach out to Compliance.

## Country Appendix

This appendix addresses additional country-specific legal requirements applicable to OpenAI. If a country in which OpenAI is operating is not listed in this Appendix it is because there is no additional law, the law does not apply to us, or the law does not include any additional terms. Unless otherwise stated, the clauses below supplement the language above.

**European Union (EU)**
*(Applies in EU Member States with 50 or more employees)*

Section II — Scope

- Where required by EU whistleblower protection laws, including Directive (EU) 2019/1937, the protections in this Policy extend to all categories of persons protected under Article 4 of the Directive, including employees, former employees, job applicants, contractors, subcontractors, suppliers, shareholders, board members, volunteers, trainees, facilitators, and legal entities owned by or connected to the reporting person.

Section IV.B.1 (Kinds of Concerns, Potential Misconduct) — EU Material Scope

- In the EU, this Policy protects good-faith reports relating to potential breaches of applicable EU law, including where there is uncertainty about the interpretation, application, or compliance status of such law.

Section IV.D (External Reporting)

- In addition to the external reporting options described in this Policy, in the EU you may also report concerns to competent national or EU authorities designated under applicable law.

- Under EU whistleblower protection laws, public disclosure of concerns (for example, to the media or civil society organizations) may be protected only in limited circumstances defined by law. These circumstances may include situations where:
  - you have reported the concern internally and/or to a competent authority, and no appropriate action was taken within the timeframes required by law; or
  - you reasonably believe the concern presents an imminent or manifest danger to the public interest; or
  - you reasonably believe that reporting to a competent authority would expose you to retaliation or that evidence may be concealed or destroyed.

- Outside of these limited circumstances, public disclosure may not be protected.

Section IV.E (How to Raise Concerns)

- Concerns may also be raised orally, including by telephone or other voice-based reporting channels. Upon request, a reporting person may ask to meet with Compliance to discuss a concern, and such a meeting will be arranged within a reasonable timeframe, consistent with applicable law.

Section IV.F (Confidentiality) — Protection of Identity

- OpenAI will not disclose the identity of a reporting person, or information from which their identity may be directly or indirectly deduced, to anyone beyond authorized personnel without the reporting person's explicit consent, except where disclosure is required by law. Where legally permitted, OpenAI will notify the reporting person before any such disclosure.

Section IV.G (How We Address Reports: Assign)

- For purposes of EU whistleblower protection laws, Compliance serves as the impartial function responsible for oversight of the receipt and follow-up of reports under this Policy, including coordination with other designated functions as appropriate.

Section IV.H — GDPR

- Any processing of personal data relating to EU whistleblower reports will be carried out in accordance with Regulation (EU) 2016/679 (GDPR) and applicable data-protection law.

**India**

Section IV.D (No Retaliation)

- The non-retaliation protections in this Policy apply fully to anyone participating in an IC process.

Section IV.F (How to Raise Concerns)

- In addition to the reporting channels listed in this policy, you may raise a sexual harassment concern directly to the Internal Committee (IC). If a report involves sexual harassment, it will be referred to the IC for handling in accordance with the Prevention of Sexual Harassment (PoSH) of Women at Workplace Act, 2013 .

Section IV.E (External Reporting)

- You may also raise sexual harassment concerns to the Local Complaints Committee (LCC) established by the District Officer.

Section IV.H (How We Address Concerns)

- For sexual harassment concerns, the IC will conduct the inquiry under PoSH procedures.
- During an IC inquiry, you may request temporary adjustments, such as changes to reporting structure, work location, or leave. Requests will be assessed in accordance with the PoSH Act.

**Ireland** (supplement to EU provisions)

Section IV.B (Examples of Potential Concerns):

- The types of issues that qualify as “relevant wrongdoings” under Irish law should also include:
  - Criminal offences
  - Misuse of public funds
  - Miscarriage of justice
  - Gross mismanagement or negligence
  - Actions endangering health, safety, or the environment
  - Attempts to conceal any of the above

Section IV.D (External Reporting)

- In Ireland, you may also make a report to the Protected Disclosures Commissioner, Data Protection Commission, or Central Bank of Ireland, or to another regulator authorized to receive such reports. For more information, visit the Office of Protected Disclosures Your Questions page.

**Korea**

Section IV.G (Confidentiality)

- We may need to confirm your identity when you report workplace or sexual harassment. We’ll only do so when necessary and will keep that information confidential.

**United Arab Emirates (UAE)**

Section IV.D (External Reporting)

- In the UAE, you may also raise concerns directly with any relevant UAE regulatory or governmental authority without notifying OpenAI first. This may include, depending on the issue:
  - Ministry of Human Resources and Emiratisation (MOHRE)
  - UAE Police or Public Prosecution
  - UAE Data Office
  - Financial services or free-zone regulators (e.g., DFSA, ADGM FSRA) where applicable

  You are protected from Retaliation for making such reports in good faith.

- For employment-related concerns, you may also access UAE dispute-resolution bodies or mediation channels. Using these external channels does not affect your rights under this Policy.

Section IV.G (Confidentiality)

- Consistent with UAE law, OpenAI will maintain confidentiality of your identity when you raise a concern. In limited situations, however, we may be required to disclose your identity to government authorities as part of an official investigation. Where possible and legally permitted, we will notify you before doing so.