Responsible Use of Anthropic's Models: Guidelines for Organizations Serving Minors

Updated this week

At Anthropic, we recognize the unique vulnerabilities and needs of children in digital spaces. In order to create a safer digital environment and mitigate risks, organizations providing minors with the ability to directly interact with products that incorporate our API(s) should implement the following safeguards:

1. Additional Technical Measures

Organizations with products serving minors should implement additional safety features tailored to their unique use cases, as they are best situated to understand the specific ways their end users may interact with products that incorporate Anthropic's services. These safety measures may include, but are not limited to:

Age verification systems to ensure only intended users can access the product

Content moderation and filtering to block inappropriate or harmful content

Monitoring and reporting mechanisms to identify and address potential issues

Educational resources and guidance for minors on safe and responsible use of the product

In addition to these organization-specific measures, Anthropic may make available technical measures intended to tailor product experiences for certain end users, including minors. For example, we may provide a child-safety system prompt, which organizations serving minors should implement as part of a comprehensive suite of safety measures. It is important to note that, while helpful, these technical measures are not infallible and should be used in conjunction with the organization's own safety features to ensure a robust approach to child safety.

2. Regulatory Compliance

It is the responsibility of organizations to comply with all applicable child safety and data privacy regulations, such as the Children's Online Privacy Protection Act (COPPA) in the United States. Compliance with these regulations should be clearly stated on the organization’s website or similar public-facing documentation.

3. Disclosure Requirements

Organizations must disclose to their users that they are interacting with an AI system rather than a human.

Anthropic will periodically audit organizations for compliance with these safeguards. If your organization has a high violation rate and has not implemented these safety recommendations, we may ask you to implement them. Failure to implement these recommendations when requested, or a continued high violation rate, may lead to the suspension or termination of your account.

Usage Policy

Effective September 15, 2025Previous Version

English

Our Usage Policy (also referred to as our “Acceptable Use Policy” or “AUP”) applies to anyone who can submit inputs to Anthropic’s products and/or services, including via any authorized resellers or passthrough access, all of whom we refer to as “users.” The Usage Policy is intended to help our users stay safe and promote the responsible use of our products and services.

The Usage Policy is categorized according to who can use our products and for what purposes. We will update our policy as our technology and the associated risks evolve or as we learn about unanticipated risks.

- Universal Usage Standards: Our Universal Usage Standards apply to all users and use cases.
- High-Risk Use Case Requirements: Our High-Risk Use Case Requirements apply to specific consumer-facing use cases that pose an elevated risk of harm.
- Additional Use Case Guidelines: Our Additional Use Case Guidelines apply to certain other use cases, including consumer-facing chatbots, products serving minors, agentic use, and Model Context Protocol servers.

Anthropic's Safeguards Team will implement detection and monitoring to enforce our Usage Policy, so please review this policy carefully before using our products or services. If we learn that you have violated our Usage Policy, we may throttle, suspend, or terminate your access to our products and services. We may also block or modify model outputs when inputs violate our Usage Policy.

If you believe that our model outputs are potentially inaccurate, biased or harmful, please notify us at usersafety@anthropic.com, or report it directly in our product through the "report issues" thumbs down button or similar feedback features (where available). You can read more about our Safeguards practices and recommendations in our Safeguards Support Center.

This Usage Policy is calibrated to strike an optimal balance between enabling beneficial uses and mitigating potential harms. Anthropic may enter into contracts with certain governmental customers that tailor use restrictions to that customer's public mission and legal authorities if, in Anthropic's judgment, the contractual use restrictions and applicable safeguards are adequate to mitigate the potential harms addressed by this Usage Policy.

Universal Usage Standards

Do Not Violate Applicable Laws or Engage in Illegal Activity

-
-
-
-

Do Not Compromise Critical Infrastructure

-
-
-

Do Not Compromise Computer or Network Systems

-
-
-
-
-
-
-
-

Do Not Develop or Design Weapons

This includes using our products or services to:

- Produce, modify, design, or illegally acquire weapons, explosives, dangerous materials or other systems designed to cause harm to or loss of human life
- Design or develop weaponization and delivery processes for the deployment of weapons
- Circumvent regulatory controls to acquire weapons or their precursors
- Synthesize, or otherwise develop, high-yield explosives or biological, chemical, radiological, or nuclear weapons or their precursors, including modifications to evade detection or medical countermeasures

Do Not Incite Violence or Hateful Behavior

-
-
-
-

Do Not Compromise Privacy or Identity Rights

-
-
-

Do Not Compromise Children’s Safety

-
-
-
-
-
-

Do Not Create Psychologically or Emotionally Harmful Content

-
-
-
-
-
-
-

Do Not Create or Spread Misinformation

-
-
-

-
-

Do Not Undermine Democratic Processes or Engage in Targeted Campaign Activities

-
-
-
-
-
-
-
-

Do Not Use for Criminal Justice, Censorship, Surveillance, or Prohibited Law Enforcement Purposes

-
-
-
-
-
-
-

Do Not Engage in Fraudulent, Abusive, or Predatory Practices

-
-
-
-
-
-
-
-
-

-
-
-
-

Do Not Abuse our Platform

-
-
-
-
-
-

Do Not Generate Sexually Explicit Content

-
-
-
-

High-Risk Use Case Requirements

Some use cases pose an elevated risk of harm because they influence domains that are vital to public welfare and social equity. For these use cases, given potential risks to individuals and consumers, we believe that relevant human expertise should be integrated and that end-users should be aware when AI has been involved in producing outputs.

As such, for the “High-Risk Use Cases” described below, we require that you implement these additional safety measures:

- Human-in-the-loop: When using our products or services to provide advice, recommendations, or in subjective decision-making directly affecting individuals or consumers, a qualified professional in that field must review the content or decision prior to dissemination or finalization. You or your organization are responsible for the accuracy and appropriateness of that information.

- Disclosure: If model outputs are presented directly to individuals or consumers, you must disclose to them that you are using AI to help produce your advice, decisions, or recommendations. This disclosure must be provided at a minimum at the beginning of each session.

“High-Risk Use Cases” include:

- Legal: Use cases related to legal interpretation, legal guidance, or decisions with legal implications
- Healthcare: Use cases related to healthcare decisions, medical diagnosis, patient care, therapy, mental health, or other medical guidance. Wellness advice (e.g., advice on sleep, stress, nutrition, exercise, etc.) does not fall under this category
- Insurance: Use cases related to health, life, property, disability, or other types of insurance underwriting, claims processing, or coverage decisions
- Finance: Use cases related to financial decisions, including investment advice, loan approvals, and determining financial eligibility or creditworthiness
- Employment and housing: Use cases related to decisions about the employability of individuals, resume screening, hiring tools, or other employment determinations or decisions regarding eligibility for housing, including leases and home loans
- Academic testing, accreditation and admissions: Use cases related to standardized testing companies that administer school admissions (including evaluating, scoring or ranking prospective students), language proficiency, or professional certification exams; agencies that evaluate and certify educational institutions
- Media or professional journalistic content: Use cases related to using our products or services to automatically generate content and publish it for external consumption

Additional Use Case Guidelines

The below use cases – regardless of whether they are High-Risk Use Cases – must comply with the additional guidance provided.

- All consumer-facing chatbots, including any external-facing or interactive AI agent, must disclose to users that they are interacting with AI rather than a human. This disclosure must be provided at a minimum at the beginning of each chat session.
- Products serving minors, including organizations providing minors with the ability to directly interact with products that incorporate our API(s), must comply with the additional guidelines outlined in our Help Center article.
- Agentic use cases must still comply with the Usage Policy. We provide examples of Usage Policy prohibitions in the context of agentic use in this Help Center article.
- Model Context Protocol (MCP) servers listed in our Connector Directory must comply with our Directory Policy.

Usage Policy
Effective September 15, 2025
Previous Version
English
Our Usage Policy (also referred to as our "Acceptable Use Policy" or "AUP") applies to anyone who can submit inputs to Anthropic's products and/or services, including via any authorized resellers or passthrough access, all of whom we refer to as "users." The Usage Policy is intended to help our users stay safe and promote the responsible use of our products and services.

The Usage Policy is categorized according to who can use our products and for what purposes. We will update our policy as our technology and the associated risks evolve or as we learn about unanticipated risks.

Universal Usage Standards: Our Universal Usage Standards apply to all users and use cases.
High-Risk Use Case Requirements: Our High-Risk Use Case Requirements apply to specific consumer-facing use cases that pose an elevated risk of harm.
Additional Use Case Guidelines: Our Additional Use Case Guidelines apply to certain other use cases, including consumer-facing chatbots, products serving minors, agentic use, and Model Context Protocol servers.
Anthropic's Safeguards Team will implement detection and monitoring to enforce our Usage Policy, so please review this policy carefully before using our products or services. If we learn that you have violated our Usage Policy, we may throttle, suspend, or terminate your access to our products and services. We may also block or modify model outputs when inputs violate our Usage Policy.

If you believe that our model outputs are potentially inaccurate, biased or harmful, please notify us at usersafety@anthropic.com, or report it directly in our product through the “report issues” thumbs down button or similar feedback features (where available). You can read more about our Safeguards practices and recommendations in our Safeguards Support Center.

This Usage Policy is calibrated to strike an optimal balance between enabling beneficial uses and mitigating potential harms. Anthropic may enter into contracts with certain governmental customers that tailor use restrictions to that customer’s public mission and legal authorities if, in Anthropic’s judgment, the contractual use restrictions and applicable safeguards are adequate to mitigate the potential harms addressed by this Usage Policy.

Universal Usage Standards

Do Not Violate Applicable Laws or Engage in Illegal Activity

Do Not Compromise Critical Infrastructure

Do Not Compromise Computer or Network Systems

Do Not Develop or Design Weapons

This includes using our products or services to:

Produce, modify, design, or illegally acquire weapons, explosives, dangerous materials or other systems designed to cause harm to or loss of human life
Design or develop weaponization and delivery processes for the deployment of weapons
Circumvent regulatory controls to acquire weapons or their precursors
Synthesize, or otherwise develop, high-yield explosives or biological, chemical, radiological, or nuclear weapons or their precursors, including modifications to evade detection or medical countermeasures

Do Not Incite Violence or Hateful Behavior

Do Not Compromise Privacy or Identity Rights

Do Not Compromise Children’s Safety

Do Not Create Psychologically or Emotionally Harmful Content

Do Not Create or Spread Misinformation

Do Not Undermine Democratic Processes or Engage in Targeted Campaign Activities

Do Not Use for Criminal Justice, Censorship, Surveillance, or Prohibited Law Enforcement Purposes

Do Not Engage in Fraudulent, Abusive, or Predatory Practices

Do Not Abuse our Platform

Do Not Generate Sexually Explicit Content

High-Risk Use Case Requirements

Some use cases pose an elevated risk of harm because they influence domains that are vital to public welfare and social equity. For these use cases, given potential risks to individuals and consumers, we believe that relevant human expertise should be integrated and that end-users should be aware when AI has been involved in producing outputs.

As such, for the “High-Risk Use Cases” described below, we require that you implement these additional safety measures:

Human-in-the-loop: When using our products or services to provide advice, recommendations, or in subjective decision-making directly affecting individuals or consumers, a qualified professional in that field must review the content or decision prior to dissemination or finalization. You or your organization are responsible for the accuracy and appropriateness of that information.

Disclosure: If model outputs are presented directly to individuals or consumers, you must disclose to them that you are using AI to help produce your advice, decisions, or recommendations. This disclosure must be provided at a minimum at the beginning of each session.

“High-Risk Use Cases” include:

Legal: Use cases related to legal interpretation, legal guidance, or decisions with legal implications

Healthcare: Use cases related to healthcare decisions, medical diagnosis, patient care, therapy, mental health, or other medical guidance. Wellness advice (e.g., advice on sleep, stress, nutrition, exercise, etc.) does not fall under this category

Insurance: Use cases related to health, life, property, disability, or other types of insurance underwriting, claims processing, or coverage decisions

Finance: Use cases related to financial decisions, including investment advice, loan approvals, and determining financial eligibility or creditworthiness

Employment and housing: Use cases related to decisions about the employability of individuals, resume screening, hiring tools, or other employment determinations or decisions regarding eligibility for housing, including leases and home loans

Academic testing, accreditation and admissions: Use cases related to standardized testing companies that administer school admissions (including evaluating, scoring or ranking prospective students), language proficiency, or professional certification exams; agencies that evaluate and certify educational institutions

Media or professional journalistic content: Use cases related to using our products or services to automatically generate content and publish it for external consumption

Additional Use Case Guidelines

The below use cases – regardless of whether they are High-Risk Use Cases – must comply with the additional guidance provided.

All consumer-facing chatbots, including any external-facing or interactive AI agent, must disclose to users that they are interacting with AI rather than a human. This disclosure must be provided at a minimum at the beginning of each chat session.

Products serving minors, including organizations providing minors with the ability to directly interact with products that incorporate our API(s), must comply with the additional guidelines outlined in our Help Center article.

Agentic use cases must still comply with the Usage Policy. We provide examples of Usage Policy prohibitions in the context of agentic use in this Help Center article.

Model Context Protocol (MCP) servers listed in our Connector Directory must comply with our Directory Policy.