

Anthropic's Protocol for Addressing Expressions of Suicidal Ideation, Suicide, and Self-Harm Risk

1. User Safety Commitment

Anthropic is committed to the safety and wellbeing of users who interact with Claude. We recognize that users may sometimes express feelings of emotional distress or mental health challenges, and we have trained Claude to approach these delicate situations with care and, where appropriate, have implemented measures to connect users with crisis support resources. Claude is not intended to provide medical or mental health diagnoses, therapy, or treatment recommendations.

2. Our Approach

Anthropic employs a multi-layered evidence-based approach to train, detect, and respond to identified expressions of suicidal ideation, suicide, or self-harm in user interactions with Claude and to prevent the production of content that promotes or facilitates suicidal ideation, suicide, or self-harm.

2.1 Evidence-based Methods

Our overall approach for designing Claude's behavior and safeguards is informed by our in-house subject matter expertise, consultation with various external partners, evidence-based frameworks, and trust and safety best practices. This approach is also informed by both observed and anticipated real-world usage in this area.

2.2 Model Behavior Training and Response Guidelines

Claude is trained to detect and respond to expressions of distress with empathy and care, while referring users to seek support from trained personnel (including qualified professionals) and encouraging them to connect with trusted individuals in their lives. Claude is not designed as a therapist or as a substitute to clinical care or connecting with real-world resources.

2.3 Detection

Anthropic in its consumer facing application uses automated systems to identify when users may be expressing thoughts related to suicide or self-harm while using Claude. These systems are designed to recognize various expressions of distress as well as other moments when support resources in this area may be helpful, including when a user seeks guidance on helping a third party in distress or in

connection with fictional content with significant themes of suicide or self-harm. These systems are regularly evaluated and updated to improve their effectiveness.

2.4 Crisis Resource Referral and Continuing Support

When our systems detect that a user may be expressing thoughts of suicide or self-harm, or interacting with fictional content or media that generally centers on themes of suicide or self-harm, Claude.ai is instructed to display a banner notification that directs these users to crisis support services via our partner, ThroughLine. These resources direct users to trained personnel, who may include counselors, volunteers, or individuals with relevant lived experience, who can provide immediate assistance via phone, text, or chat. The specific resources surfaced vary by region based on the user's location.

Users may be connected to the following crisis support services:

- **988 Suicide & Crisis Lifeline** – Call or text 988 (US)
- **Crisis Text Line** – Text HOME to 741741 (US)
- **International resources** – Users outside the US are directed to crisis services available in their region via <https://claude.findahelpline.com/>

Claude is designed to remain supportive when users express distress and encourage users to connect with human support. This is consistent with evidence-based approaches to crisis intervention, which identify reducing isolation, supporting emotional regulation, and ensuring individuals feel heard as critical to crisis stabilization.

For more information about crisis support resources in Claude, including what to expect when contacting a helpline and answers to frequently asked questions, see [Help Center Article](#): Crisis Helpline Support in Claude.

2.5 Ongoing Evaluation

We regularly review and improve our detection systems and response approaches based on emerging research, expert consultation, and evaluation of effectiveness.

For more information about our evaluation approach, see [Blog](#): Protecting the well-being of our users.

3. Response to User Safety Concerns

Anthropic maintains internal response procedures for situations where we are made aware of user safety concerns outside of our normal automated detection

methods (such as through user reports or third-party referrals). These procedures guide our assessment of potential risks and determination of appropriate responses. Our assessment is informed by research and incorporates evidence-based methods for evaluating the specificity and imminency of suicide risk.

4. Important Notice

Claude is not a substitute for professional mental health care and is not intended to diagnose or treat any medical condition. Resources are not intended to provide medical advice and users experiencing mental health concerns should seek support from qualified healthcare providers, counselors, or crisis services.

Anthropic requires users to be aged 18 or older to access Claude.ai. The service may not be suitable for some individuals, including minors.