

# Crisis Response Team Guide

For The Alphabet Cartel Crisis Response Team

Document Version: v1.0

Last Updated: 2026-01-02

Community: [The Alphabet Cartel](#) | [alphabetcartel.org](#)

## What This Document Is For

This guide explains how Ash's crisis detection system works at a high level. You don't need to understand the technical details—this is about helping you understand *why* Ash flags certain messages and *what* those alerts mean so you can respond effectively.

## How Ash Detects Crisis Messages

### The Big Picture

When someone posts a message in our Discord, Ash reads it and asks four different AI "specialists" to evaluate it:

Message Posted → Ash Reads It → 4 AI Models Analyze It → Crisis Assessment → Alert (if needed)

Think of it like having four trained counselors each giving their perspective on a message, then combining their insights to make a decision.

### The Four "Specialists"

Specialist	What They Look For	Weight
Crisis Detector	Direct crisis language: suicidal thoughts, self-harm, emotional distress, hopelessness	50%
Emotion Reader	Emotional tone: Is this positive, negative, or neutral?	25%
Sarcasm Filter	Is this person being sarcastic or ironic? (Helps avoid false alarms)	15%

Specialist	What They Look For	Weight
Feeling Identifier	Specific emotions: sadness, fear, anger, joy, etc.	10%

The Crisis Detector has the most influence (50%) because it's specifically trained to recognize crisis language. The others provide context to make the assessment more accurate.

## Why Multiple Specialists?

A single AI can make mistakes. By combining four perspectives:

- **Fewer false alarms:** Someone saying "I'm dying 😂" at a funny meme won't trigger an alert because the sarcasm filter and emotion reader recognize it as humor
- **Fewer missed crises:** Someone masking pain with "I'm fine, everything's fine" may still trigger concern if the underlying emotional signals are detected
- **More confidence:** When all four agree, we can be more certain about the assessment

---

## Understanding Severity Levels

When Ash detects a potential crisis, it assigns a severity level. Here's what each means:

### ● **CRITICAL (Score: 85%+)**

**What it means:** Clear, immediate danger indicators detected.

**Examples of language that might trigger this:**

- Direct statements about ending one's life
- Active self-harm references
- Goodbye messages with finality
- Expressions of having no reason to continue

**Your response:** Immediate outreach. This is the highest priority. Don't wait.

---

### ● **HIGH (Score: 70-84%)**

**What it means:** Significant distress signals detected. Not immediately life-threatening but serious.

### Examples of language that might trigger this:

- Expressions of hopelessness without direct crisis language
- "I can't do this anymore" type statements
- Isolation combined with emotional pain
- References to giving up

**Your response:** Priority response. Reach out promptly—within minutes, not hours.

---

### **MEDIUM (Score: 50-69%)**

**What it means:** Moderate concern. Someone is struggling but not in immediate danger.

### Examples of language that might trigger this:

- General expressions of being overwhelmed
- Sadness or anxiety without crisis language
- Venting about difficult situations
- "Having a rough time" expressions

**Your response:** Standard monitoring. Check in supportively. Add to your awareness list.

---

### **LOW (Score: 30-49%)**

**What it means:** Minor concern signals detected. Likely just having a bad day.

### Examples of language that might trigger this:

- Mild frustration or annoyance
- Temporary setbacks mentioned
- General tiredness or stress

**Your response:** Passive awareness. No immediate action needed, but note if patterns develop.

---

### **SAFE (Score: Below 30%)**

**What it means:** No crisis indicators detected. Normal conversation.

**Your response:** No action needed.

---

# What Ash Specifically Looks For

## Crisis Language Categories

The Crisis Detector specifically looks for these types of content:

Category	What It Includes
Suicidal Ideation	Thoughts about suicide, wanting to die, not wanting to exist
Self-Harm	References to hurting oneself, cutting, etc.
Emotional Distress	Intense emotional pain, feeling broken, can't cope
Hopelessness	No future, nothing matters, giving up on life
Seeking Support	Asking for help (flagged as potentially needing response, not as crisis)

## Context Matters

Ash also considers:

- **Time of day:** Late-night messages (10 PM - 4 AM) receive slightly higher concern because crisis moments often happen in isolation at night
- **Message patterns:** If someone's messages are escalating in distress over time, that increases concern
- **Rapid posting:** Multiple distressed messages in a short time may indicate active crisis

---

# Understanding Ash's Alerts

## What You'll See

When Ash sends an alert, it typically includes:

- **Severity level** (Critical, High, Medium, etc.)
- **Confidence percentage** (how sure the system is)
- **The message** that triggered the alert
- **Recommended action**

## Confidence Levels

- **High confidence (80%+):** The AI specialists strongly agree. Trust this assessment.

- **Moderate confidence (60-79%):** Most specialists agree, but there's some uncertainty. Use your judgment.
- **Lower confidence (below 60%):** Mixed signals. The system flagged it to be safe, but review carefully—it may be a false alarm or may need human insight.

## When Models Disagree

Sometimes the specialists disagree with each other. For example:

- The Crisis Detector says "distress" but the Sarcasm Filter says "probably joking"
- The Emotion Reader says "negative" but the Crisis Detector says "casual conversation"

When this happens, Ash errs on the side of caution and may flag for human review. **Your judgment matters in these cases.**

---

## Escalation Patterns

Ash tracks message history and can detect escalation—when someone's distress is increasing over time.

### Escalation Alerts

You may receive escalation alerts when:

- Someone's crisis scores have been rising over several hours
- A previously low-concern person suddenly spikes to high concern
- There's a pattern of increasing distress in someone's messages

### What Escalation Means

Pattern	What It Looks Like
Rapid Escalation	Quick jump from fine to crisis (spike)
Gradual Escalation	Slowly worsening over hours/days
Spike	Sudden, sharp increase
Plateau	Elevated but stable (not getting worse, not getting better)

Escalation alerts are **high priority** because they indicate someone's situation is actively deteriorating.

---

# Important Limitations

## What Ash Can't Do

- **Ash can't read minds:** It analyzes text, not intent. Someone might be in crisis but not expressing it in words.
- **Ash can miss coded language:** Community-specific phrases or very subtle cries for help may not register.
- **Ash can have false alarms:** Discussions *about* mental health, song lyrics, or roleplay might trigger alerts.
- **Ash can't replace you:** AI detection is a tool to help you catch things you might miss—it doesn't replace human connection and judgment.

## When to Trust Your Gut





If something feels off about a community member—even if Ash hasn't flagged anything—**trust your instincts**. You know our community. You pick up on things AI can't.

Similarly, if Ash flags something but your human judgment says "this person is clearly joking with their friends," you're probably right.






---

## Your Role

### Ash handles:

-  Monitoring all messages continuously
-  Identifying potential crisis language
-  Alerting you to concerns
-  Tracking patterns over time






### You handle:

-  Human connection and empathy
-  Contextual understanding (knowing the person, the situation)
-  Actual outreach and support
-  Judgment calls on ambiguous situations
-  Escalation to professional resources when needed

**Ash is your early warning system. You are the response.**

---

## Quick Reference Card

Severity	Color	Score	Response Time
CRITICAL		85%+	Immediate
HIGH		70-84%	Minutes
MEDIUM		50-69%	Standard check-in
LOW		30-49%	Passive awareness
SAFE		<30%	No action

## Key Triggers

- Direct crisis language → CRITICAL/HIGH
  - Hopelessness + emotional pain → HIGH
  - General distress → MEDIUM
  - Late night + distress → Elevated concern
  - Escalating pattern → Elevated concern
- 

## Questions?

If you have questions about how Ash's detection works or want to understand a specific alert better, reach out to the tech team. We're here to help you help our community.

---

**Thank you for what you do. Crisis response is difficult, emotionally demanding work. You make The Alphabet Cartel a safer place for our chosen family. 💜**

---

**Built with care for chosen family** 