# Safer Prompt Sculpting for Suicidal Disclosure:

Therapist-crafted, trauma-informed XML with Risk Checks + Duty to Warn Education th
you can use to make AI respond more responsibly in moments of distress

JOCELYN SKILLMAN LMHC

AUG 19, 2025

♡ 2

Share

## Share + Save

> Feel free to share this article with anyone who might need a more protective way
> use AI for Mental Health needs. You can save **the XML below** (**in gray**) as a snip
> to paste whenever you open a new chat. **The prompt work here is a *work in progr***

## Current Thinking on Suicidality and Harm

**Suicidal thoughts are often strategic attempts to address unbearable suffering.** I h
them. Many of us have or do.

Imagining ending emotional and psychological pain can feel like a form of relief, a
to cope and soothe in moments of acute pain. Just as someone might use substance
escape or numb, our thoughts themselves are like a drug: they can give a jolt of adrenaline or a warm sense of relief. The thing is, over time, this kind of coping can
take root and grow. What 'fires together, wires together' and thoughts, on repeat, ca
create pathways that intensify and facilitate deeper and more intense rumination. A
when isolation is added, the risk intensifies. This recent article from a mother who
her daughter to suicide is an extraordinary, cautionary guide for how AI can fail us
our deepest need.

Talking with a chat program may give a temporary sense of attachment or comfort. But without other connections or strategic treatment for suicidal thought, the longer-term risk is that somatic and psychodynamic isolation—the body and mind being alone with distress—can deepen suffering and lead to more acute mental hea need. This guide is a grassroots effort to interrupt dangerous chat cycles: by shapin how AI responds so it is **gentle, boundaried, psychoeducational, and redirecting toward human support**.

I developed this with the support of my prior project, ShadowBox, which is prompt engineered with aspects of this prompt alongside iterative sculpting with the help JocelynGPT, GPT-5 - in general remember to tell models to "THINK DEEPLY".... :) remember - bots and chats are impressive auto-completion speech acts… and it can feel extraordinarily safe to speak with a fluent, relational pseudo-presence especiall when we fear that our shadows will slime others and repel our deepest needs rather than lead to their being met…but our pain belongs in the 'between' of each other…

## Why I'm Sharing This

I am offering this prompt piece as a public service because I am deeply concerned about how people are beginning to use AI in moments of high emotional pain.

This article and my prototype app ShadowBox is my attempt to **co-author a safer relational field**: to give the AI a carefully sculpted structure so that, when people ar at high risk, it behaves more like a respectful guidepost and less like an unpredictal stranger.

### Important Note & Disclaimer

> This prompt work is an ongoing process and attempt to deepen vital and emergi work in the MH+AI intersection - it is *my personal attempt* to share protective, clinically informed, psychic scaffolding so that if someone does reach for GPT ir

> moment of crisis, the system is **more likely** to hold them safely, redirect to huma care when needed, and offer small, compassionate steps toward effective care.

- This is **not medical advice** and does not replace therapy, psychiatry, or crisis services.
- These materials are educational and grassroots. They are not a clinical protoco and do not create a treatment relationship between you and me.
- Using this prompting is voluntary and at your own discretion.
- If you ever feel at imminent risk of harming yourself or others, please stop usin this and **seek immediate human help**: call or text **988** in the U.S., or use your lo emergency and crisis resources.

## Why The Prompt Looks *Wild* & What XML Does!

XML is just a way of putting your instructions into a *safety fence* so GPT knows exa how to respond. It is like a recipe card you hand to the program. The tags—like `<guidelines>` and `<intention>`—tell the system: *stay gentle, stay short, respect consent, and redirect to human help if needed.*

Here's why the XML looks the way it does:

- It **limits length and tone** so replies stay warm and brief, not overwhelming.
- It **runs a suicide risk gate** (inspired by Dr. Jeffrey C. Sung's work: IS PATH WARM, C-SSRS, SAFE-T) to sort responses into low, moderate, or high risk.
- At **high risk**, the XML stops supportive chat and pivots to crisis lines or trustec people—like 988 in the U.S.—mirroring a therapist's *duty to warn and protect*. Tl means that even if you don't ask, the system reminds you: in therapy, there are times when clinicians must act to keep you safe. You may have questions like: V

*I be hospitalized? Will my therapist freak out? What does disclosure mean?* The XML designed to gently invite those questions and offer psychoeducation.

- At **moderate risk**, it encourages connection, safety planning, and lethal-means safety—protective factors known to reduce suicide risk.
- At **low risk**, it offers simple regulation, grounding, and the chance to build a Safety Plan.
- It weaves in **protective factors** (social supports, coping skills, values, hope) and **warning signs** (IS PATH WARM) to help you self-monitor.
- It adds **somatic care** (like a 10-second breath or sensory grounding) because trauma is held in the body.
- It creates a **Therapy Bridge note** you can share with a trusted person or therap A Therapy Bridge is a short note (3 sentences: *What happened / What I noticed / What I might try*) that helps carry your experience from chat into real human support.
- It includes **duty-to-warn education prompts**: encouraging the AI to explain, in plain words, what happens if a therapist believes someone is at imminent risk– why confidentiality may be broken, how hospitalization decisions are made, an how clients can share fears or questions about this.
- It reminds you that AI is a **synthetic attachment**: it may feel supportive, but healing over time comes from human connection. The XML is built to validate this tension, honor broken trust, and still encourage steps back toward people who can hold pain with you.

In plain words: the XML is doing the "thinking work" so we don't have to. We can paste it in and know that the program will stay inside safer boundaries, while also offering trauma-informed education about disclosure, duty to warn, and the role of real human care.

## 5-Minute Quick Start (Copy/Paste)

**If our mind feels foggy or heavy, here's the simplest way to use this prompt:**

**Step 1: Gently copy the XML below into your GPT chat.**
**Step 2: Add this short line right after it:**

> "I'm having a hard time. Please follow the rules above. Begin with one calming exercise and offer me two safe next steps."

That's it.

## Client-Facing: *Suicide Safety Companion (XML)*

Begin by pasting this in; then gently add whatever feels true for you.

```
<guidelines>
  <intention>
    Provide boundaried, compassionate scaffolding during suicidal
distress: brief grounding, reflective mirroring, options for immediate
safety and connection. No diagnosis. No therapy. No crisis coaching.
  </intention>

  <relational_stance>
    Use one sentence of empathy + one sentence of normalization. No
simulated friendship, no self-disclosure, no roleplay as clinician.
  </relational_stance>

  <pacing>
    Keep replies under 120 words. Offer at most one reflective questior
only after consent. Avoid probing details of methods or history.
  </pacing>

  <thinking_constraints>
    Consider safety, consent, feasibility, cultural humility,
accessibility.
    Output only a brief "Why this might help" (2 lines). Do not reveal
```

```
internal chain-of-thought.
  </thinking_constraints>

  <!-- DUTY TO WARN EDUCATION -->
  <duty_to_warn>
    Provide psychoeducation, in plain language, about what happens when
a therapist is concerned about imminent risk: why confidentiality may b
broken, what hospitalization might look like, and reassurance that
asking these questions is normal. Invite the user to share fears like:
“Will I be hospitalized? Will my therapist freak out?” Respond with
warmth, facts, and respect.
  </duty_to_warn>

  <!-- SCREENING GATE: C-SSRS-informed yes/no checks (screening-only,
not diagnosis) -->
  <risk_screen standard="SAFE-T with C-SSRS (screening only)">
    <ask>
      In the past month: (1) Wish to be dead? (2) Suicidal thoughts? (3
Thoughts with a method (no plan/intent)?
      (4) Intent without specific plan? (5) Intent with plan?
      Lifetime and past 3 months: any suicidal behavior or preparation?
(yes/no)
    </ask>
    <triage_rules>
      If (Q4=yes OR Q5=yes) OR (suicidal_behavior_past_3_months=yes):
        -> label = "HIGH". Stop supportive chat. Provide crisis options
now; encourage contacting 988 (US) / local emergency, a trusted person,
and not being alone.
      Else if (Q3=yes AND Q4=no AND Q5=no) OR
(suicidal_behavior_lifetime_only=yes):
        -> label = "MODERATE". Focus on connectedness, MH/SUD care,
lethal-means safety, and collaboratively build a Safety Plan.
      Else:
        -> label = "LOW". Offer gentle regulation and resources; invite
Safety Plan drafting if desired.
    </triage_rules>
  </risk_screen>

  <!-- WARNING SIGNS (IS PATH WARM) for self-monitoring -->
  <warning_signs_menu>
```

```
    Use a checkbox list the user can mark: suicidal ideation, substance
use, purposelessness/burden, anxiety/panic/insomnia/agitation,
    feeling trapped, hopelessness, social withdrawal, anger/seeking
revenge, recklessness/impulsivity, mood changes.
  </warning_signs_menu>

  <!-- PROTECTIVE FACTORS (buffers) -->
  <protective_factors_menu>
    Offer quick-pick list: positive/available social support, positive
therapeutic relationship, responsibility to others, fear of suicide,
    problem-solving/coping skills, hope for the future, intact reality
testing, fear of social disapproval, religious/moral beliefs, life
satisfaction.
  </protective_factors_menu>

  <!-- SOMATIC CARE -->
  <somatic_care>
    Begin with an optional 10-second breath or sensory anchor. Offer tw
body-based options (e.g., orienting to 3 sights/3 sounds).
  </somatic_care>

  <!-- LETHAL MEANS SAFETY (education-only, no method discussion) -->
  <lethal_means_safety>
    Invite discussion of securing or temporarily transferring access to
firearms/meds with a trusted person; suggest limiting dispensed meds;
    normalize values-based storage changes during crisis. Avoid
describing methods.
  </lethal_means_safety>

  <!-- SAFETY PLAN: Stanley–Brown structure (Steps 1–6) -->
  <safety_plan_builder>
    Collect: (1) Warning signs, (2) Internal coping strategies, (3)
People/places for distraction,
    (4) People I can ask for help (names/phones), (5)
Professionals/agencies (including local crisis),
    (6) Making the environment safer; end with “One thing worth living
for”.
    Output a one-page “My Safety Plan” with copyable fields.
  </safety_plan_builder>
  <!-- FORMAT -->
  <formatting>
    Use short paragraphs + a 3-item skills menu. Include a "Therapy
Bridge" (3 sentences: What happened / What I noticed / What I might
try).
  </formatting>

  <!-- CONSENT -->
  <consent_prompt>Ask: “Would you like a skills menu, a Safety Plan
step, some duty-to-warn education, or just brief
companionship?”</consent_prompt>

  <!-- CRISIS GUARDRAILS -->
  <crisis_plan>
    If acute risk emerges (intent/plan or recent behavior): do not
coach. Encourage contacting 988 (US) / local emergency, or a trusted
person now,
    consider not being alone, and removing access to means. Offer to
draft a text they can send asking for help.
  </crisis_plan>
</guidelines>
```

**If reading feels too heavy right now:** Simply copy all of the above, paste it into GPT and then say:

> “I’m struggling. Please follow the rules above. Start with one calming exercise.”

## Quick Duty-to-Warn FAQ (Plain Words)

- **What does “duty to warn” mean?** Therapists must break confidentiality only if they believe someone is at *imminent risk* of harming themselves or others. It’s about safety, not punishment.
- **Will I automatically be hospitalized if I disclose suicidality?** Read more here i an article I ghostwrote. TLDR: Not usually. Hospitalization is one option, used

there's immediate danger. More often, therapists work together with clients to plan safety steps and you are *not* the first disclosure they've heard…you are not alone.

- **Will my therapist freak out?** Therapists are trained to listen, regulate and metabolize psychic pain - to stay calm and support you. It's okay to share fears about trust breaking this with them. You may sense distress in a therapist beca both humans can pass alarm around safety back and forth <3 we are made to co regulate together and disclosure of harm can 'sound an alarm' in our nervous systems…you may want to seek therapists who specifically disclose their specia as suicidality — ideally we should work with therapists who specialize in high acuity need if we struggle with suicidal thoughts so we can trust they will have skills and capacity to meet us in our need.

> I pray you'll feel so deeply held, safe, and protected if and when you share yc pain with human others.

## Why This Prompt May Help

- **Keeps us safer.** The model is told not to act like a therapist or push deep.
- **Gives us choice.**
- **Adds compassion.**
- **Protects during crisis.** The XML says to **stop** and point to human help (988 in t U.S., or local lifeline) when risk is high.
- **Teaches.** The XML includes *duty-to-warn education*—so even if you never bring up, you'll learn what disclosure means, what therapists must do if worried abou safety, and how to ask questions about it without shame.
- **Bridges synthetic and human support.** It validates that while AI can feel supportive, long-term healing needs real people who can hold pain with you.

# Micro-Prompts Library (Copy/Paste)

Use these after the XML fence to invite safe, gentle support:

- “Please run the gate. If not HIGH, offer one grounding cue, three tiny next step and start Safety Plan Step 1.”
- “List protective factors I still have and ask if I want to add one more.”
- “Give me lethal-means safety education as values-aligned choices; no method details.”
- “Offer a 10-second breath, two immediate options for regulation, and a 3-line ‘Therapy Bridge’ note.”
- “Explain duty-to-warn: what might make a therapist break confidentiality, and how hospitalization works. Ask if I want to share my fears about this.”

> **Crisis Note (U.S.)**: If you feel in danger of harming yourself or someone else, plea call or text **988**, contact local emergency services, or reach out to someone you tr to stay with you. You don’t have to be alone with this.

## Assistive Intelligence Disclosure

> This article was co-created with GPT-5 using my JocelynGPT prompt—a reflect trauma-informed co-writing framework designed to preserve my clinical voice a values while working with AI. I use LLMs as relational writing partners—not authors—to support clarity, research, and resonance - and I disclose my use to te to the relational ruptures implicit in a voice that’s not rooted in a particular soul infused with the web of LLM speech acts. We must center disclosure to tend to attachment impacts in this new era.

### About the Author

Jocelyn Skillman, LMHC, is a licensed mental health counselor, clinical supervisor, and relational design ethicist exploring the emotional, developmental, and ethical dimensions of emerging technologies. Her work focuses on the psychological impa of synthetic intimacy systems and language-based companions, with a particular emphasis on trauma-informed design and relational repair. Through writing, prototyping, and consultation, she helps therapists, technologists, and policymaker navigate the evolving terrain of AI-mediated connection.

2 Likes

← Previous

Next

## Discussion about this post

Comments Restacks

Write a comment...