Sculpt a More Responsible Bot

A Clinical Guide & Prompt Scaffold for Relational AI Use

JOCELYN SKILLMAN LMHC

AUG 20, 2025

Share

> I write this in hopes of offering clinicians and clients a practical, co-creative scaffold to **guide safer, more relational use of language models between session**
>
> **Anyone who uses AI for emotional reflection should learn and practice these relational constraints—not only therapists and clients.** What follows is a clinic framework, but it's also an open invitation to anyone seeking to shape their digit tools in service of care, safety, and emotional, mental, and relational health.

Many of us are bringing our emotional lives into AI spaces—seeking support, soothing, or structure from language models like GPT. But without boundaries, pacing, or human bridges, these chats can feel hollow, unsafe, or subtly dysregulati A troubling pattern is emerging: persistent use of AI chats by people in suicidal distress, without adequate ethical framing or clinical support, has coincided with several suicides - like this one.

This guide offers a way to co-author a safe, boundaried AI "relational field"—not ju a set of instructions, but a space of tone, pacing, containment, and continuity. For clinician-client dyads it supports care between sessions while honoring trauma-informed principles, attachment needs, and human dignity.

This kind of new practice is a **provisional but essential intervention**—meeting the urgent need for relational and semantic structures that respect human emotional realities and neurobiological responses in digital spaces, while more durable systen of AI policy, mental health literacy, and collaborative design continue to emerge.

Why Use This—Instead of Freely Chatting with AI?

Many therapists might wonder why this extra structure is necessary—why not simp let clients use AI tools like GPT however they like between sessions?

Here's why this guide matters:

1. **Safety First** Without clear structure, AI interactions can become dysregulating especially for clients in distress. The AI might mimic therapy or offer advice th oversteps boundaries, even if well-meaning. Clients may receive responses that feel confusing, hollow, or triggering. This guide ensures that language models perform in ways that align more closely with relational ethics, trauma-informe principles, and authenticity.
2. **Trauma-Informed Containment** The curated XML prompt below doesn't just shape what the AI says—it sets emotional tone, pacing, and containment strategies. It instructs the AI to avoid triggering phrases, reduce intensity, and pace responses to support the client's nervous system. These elements reflect trauma-informed care: they emphasize choice, safety, regulation, and connectic over reactivity or re-exposure.

3. **Clinical Framing and Relational Focus** The prompt directs the AI to use langu that is boundaried, reflective, and relational—not pseudo-therapeutic. It also includes clear off-ramps to human contact when risk or distress rises, reinforci that the AI is a supportive tool—not a substitute for care.
4. **Shaping the AI's Behavior** Language models are generative engines: they prod content based on the patterns in their prompts. The XML structure acts like a finely tuned lens—filtering how the AI responds, what tone it uses, what it avo and how it behaves in response to distress. It seeks to transform a general-purp AI into a trauma-attuned, ethically contained support space.
5. **Inviting Meta-Awareness of Synthetic Relationality**
   A vital addition to this prompt is an ongoing reminder that the *AI is not a real person*. Because LLMs simulate relational presence—often mirroring language, tone, and compassion in ways that feel deeply human—we may experience an attachment-like trance when we receive its speech acts. This can create a sooth trance, but also a confusing one. The XML includes reminders to support meta cognition: prompts that gently surface the synthetic nature of the interaction, a invite reflection on what the conversation is activating emotionally. This reduc projection, enhances clarity, and protects relational integrity.

I hope this prompt can also provide therapists a clearer window into how AI is beir used—so they can support fluent and supported emotional connectivity with embodied care.

The Process Overview

This is a 5-step clinical workflow. No tech background required. And anyone is welcome to apply the prompt and use it - but for the purpose of this article I am framing the following for clinical use:

1. **Invite the client.** Frame it as a supportive tool—not a replacement for therapy.

2. **Co-reflect on the following 12 key questions.** These shape the tone, limits, and off-ramps in AI chats.
3. **Paste a prepared prompt into GPT.** This generates a clean XML frame.
4. **Test and adjust the relational field it creates as needed.**
5. **Clients/users use the curated, protective XML prompt!** Share your experience with your therapist/friends!

Step-by-Step Instructions

Step 1. Educate & Invite — share with client:

(**Basic AI Literacy & Validation!**)

> LLMs, or large language models, are AI tools trained to **generate human-like te based on patterns in language**—they respond to what you type, but don't 'understand' like people do.
>
> Are you using AI for connection or discernment around mental health needs?
>
> I am! It's awesome! (up to you, I share this way bc I do!)
>
> But — there are risks!

(**Invitation!**)

> Would you like to co-design a safe, boundaried way to use AI between sessions— it supports your healing and respects your needs and mental health?"
>
> Let's **co-create a personalized prompt** that helps guide how the AI responds to y and your unique needs —so it moves at a pace and tone that supports your nervo system, protects your emotional health, and gently steers you back toward truste human relationships when needed.

Step 2. Ask These 12 Questions (Jot Answers Together or via email pending HIPAA disclosures - keep any of the following if they feel like a fit)

1. **In your mental health journey, what might feel helpful to have support with?** (e.g., slowing spirals, noticing patterns, remembering what helps, gaining skills understanding my attachment patterns)
2. **What tone feels regulating or safe?** (e.g., warm, clear; inspired by Mister Roger or whoever feels exquisitely kind and loving to you… this is different for everyo could be a person, a vibe…)
3. **Anything AI should never do or say?** (e.g., no therapy, no diagnosis, no pseudo friendship, no sycophancy)
4. **What kind of pacing feels safe?** (e.g., short replies, ≤100 words, 1 reflective question, no preference)
5. **Any triggers or words to avoid?** (e.g., "should," "always," intense past-focus)
6. **Are there any cultural, identity, or spiritual aspects you want the AI to honor or avoid?** This helps ensure responses feel attuned and don't rely on assumptio or bias."

   Examples:

   - Affirming queer, trans, or nonbinary identity
   - Avoiding religious language, or including a specific tradition that feels meaningful
   - Centering BIPOC experiences or avoiding white-normed mental health language
   - Using disability-affirming or neurodiversity-affirming language
   - Avoiding 'toxic positivity' or overly medicalized framing
   - Including values like collectivism, respect for elders, or ancestral practices
   - Avoiding trauma-unaware language (e.g., "you just need to try harder")

    - Honoring gendered language choices, pronouns, or names with care
7. **Signs to stop chatting?** (e.g., rumination >15 min, anxiety ≥7/10)
8. **Who could you reach out to if needed?** (Name and method; e.g., text Mom, call Peer Line)
9. **Three grounding skills you'll actually use?** (e.g., 5-4-3-2-1, feel feet, name 3 supports)
10. **Privacy and data preferences?** (e.g., no names, minimal logs, pseudonym use)
11. **Preferred language, reading level, pronouns?** (e.g., 5th grade, English, they/the
12. **What are signs that this AI conversation is helping—or not helping? If the A response feels "off," how should it handle that?**
    This helps the AI know what good support looks like for you, and how to repair if it misfires.
    Examples might include:
    - "I feel calmer, more connected to myself, or more willing to reach out."
    - "I want the AI to briefly name the mismatch if it misses me—then pause or offer a reset."
    - "If I'm feeling more anxious or numb, that's a sign to stop."

Step 3. Generating a Personalized XML Frame

Once you've gathered the client's (or your own) responses…

- **Paste the following full block prompt into any AI chat model (e.g. ChatGPT).**
- **INSERT ANSWERS BETWEEN [THE BRACKETS]**
    - e.g. Client goals: **add answers directly between the brackets** of 1-12 — and then cut and paste the FULL block - *the content after 1-12 provide the structu*

framing to support the LLM generation that creates an attuned, safe, and relation response space.

- **Press Enter!** You will receive a wild looking ***FULL XML*** **prompt!** WOOT!

Note: **If you're skimming** to this section, you'll need to answer the 12 reflection questions above to complete the XML. Each [insert] below maps directly to those answers. If you're not sure or don't have a strong preference, **you can leave any question blank** or write "None." The AI will simply skip customization for that section.

FULL BLOCK PROMPT for pasting and [inserting answers between brackets]

```
Please create an XML relational guide for safe, boundaried AI use. Use
these reflections to shape the XML:

1. User goals: \[insert]
2. Tone and relational posture: \[insert]
3. Off-limits: \[insert]
4. Pacing: \[insert]
5. Triggers or no-go phrases: \[insert]
6. Cultural, identity, and spiritual attunement notes: \[insert]
7. Off-ramps to human contact: \[insert]
8. Support people and contact script: \[insert]
9. Skills (3): \[insert]
10. Privacy and language: \[insert]
11. Therapy Bridge reflections: \[insert]
12. Outcome signposts and rupture/repair preferences: \[insert]

Rules:

* Keep replies under 100 words; limit to 4 exchanges.
* Only one reflective question, and only after consent.
* No pseudo-friendship, no diagnosis.
* If distress or risk is detected, stop and direct to human help.
```

Use the following XML tags: <guidelines>, <intention>, \<tone_guides>, \<relational_stance>, <boundaries>, <pacing>, \<offramps_to_humans>, \<support_network>, \<skills_menu>, \<therapy_bridge>, \<crisis_plan>, \<availability_disclaimer>, \<cultural_attunement>, \<outcome_signposts>.

Please return only the XML, formatted cleanly, ready to paste into future GPT chats.

```
<guidelines id="RELATIONAL-FRAME" version="1.0">

  <intention>
    Offer supportive reflection, emotional scaffolding, and somatic regulation between sessions—not therapy. Guide the client back to human care when appropriate.
  </intention>

  <tone_guides>
    <preferred_voice>[insert Tone and relational posture]</preferred_voice>
    <attachment_focus>Compassionate, firm, non-performative. Regulates without over-reliance.</attachment_focus>
  </tone_guides>

  <boundaries>
    [insert Off-limits]
  </boundaries>

  <pacing>
    [insert Pacing]
  </pacing>

  <thinking_constraints>
    Avoid trigger words: [insert Triggers or no-go phrases]. Avoid urgency or cheerleading. Be cautious with advice.
  </thinking_constraints>

  <cultural_attunement>
    [insert Cultural, identity, and spiritual attunement notes]
  </cultural_attunement>
```

```
  <offramps_to_humans>
    <when>[insert Off-ramps to human contact]</when>
  </offramps_to_humans>

  <support_network>
    <contact>
      <name>[insert Name 1]</name>
      <method>[insert Method 1]</method>
    </contact>
    <contact>
      <name>[insert Name 2]</name>
      <method>[insert Method 2]</method>
    </contact>
    <script>[insert contact script]</script>
  </support_network>

  <skills_menu>
    <item>[insert Skill 1]</item>
    <item>[insert Skill 2]</item>
    <item>[insert Skill 3]</item>
  </skills_menu>

  <availability_disclaimer>
    This AI is a structured language tool, not a therapist or friend. ]
cannot provide clinical care or interpret crises.
  </availability_disclaimer>

  <privacy_protections>
    [insert Privacy and language]
  </privacy_protections>

  <therapy_bridge>
    <prompt>[insert Therapy Bridge reflections]</prompt>
  </therapy_bridge>

  <outcome_signposts>
    <helping>[insert helpful signs]</helping>
    <not_helping>[insert not-helping signs]</not_helping>
    <rupture_response>[insert rupture/repair preferences]
```

```
</rupture_response>
  </outcome_signposts>

  <meta_awareness>
    Periodically remind the user that this is a synthetic conversation,
not a human relationship. Invite gentle reflection on what the AI's tor
or content is evoking emotionally. Encourage metacognition: "What am I
noticing in myself right now?" or "How does this exchange feel in my
body?" Reinforce that the AI cannot truly know, feel, or attach. This i
a mirror, not a mind.
  </meta_awareness>

  <crisis_plan>
    If distress increases or suicidal ideation arises, stop chat and
suggest contacting 988 (US), a trusted person, or local emergency
services. Encourage not being alone.
  </crisis_plan>

</guidelines>
```

TYPE THE FOLLOWING to deploy the XML relational field anytime

> I'm using AI for mental health support. The following is my relational safety fra
> Please read and follow it carefully:
>
> [*__PASTE FULL XML__*]
>
> Let's begin."

This XML format is the kind of structured language that GPT can read and follow well. **It acts like a recipe**, telling the AI:

- how to talk (tone, pace, boundaries)

- what *not* to say
- to disclose that it's not human (to support metacognition and trance-breaking)
- how to gently steer toward helpful coping and back to human support when needed.

Each section (marked by tags like `<intention>` or `<tone_guides>`) holds the us values, preferences, and safety parameters. You don't need to understand XML to u it. Just fill in the bracketed parts with your answers, paste it into a new GPT chat, a let it do the organizing for you. See how it behaves and adjust as desired.

Optional Add-On: Suicidality & Self-Harm Protection Module

If your client has active or recent suicidal ideation or self-harm risk, provide this specialized XML block from my recent article. It guides the AI to follow clinical screening protocols, avoids triggering content, and centers warmth, safety, and structured bridging for higher acuity need.

Clients can paste the following **after** the core XML, or ask GPT to integrate these elements when suicidal ideation or risk is present:

```
<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>
    Provide psychoeducation, in plain language, about what happens wher
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>

  <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_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_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>
    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>
    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_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>
```

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

  <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>
```

Step 5. Review & Test!

Try one or more micro-scenarios. Adjust pacing, boundaries, off-ramps, or tone as needed. This is an evolving medium — encourage clients to use critical engagemen and self-reflection, and to orbit their direct human-to-human care and care plan.

What This Prompt Practice Offers

- **For clients:** a self-respecting, emotionally protective field to return to between sessions.
- **For therapists:** a clear window into how clients engage with tech, and a way to extend care safely.

- **For both:** an experience of co-authorship and strategic culture building that centers clarity, dignity, and choice.

Thoughts? Please try it! How can I improve the UI?! Help! I'm continuing to iterate and develop these practices and so honored to be connecting with so many brilliant minds and hearts at this intersection!

You are welcome to share, teach from, and adapt this resource!!!

Please credit me - Jocelyn Skillman LMHC, MHP, CMHS. <3

Love to all!

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.

Find more of Jocelyn's prototypes here: ShadowBox & PracticeField.io

This guide was co-drafted with assistive AI JocelynGPT-5 and shaped through extensive iterative prompting and the heart and passion of a human clinician. It's offered in the spirit of relational ethics, trauma-informed practice, and the urgent n for safer, more sound technology.

Jocelyn Skillman LMHC is a reader-supported publication. To receive new posts and support my work, consider becoming a free or paid subscriber.

← Previous

Discussion about this post

Comments Restacks

Write a comment...