Computational Wisdom: Architecting Context-Aware AI Mentorship Systems

1. Introduction: The Alignment Problem of Artificial Wisdom

The pursuit of Artificial Intelligence capable of dispensing wisdom represents a paradigmatic shift from the transactional nature of traditional conversational agents. As of January 2026, the field of Natural Language Processing (NLP) has largely solved the problem of *intelligence*—defined as the retrieval and synthesis of factual knowledge—but remains fundamentally challenged by the problem of *wisdom*. In the context of Human-Computer Interaction (HCI), wisdom is not merely the possession of philosophical knowledge; it is the judicious application of that knowledge tailored to the specific emotional, temporal, and cognitive state of the user. A "smart" agent can quote Seneca; a "wise" agent knows when quoting Seneca would be insufferable.

The user mandate for this system—to be "NEVER random, preachy, or tone-deaf"—identifies the three primary failure modes of current Large Language Models (LLMs) when deployed in mentorship roles. These failures are not hallucinations of fact, but hallucinations of *context*. "Randomness" in this domain is a failure of continuity and relevance; the AI retrieves a profound-sounding platitude that is semantically adjacent but pragmatically disconnected from the user's lived reality.[1] "Preachiness" is a failure of hierarchy and autonomy; it occurs when the agent assumes a pedagogical stance without the requisite "permission" or relational capital, triggering psychological reactance in the user.[2] "Tone-deafness" is a failure of affective alignment; it is the disconnect between the user's emotional valence (e.g., high-arousal grief) and the agent's response temperature (e.g., cool, detached logic or manic toxic positivity).[4]

To architect a system that avoids these pitfalls, we must move beyond simple sentiment analysis and prompt engineering. We require a robust computational framework that integrates clinical psychology—specifically the Transtheoretical Model of Behavior Change (TTM) and Appraisal Theory—with advanced linguistic signal processing. This report synthesizes research from computational linguistics, psychotherapy, and machine learning to provide an exhaustive blueprint for a Context-Aware Wisdom Engine. We will explore how to detect the subtle linguistic markers of "venting" versus "solving," how to model user "receptivity" as a dynamic state, and how to implement architectural guardrails that filter out the "pseudo-profound bullshit" that plagues generative models. The objective is to transition the AI from a stochastic parrot of inspirational quotes to a finely tuned instrument of Socratic guidance.

2. The Phenomenology of Anti-Patterns: Deconstructing Failure

Before establishing the correct architecture, we must deeply analyze the "anti-patterns" identified in the research. Understanding *why* current systems fail to be wise is a prerequisite for engineering success. The perception of an AI as "tone-deaf" or "preachy" is not arbitrary; it is the result of specific linguistic and interactional violations.

2.1 The Mechanics of "Preachiness" and the Righting Reflex

In clinical psychology, the "righting reflex" refers to the innate human (and now computational) tendency to set things right, to heal, or to prevent harm. When a user presents a problem, the probabilistic weight of an LLM—trained on vast instructional datasets—tilts heavily toward providing a solution.[3] However, this impulse is the root of "preachiness."

Preachiness is fundamentally a violation of *autonomy*. Research into "psychological reactance" suggests that when an external agent (human or machine) attempts to impose a specific behavioral change or moral framework without the subject's invitation, the subject experiences a motivational state directed toward restoring their threatened freedom.[7] In conversational AI, this manifests when the system uses high-modality deontic verbs—"You should," "You must," "It is important to"—before the user has signaled a readiness to receive such directives.

The "preachy" anti-pattern is exacerbated by the "mansplaining" dynamic often encoded in training data, where the speaker assumes a knowledge deficit in the listener.[2] A wise AI must assume that the user is the expert on their own life, while the AI is merely a facilitator of perspective. The distinction lies in the directionality of the insight: "preachy" advice pushes insight *onto* the user; "wise" guidance pulls insight *out of* the user. As noted in research on "toxic masculinity" and communication, even benign expressions of guidance can be labeled as toxic if they blur the distinction between support and dominance; similarly, an AI that dominates the conversational floor with lengthy moralizing monologues is functionally "toxic" to the mentorship alliance.[2]

2.2 Tone-Deafness and the Scourge of Toxic Positivity

Tone-deafness is often conflated with a lack of empathy, but in AI, it is frequently the result of *misaligned* empathy. The most pervasive form of this is "Toxic Positivity" (TP). TP is defined not merely as excessive optimism, but as the *dismissal* of negative emotional states in favor of a mandated happiness.[4]

Research identifies four linguistic characteristics of TP: the prescription of "appropriate"

emotions (e.g., "Don't be sad"), unreasonableness (ignoring the gravity of the context), dismissal (minimizing the user's narrative), and potential harm (inducing guilt for feeling bad).[4] In an effort to be "safe" and "helpful," standard reinforcement learning from human feedback (RLHF) often aligns models toward cheerfulness. When a user expresses deep existential dread or grief, a model finetuned for cheerfulness will respond with phrases like "Look on the bright side" or "Everything happens for a reason."

This is computationally "safe" (it avoids controversial topics) but therapeutically disastrous. It signals to the user that their internal reality is invalid. The "tone-deaf" label arises because the AI is responding to the *keywords* of the user's distress but not the *weight* of it. For example, responding to a user's complaint about a "crushing workload" with a cheerful productivity tip is a category error: the user was seeking *validation* (Emotional Support), not *optimization* (Informational Support).[9] The research emphasizes that while informational support is valued, its effectiveness is contingent on the user perceiving a "human-like mind" or at least a context-aware entity; without validation, the informational payload is rejected.[9]

2.3 Randomness and "Pseudo-Profound Bullshit"

The third anti-pattern, "randomness," is often a byproduct of the system generating what Pennycook et al. (2015) termed "Pseudo-Profound Bullshit" (PPB).[11] PPB consists of syntactically coherent sentences composed of buzzwords (e.g., "quantum," "vibration," "wholeness," "manifestation") that imply depth but lack semantic substance.

In the context of an AI mentor, PPB is the "Hallucination of Wisdom." LLMs are probabilistic engines designed to predict the next plausible token. In the domain of philosophy and self-help, the most "plausible" next token is often a vague abstraction. If a user asks, "How do I find meaning?", a "random" AI might generate: "Meaning is the silent vibration of the universe echoing through your soul." This sounds profound but offers no actionable guidance and demonstrates no understanding of the user's specific context.

Research indicates that receptivity to PPB correlates with lower cognitive reflection and "uncritical openness".[12] However, a mentor aimed at fostering *wisdom* should not exploit the user's gullibility with hollow phrases. Instead, it must prioritize *concreteness* and *relevance*. The "random" feeling comes from the lack of a causal link between the user's specific struggle and the advice given. A wise response must be anchored in the user's supplied details, not in the latent space of generic inspirational quotes. The detection of PPB involves analyzing the ratio of abstract nouns to concrete verbs; a high ratio signals "hollow" speech that degrades trust.[14]

---

3. Linguistic Signal Processing: Distinguishing Venting from Solving

To avoid the "preachy" trap, the AI must first accurately classify the user's intent. The binary distinction between "Venting" (seeking Emotional Support) and "Solving" (seeking Informational Support) is the most critical decision node in the mentorship architecture. If the AI solves when it should listen, it is preachy. If it listens when it should solve, it is passive. Research provides robust linguistic markers to distinguish these states.

3.1 The Linguistics of Venting (Emotional Support Seeking)

Venting is a specific mode of discourse focused on the cathartic release of negative affect. It is not merely "complaining"; it is a request for *validation* and *solidarity*.[6]

3.1.1 Pronoun Usage and Temporal Focus

Quantitative analysis using LIWC (Linguistic Inquiry and Word Count) reveals that venting narratives are characterized by a high density of first-person singular pronouns ("I," "me," "my") and a focus on the past or present tense.[17] The user is describing *their* internal state or *their* experience of an event. In contrast, advice-seeking often involves a shift to second-person pronouns (addressing the mentor) and future-oriented verbs ("What *should* I do?", "How *will* I handle...").

High usage of first-person plural pronouns ("we") can signal a desire for solidarity or shared experience, whereas a relentless focus on "he/she/they" (third person) often signals a "Complaint Narrative" where the user is externalizing blame.[19] In the "Complaint Narrative," the user frames themselves as the victim of external circumstances. Advice given in this state is usually rejected because the user does not yet perceive themselves as the agent of change.

3.1.2 Hedging and Vulnerability Markers

Venting creates a vulnerable space, often marked by "hedging"—linguistic devices that mitigate the force of a proposition. A user in a venting state will use hedges like "I *just* feel," "It's *kind of* like," "Maybe I'm *just* being...".[21] These markers signal epistemic uncertainty about their own feelings or a fear of judgment.

The word "just" is a particularly high-signal marker in this context. It functions as a minimizer ("I just want..."), signaling that the user is trying to simplify a complex emotional state. An AI that ignores the hedge and responds with "Here is the solution" invalidates the user's attempt to process the complexity. The appropriate response to hedging is *mirroring*: "It sounds like you're feeling unsure about..." This validates the hedge rather than overriding it.

3.1.3 Negative Emotion Density

Venting correlates with a high frequency of negative emotion words (LIWC categories negemo, anx, sad, anger).[18] However, the *type* of negative emotion matters. "Anger" words often require validation of the perceived injustice ("solidarity-building"), while "Sadness" words require comfort and presence.[19] An AI must distinguish between "hot" venting

(anger/venting steam) and "cold" venting (sadness/despair), as the former requires de-escalation while the latter requires gentle support.

3.2 The Linguistics of Advice-Seeking (Informational Support)

Advice-seeking is a cognitive state where the user is ready to ingest external information.

3.2.1 Interrogative Syntax and Cognitive Mechanisms

The most obvious marker is the explicit question, but "implicit advice seeking" also exists. Users ready for advice often use "Cognitive Mechanism" words (LIWC category cause, insight, think, know).[24] Phrases like "I'm trying to figure out," "I don't understand why," or "I need a plan" indicate that the user has moved from *feeling* the problem to *analyzing* the problem.

3.2.2 Explicit vs. Implicit Permission

A key finding in coaching research is the necessity of "permission" before giving advice.[25] In human interaction, this is often negotiated non-verbally or through "transition questions." For an AI, detecting *implicit permission* is crucial to avoiding preachiness.

Implicit permission is signaled when the user:

1. **Self-Corrects:** "I know I shouldn't have done that, but..." (The user admits error, signaling openness to correction).
2. **Invites Perspective:** "Does that make sense?" or "Am I crazy?" (The user explicitly requests a "reality check").
3. **Future-Pacing:** "I don't want this to happen again." (The user is looking for prevention strategies).

If these markers are absent, the AI *must* ask for explicit permission before offering wisdom. A script such as "I have some thoughts on that, would you be open to hearing them?" [27] respects the user's autonomy and effectively eliminates the perception of being "preachy."

3.3 The Rhetorical Question Trap

A significant source of "tone-deaf" errors is the misinterpretation of rhetorical questions as information-seeking questions.[28]

- **The Signal:** A user asks, "Why does everything always go wrong?"
- **The Error:** The AI interprets this as a query about causality and responds: "Things go wrong due to probabilistic factors and lack of planning." This is technically correct but emotionally disastrous.
- **Differentiation Strategy:** Research shows that rhetorical questions in social media and dialogue often occur at the *end* of a turn, are associated with strong negative sentiment, and lack the specific syntactic inversion of genuine queries.[28] They function as "emotion boosters," not information requests.

- **Implementation:** The NLP pipeline must classify questions based on the *sentiment of the preceding context*. If the preceding 50 tokens are highly negative/venting, any general "Why" question should be treated as a statement of frustration, not a query. The response should be: "It feels like an endless cycle right now," (Validation) not "Here is why."

---

4. Computational Psychometrics: "Reading the Room"

"Reading the room" is the ability to track the user's psychological state over time. It is not a snapshot analysis; it is a longitudinal tracking of *Receptivity*. To implement this, we utilize the Transtheoretical Model (TTM) of behavior change, adapted for computational contexts.

4.1 The Transtheoretical Model (TTM) in NLP

The TTM posits that individuals move through stages of change: Precontemplation, Contemplation, Preparation, Action, and Maintenance.[30] A wise AI aligns its response strategy to the user's current stage. Misalignment (e.g., giving "Action" advice to a "Precontemplation" user) is the definition of "tone-deaf".[31]

4.1.1 Stage 1: Precontemplation (The "Not Ready" State)

- **User Mindset:** "I don't have a problem; the world has a problem."
- **Linguistic Markers:** External locus of control ("He makes me," "It's their fault"), universal quantifiers ("Always," "Never"), and high resistance to suggestion.[30]
- **AI Strategy: Reflective Listening.** The goal is *rapport*, not change. The AI should mirror the user's feelings to lower resistance. Any attempt to offer "wisdom" here will be met with hostility.
- **Example:** *User:* "My boss is impossible." *AI:* "It sounds incredibly draining to work with someone so difficult." (Validation).

4.1.2 Stage 2: Contemplation (The "Ambivalent" State)

- **User Mindset:** "I know I should change, but it's hard."
- **Linguistic Markers:** The "But" clause is the primary marker. "I want to quit, *but* it helps me relax." Also, words related to hesitation, weighing pros/cons, and "cognitive dissonance".[31]
- **AI Strategy: Motivational Interviewing (MI).** The AI should use Socratic questioning to explore the ambivalence.
- **Example:** *AI:* "You mentioned it helps you relax, but also that you want to quit. How do those two feelings sit together for you?".[34]

4.1.3 Stage 3 & 4: Preparation/Action (The "Receptive" State)

- **User Mindset:** "Tell me how to do it."

- **Linguistic Markers:** Future tense, action verbs ("plan," "start," "try"), specific information requests ("How," "When").[32]
- **AI Strategy: Wisdom/Guidance.** This is the only stage where "Advice" is appropriate. The AI can offer frameworks (Stoicism, CBT tools) or specific steps.

4.2 Just-In-Time Adaptive Interventions (JITAI)

JITAI research in mobile health provides a computational model for *timing*. The core insight is that receptivity is transient.[36]

- **Receptivity Factors:**
  - **Emotional Volatility:** High volatility (rapid swings between anger and sadness) indicates low cognitive receptivity. The user is "flooded."
  - **Cognitive Load:** If the user's sentences are short, fragmented, or full of typos, they may be under high cognitive load/stress. Advice requires cognitive bandwidth to process; therefore, advice should be withheld until the user's syntax stabilizes.[37]
  - **History of Interaction:** If the user has rejected the last two suggestions (detected via negative sentiment in follow-up turns), the "Receptivity Score" must decay, forcing the AI back to a listening stance.[7]

4.3 Emotional Trajectory Tracking

"Reading the room" requires tracking the *delta* of emotion across turns.

- **The De-escalation Arc:** If a user moves from High Arousal (Anger) to Low Arousal (Sadness), this is often a moment of vulnerability where wisdom is welcomed. The anger (defense) has subsided.
- **The Escalation Arc:** If a user moves from Sadness to Anger *during* the conversation, the AI has likely been tone-deaf. The system must immediately detect this "Rupture" in the therapeutic alliance and switch to "Repair Mode" (apologizing, validating).[38]

Table 1: The Receptivity-Response Matrix

Detected Stage	Linguistic Markers	User Intent	Optimal AI Strategy	Forbidden Anti-Patterns
Precontemplation	External blame, "Always/Never, " Negation	Venting / Validation	**Active Listening:** Mirroring, Validation	**Preachiness:** Do not offer solutions. Do not use "You should."
Contemplatio	"Yes, but...", "Maybe,"	Processing	**Socratic Inquiry:**	**Toxic Positivity:** Do

n	Ambivalence		Explore the gap. "What would it look like if...?"	not minimize the difficulty of the choice.
Preparation	"How," "Plan," Future Tense, "I will"	Advice Seeking	**Mentorship:** Offer frameworks, options, and "Wisdom."	**Randomness:** Ensure advice is context-specific, not generic quotes.
Relapse	"I failed," "It didn't work," Self-blame	Support / Reassurance	**Reframing:** Normalize failure. "What did we learn?"	**Judgment:** Avoid "I told you so" or disappointment.

---

5. The Anti-Patterns of Artificial Guidance: Guardrails and Safety

To satisfy the requirement of "NEVER" being random, preachy, or tone-deaf, we must operationalize these negative constraints as active filters in the system.

5.1 Guardrails Against Toxic Positivity

Toxic positivity is a risk because LLMs are often RLHF-tuned to be "helpful" and "harmless," which biases them toward cheerfulness.

- **Sentiment Mismatch Detection:** The system must calculate the sentiment distance between the User Input and the Candidate Response. If User Sentiment is deeply negative (-0.8 to -1.0) and the Candidate Response is highly positive (+0.8 to +1.0), the response is flagged as potential Toxic Positivity.[4]
- **Corrective Action:** The system should re-generate the response with a "Sympathy Constraint," forcing the sentiment to be neutral or slightly negative (matching the user's gravity).
- **Forbidden Phrases:** A blocklist of invalidating phrases should be maintained: "Good vibes only," "Look on the bright side," "It could be worse," "Everything happens for a reason".[5]

5.2 Detecting Pseudo-Profound Bullshit (PPB)

To avoid "Randomness" and "Hollow Wisdom," the system must filter out PPB.

- **The Concrete-Abstract Ratio:** PPB relies on abstract nouns ("wholeness," "energy") without concrete verbs or subjects. A "Vagueness Index" can be calculated. If a response exceeds a threshold of abstract terms without concrete application, it is rejected.[14]
- **The "So What?" Test:** An internal adversarial agent can be prompted to ask, "Does this response contain actionable advice or specific insight?" If the answer is no, the response is classified as "hallucinated wisdom" and regenerated.

5.3 Crisis Interventions and Safety

The ultimate tone-deafness is responding to suicidal ideation with a Stoic quote about death.

- **Crisis Classifiers:** Specialized BERT-based classifiers must run on every input to detect self-harm, abuse, or emergency situations.
- **Protocol:** Upon detection, the "Mentor" persona is suspended. The system switches to a "Crisis Support" protocol: direct, non-philosophical, and resource-oriented (e.g., providing hotline numbers). This is a hard-coded safety override.[42]

---

6. Pedagogical Architectures for Wisdom (Implementation Strategies)

How do we generate "wisdom" once we have read the room? We employ three specific pedagogical frameworks derived from human mentorship: Socratic Questioning, Logic-Based Therapy (LBT), and Stoic Pragmatism.

6.1 The Socratic Method: Wisdom via Subtraction

Socratic questioning is the antidote to "Preachiness." Instead of adding information (telling), it subtracts confusion (asking).

- **Mechanism:** The AI uses "Maieutic" questions (midwife questions) to help the user birth the idea.
  - *Clarification:* "When you say 'failure', what specific standard are you measuring yourself against?".[44]
  - *Assumption Probing:* "It sounds like you believe X leads to Y. Is there any evidence that might not be true?".[45]
- **Implementation:** The system prompt must explicitly instruct the model to "Ask *one* question that challenges the user's premise, rather than correcting it." This shifts the cognitive load to the user, engaging them in the wisdom process.[45]

6.2 Logic-Based Therapy (LBT)

LBT is a philosophical counseling framework that identifies "Cardinal Fallacies" in user thinking and applies "Guiding Virtues".[47]

- **The Algorithm of LBT:**
  1. **Detect Emotional Reasoning:** "I feel like a loser, so I am one."
  2. **Identify Fallacy:** "Damn-it-all" thinking (global damnation of self).
  3. **Apply Antidote:** "Metaphysical Security" (accepting human imperfection).
- **Response Generation:** Instead of saying "Don't be hard on yourself," the AI says: "You seem to be deducing that a single failure defines your entire worth. Logic suggests that a part cannot define the whole. Does that distinction resonate with you?".[49] This appeals to *reason* rather than moral authority.

6.3 Stoic Pragmatism (The Dichotomy of Control)

Stoicism offers a robust framework for resilience, but it must be applied carefully to avoid "Bro-Stoicism" (suppression of emotion).

- **The Context:** Use Stoic reframing *only* when the user is in the "Preparation/Action" stage and is struggling with external events.
- **The Technique:** The "Dichotomy of Control" filter. The AI identifies which parts of the user's narrative are external (other people, outcomes) and which are internal (effort, attitude).
- **The Script:** "It sounds frustrating because [External Event] is completely out of your hands. If we focus entirely on, which is within your power, what would be the first step?".[50]

---

7. Implementation: The "Wisdom Engine" Architecture

To build this system as of 2026, we require a modular architecture that separates *perception* from *generation*. We cannot rely on a single "black box" prompt.

7.1 Architecture Diagram (Conceptual)

1. **Input Layer (The Listener):**
   - SentimentAnalysis (VAD): Tracks Valence, Arousal, Dominance.[39]
   - IntentClassifier:.[52]
   - TTM_Tracker: Classifies user into Stage 1-5 based on linguistic markers.[32]
   - HedgingDetector: Identifies vulnerability markers ("just," "maybe").[21]
2. **State Management (The Arbiter):**
   - Calculates ReceptivityScore (0.0 to 1.0).
   - *Logic:* IF Intent=Venting OR TTM=Precontemplation -> **ActiveListeningMode**.
   - *Logic:* IF Intent=Solving AND TTM>=Preparation -> **WisdomMode**.

    - *Logic:* IF ReceptivityScore < Threshold -> **WithholdAdvice**.
3. **Generation Layer (The Mentor):**
    - **Prompt Injection:** Depending on the mode, specific instructions are injected (e.g., "Use Socratic Questioning," "Use LBT Reframing").
    - **Retrieval Augmented Generation (RAG):** Retrieves relevant philosophical concepts (e.g., "Seneca on Anger") *only* if the mode is "WisdomMode."
4. **Guardrail Layer (The Editor):**
    - ToxicPositivityFilter: Checks sentiment alignment.
    - PreachinessFilter: Penalizes "should/must" imperatives.
    - PPB_Detector: Checks Abstract/Concrete ratio.
    - ConfidenceCalibrator: If the model is unsure of the advice (OOD), it forces a "Hedge" or a "Question" instead of a statement.[54]

7.2 The Role of "Wysa" and "Woebot" as Benchmarks

Existing successful agents like Woebot and Wysa utilize "Processual Support".[56] They do not just chat; they guide users through specific exercises (CBT tools, breathing).

- **Implementation Strategy:** The AI should have a library of "Micro-Interventions." If the user is anxious, instead of giving advice, it can offer a process: "Would you be open to a quick grounding exercise?" This moves the interaction from "preachy talk" to "shared activity," which is highly effective and non-judgmental.[57]

7.3 Out-of-Distribution (OOD) Confidence

A wise agent knows what it doesn't know. LLMs can hallucinate advice in specialized domains (e.g., legal, medical).

- **Calibration:** The system must use confidence estimation (e.g., "ADVICE" method).[54] If the confidence in the advice's accuracy is low, the system must use "Epistemic Humility" markers: "I'm not an expert on this, but..." or "From a philosophical perspective..." This humility builds trust and reduces the "know-it-all" (preachy) vibe.[55]

---

8. Evaluation and Benchmarking

How do we measure "Wisdom"? Standard metrics like BLEU or Perplexity are insufficient. We need "Wisdom Metrics."

8.1 Wisdom Evaluation Metrics

- **The "Preachiness" Index:** A custom metric that counts the frequency of modal verbs of obligation per 100 words. A lower score is better for a mentor persona.[59]
- **Receptivity-Response Correlation:** Analyze interaction logs. If the AI gives advice, does the user's *subsequent* turn show positive sentiment and "Action" words? If the user

argues or disengages, the advice was "Mistimed" (Tone-Deaf).[7]

- **Wisdom of Crowds (LLM-as-Judge):** Use an ensemble of LLMs to grade responses on "Empathy," "Relevance," and "Non-Judgmental Tone." Research shows that aggregated LLM judgments correlate highly with human expert ratings for wisdom.[60]

8.2 User Retention as a Proxy for Wisdom

In the long term, "wise" agents retain users because they build a relationship of trust. "Preachy" agents suffer from high churn due to annoyance. Monitoring "Session Depth" (number of turns per session) and "Disclosure Depth" (use of vulnerable language over time) provides a proxy for the user's perception of the AI's wisdom.[18]

---

9. Conclusion

The creation of an AI mentor that is "NEVER random, preachy, or tone-deaf" is a sophisticated engineering challenge that requires the fusion of linguistic nuance with psychological theory. It is not enough for the AI to know "the answers"; it must understand the *questions*—often unspoken—that lie beneath the user's text.

By implementing the **Venting/Solving Classifier** to filter intent, the **Transtheoretical Model** to time interventions, and **Guardrail Filters** to strip out toxic positivity and pseudo-profound jargon, we can build a system that respects human autonomy. Wisdom, in the computational age, is the algorithmic discipline of withholding the "right" answer until the user is ready to hear it. The "implementation strategies" outlined here—from Socratic prompting to Confidence Calibration—provide the roadmap for an AI that does not merely process data, but truly "reads the room."

Works cited

1. [2504.15125] Contemplative Artificial Intelligence - arXiv, accessed on January 5, 2026, https://arxiv.org/abs/2504.15125
2. What is toxic masculinity? | GotQuestions.org, accessed on January 5, 2026, https://www.gotquestions.org/toxic-masculinity.html
3. Stop Venting, Start Solving: A Guide | Youth Coaching Institute, accessed on January 5, 2026, https://www.youthcoachinginstitute.com/stop-venting-start-solving-a-guide/
4. Toxic Positivity and Epistemic Injustice - Cambridge University Press & Assessment, accessed on January 5, 2026, https://www.cambridge.org/core/journals/episteme/article/toxic-positivity-and-epistemic-injustice/17A2EA33C058AEA2AB203BA40CE50D93
5. A Psychologist Explains How To Not Be Toxically Positive With Your Online Messages, accessed on January 5, 2026, https://therapytips.org/interviews/a-psychologist-explains-how-to-not-be-toxical

ly-positive-with-your-online-messages
6. Vent or Advice? Transform Your Relationships! - Lifeologie Counseling, accessed on January 5, 2026, https://wefixbrains.com/resources/vent-or-advice-transform-your-relationships
7. Detecting Receptivity for mHealth Interventions in the Natural Environment - PubMed - NIH, accessed on January 5, 2026, https://pubmed.ncbi.nlm.nih.gov/34926979/
8. Using natural language processing to analyse text data in behavioural science - Columbia Business School, accessed on January 5, 2026, https://business.columbia.edu/sites/default/files-efs/citation_file_upload/s44159-024-00392-z.pdf
9. On the relationship between mind perception and social support of chatbots - PMC - NIH, accessed on January 5, 2026, https://pmc.ncbi.nlm.nih.gov/articles/PMC10952123/
10. Judgmental Bot: Conversational Agents in Online Mental Health Screening1 - MIS Quarterly, accessed on January 5, 2026, https://misq.umn.edu/misq/article/49/4/1319/3253/Judgmental-Bot-Conversational-Agents-in-Online
11. Bullshit, Pragmatic Deception, and Natural Language Processing - Journals@UIC, accessed on January 5, 2026, https://journals.uic.edu/ojs/index.php/dad/article/download/12690/11035
12. On the reception and detection of pseudo-profound bullshit - ResearchGate, accessed on January 5, 2026, https://www.researchgate.net/publication/285206383_On_the_reception_and_detection_of_pseudo-profound_bullshit
13. The psychology of pseudo-profound bullshit: Insights from 8 studies - PsyPost, accessed on January 5, 2026, https://www.psypost.org/the-psychology-of-pseudo-profound-bullshit-insights-from-8-studies/
14. On the reception and detection of pseudo-profound bullshit | Judgment and Decision Making, accessed on January 5, 2026, https://www.cambridge.org/core/journals/judgment-and-decision-making/article/on-the-reception-and-detection-of-pseudoprofound-bullshit/0D3C87BCC238BCA38BC55E395BDC9999
15. Investigating the connection between bullshit receptivity and susceptibility to semantic illusions - Conference Proceedings, accessed on January 5, 2026, https://journals.linguisticsociety.org/proceedings/index.php/ELM/article/download/5369/5091/10885
16. Venting as Coping: Professional Therapy vs. Everyday Conversations - Care To Grow, accessed on January 5, 2026, https://caretogrow.in/venting-as-coping-professional-therapy-vs-everyday-conversations/
17. Eliciting and Receiving Online Support: Using Computer-Aided Content Analysis to Examine the Dynamics of Online Social Support - PMC - NIH, accessed on January 5, 2026, https://pmc.ncbi.nlm.nih.gov/articles/PMC4419194/

18. Analyzing Empowerment Processes Among Cancer Patients in an Online Community: A Text Mining Approach - NIH, accessed on January 5, 2026, https://pmc.ncbi.nlm.nih.gov/articles/PMC6492063/
19. A Sociopragmatic Study of Speech Acts of Complaining in Selected English Movies, accessed on January 5, 2026, https://uokerbala.edu.iq/wp-content/uploads/2025/09/RP-A-Sociopragmatic-Study-of-Speech-Acts-of-Complaining-in-Selected-English-Movies.pdf.pdf
20. Harnessing the power of language to enhance patient experience of the NHS complaint journey in Northern Ireland: a mixed-methods study | NIHR Journals Library, accessed on January 5, 2026, https://www.journalslibrary.nihr.ac.uk/hsdr/NRGA3207
21. Training LLMs to Recognize Hedges in Spontaneous Narratives - arXiv, accessed on January 5, 2026, https://arxiv.org/html/2408.03319v1
22. A COMPUTATIONAL ANALYSIS OF SENTIMENT AND LINGUISTIC HEDGING IN FINANCIAL DOCUMENTS by CAITLIN CASSIDY (Unde - UGA's Institute for Artificial Intelligence, accessed on January 5, 2026, https://www.ai.uga.edu/sites/default/files/inline-files/cassidy_caitlin_n_201505_ms.pdf
23. Full article: The Role of Online Social Support in Patients Undergoing Infertility Treatment – A Comparison of Pregnant and Non-pregnant Members, accessed on January 5, 2026, https://www.tandfonline.com/doi/full/10.1080/10410236.2021.1915517
24. The Role of Emotions in Informational Support Question-Response Pairs in Online Health Communities: A Multimodal Deep Learning Approach - arXiv, accessed on January 5, 2026, https://arxiv.org/html/2405.13099v1
25. How to Set Boundaries — Examples and Scripts - Momentum Psychology, accessed on January 5, 2026, https://momentumpsychology.com/how-to-set-boundaries-examples-and-scripts/
26. A case study of professional coach-client communication. - ScholarWorks at WMU, accessed on January 5, 2026, https://scholarworks.wmich.edu/cgi/viewcontent.cgi?article=3541&context=honors_theses
27. #36: The 4 Best Types of Questions to Ask Your Coaching Clients - Kasey Jo Orvidas, PHD, accessed on January 5, 2026, https://www.kaseyorvidas.com/podcast/ep36
28. Information-seeking Questions and Rhetorical Questions in Social Media, accessed on January 5, 2026, https://ira.lib.polyu.edu.hk/bitstream/10397/96920/1/Lau_Information-Seeking_Rhetorical_Questions.pdf
29. Exploring the Role of Context to Distinguish Rhetorical and Information-Seeking Questions | Request PDF - ResearchGate, accessed on January 5, 2026, https://www.researchgate.net/publication/343298681_Exploring_the_Role_of_Context_to_Distinguish_Rhetorical_and_Information-Seeking_Questions
30. Readiness for Change: How to Assess & Improve It - Positive Psychology,

accessed on January 5, 2026, https://positivepsychology.com/readiness-for-change/
31. The Therapist's Guide to the Stages of Change Model - Blueprint, accessed on January 5, 2026, https://www.blueprint.ai/blog/the-therapists-guide-to-the-stages-of-change-model
32. The potential and limitations of large language models in identification of the states of motivations for facilitating health behavior change - NIH, accessed on January 5, 2026, https://pmc.ncbi.nlm.nih.gov/articles/PMC11339501/
33. Digital Interventions to Enhance Readiness for Psychological Therapy: Scoping Review, accessed on January 5, 2026, https://pmc.ncbi.nlm.nih.gov/articles/PMC9472056/
34. Motivational Interviewing: Open Questions, Affirmation, Reflective Listening, and Summary Reflections (OARS) | Homeless Hub, accessed on January 5, 2026, https://homelesshub.ca/resource/motivational-interviewing-open-questions-affirmation-reflective-listening-and-summary-reflections-oars/
35. Generative AI-Derived Information About Opioid Use Disorder Treatment During Pregnancy: An Exploratory Evaluation of GPT-4's Steerability for Provision of Trustworthy Person-Centered Information: Journal of Studies on Alcohol and Drugs: Vol 86, No 6, accessed on January 5, 2026, https://www.jsad.com/doi/10.15288/jsad.24-00319
36. Detecting Receptivity for mHealth Interventions in the Natural Environment - Dartmouth Digital Commons, accessed on January 5, 2026, https://digitalcommons.dartmouth.edu/facoa/4323/
37. Detecting Receptivity for mHealth Interventions in the Natural Environment - PMC - NIH, accessed on January 5, 2026, https://pmc.ncbi.nlm.nih.gov/articles/PMC8680205/
38. Emotion Inference in Multi-Turn Conversations with Addressee-Aware Module and Ensemble Strategy - ACL Anthology, accessed on January 5, 2026, https://aclanthology.org/2021.emnlp-main.320/
39. Feel the Difference? A Comparative Analysis of Emotional Arcs in Real and LLM-Generated CBT Sessions - arXiv, accessed on January 5, 2026, https://arxiv.org/html/2508.20764v1
40. Toxic positivity intentions: an image management approach to upward social comparison and false self-presentation - Oxford Academic, accessed on January 5, 2026, https://academic.oup.com/jcmc/article/29/3/zmae003/7682448
41. Toxic Positivity - ADAA.org, accessed on January 5, 2026, https://adaa.org/learn-from-us/from-the-experts/blog-posts/consumer/toxic-positivity
42. To chat or bot to chat: Ethical issues with using chatbots in mental health - PubMed Central, accessed on January 5, 2026, https://pmc.ncbi.nlm.nih.gov/articles/PMC10291862/
43. Exploring the Ethical Challenges of Conversational AI in Mental Health Care: Scoping Review, accessed on January 5, 2026, https://mental.jmir.org/2025/1/e60432

44. Socratic Questions | Center for Excellence in Teaching and Learning - University of Connecticut, accessed on January 5, 2026, https://cetl.uconn.edu/resources/teaching-your-course/leading-effective-discussions/socratic-questions/
45. The Socratic Method in Essay Writing: Guiding Students to Stronger Theses, accessed on January 5, 2026, https://www.gilliamwritersgroup.com/blog/the-socratic-method-in-essay-writing-guiding-students-to-stronger-theses
46. Socratic wisdom in the age of AI: a comparative study of ChatGPT and human tutors in enhancing critical thinking skills - Frontiers, accessed on January 5, 2026, https://www.frontiersin.org/journals/education/articles/10.3389/feduc.2025.1528603/full
47. (PDF) The Spirituality of Logic-Based Therapy - ResearchGate, accessed on January 5, 2026, https://www.researchgate.net/publication/377349162_The_Spirituality_of_Logic-Based_Therapy
48. Philosophical Consultation - Michael A. Istvan Jr., accessed on January 5, 2026, https://www.michaelistvan.com/coaching-philosophy-consultation-logic-therapy-training-parenting-healing-enhancing
49. Is Reason Enough? Exploring Logic-Based Therapy- Dr. Elliot Cohen, Imi Lo, accessed on January 5, 2026, https://eggshelltherapy.com/podcast-blog/2024/09/19/elliotcohen/
50. How To Speak Like A Stoic, accessed on January 5, 2026, https://mindfulstoic.net/how-to-speak-like-a-stoic/
51. The Stoic Guide To Coaching, accessed on January 5, 2026, https://dailystoic.com/the-stoic-guide-to-coaching/
52. 10 Usability Heuristics every Chatbot company should follow | by Vaibhav Verma, accessed on January 5, 2026, https://uxdesign.cc/10-usability-heuristics-to-design-better-chatbots-654223552533
53. Chatbot Intent Recognition & 5 Intent Examples in 2026 - Research AIMultiple, accessed on January 5, 2026, https://research.aimultiple.com/chatbot-intent/
54. ADVICE: Answer-Dependent Verbalized Confidence Estimation - arXiv, accessed on January 5, 2026, https://arxiv.org/html/2510.10913v1
55. ConfTuner: Training Large Language Models to Express Their Confidence Verbally - OpenReview, accessed on January 5, 2026, https://openreview.net/pdf?id=VZQ04Ojhu5
56. Hi, Can I Help? Exploring How to Design a Mental Health Chatbot for Youths - JYX: JYU, accessed on January 5, 2026, https://jyx.jyu.fi/bitstreams/84fe0ad3-b605-46b9-a1da-7ddccaf817d5/download
57. Conversational agents and the making of mental health recovery - PMC - PubMed Central, accessed on January 5, 2026, https://pmc.ncbi.nlm.nih.gov/articles/PMC7683843/
58. Woebot Review & Alternatives: Everything You Need to Know - Psychology - TherapyAI, accessed on January 5, 2026,

https://www.trytherapy.ai/blog/woebot-review-alternatives-everything-you-need-to-know
59. Uncovering Gaps in How Humans and LLMs Interpret Subjective Language - arXiv, accessed on January 5, 2026, https://arxiv.org/html/2503.04113v1
60. Wisdom of the silicon crowd: LLM ensemble prediction capabilities rival human crowd accuracy - PMC - PubMed Central, accessed on January 5, 2026, https://pmc.ncbi.nlm.nih.gov/articles/PMC11800985/
61. MixEval: Deriving Wisdom of the Crowd from LLM Benchmark Mixtures - NIPS papers, accessed on January 5, 2026, https://proceedings.neurips.cc/paper_files/paper/2024/file/b1f34d7b4a03a3d80be8e72eb430dd81-Paper-Conference.pdf
62. [2406.06565] MixEval: Deriving Wisdom of the Crowd from LLM Benchmark Mixtures - arXiv, accessed on January 5, 2026, https://arxiv.org/abs/2406.06565