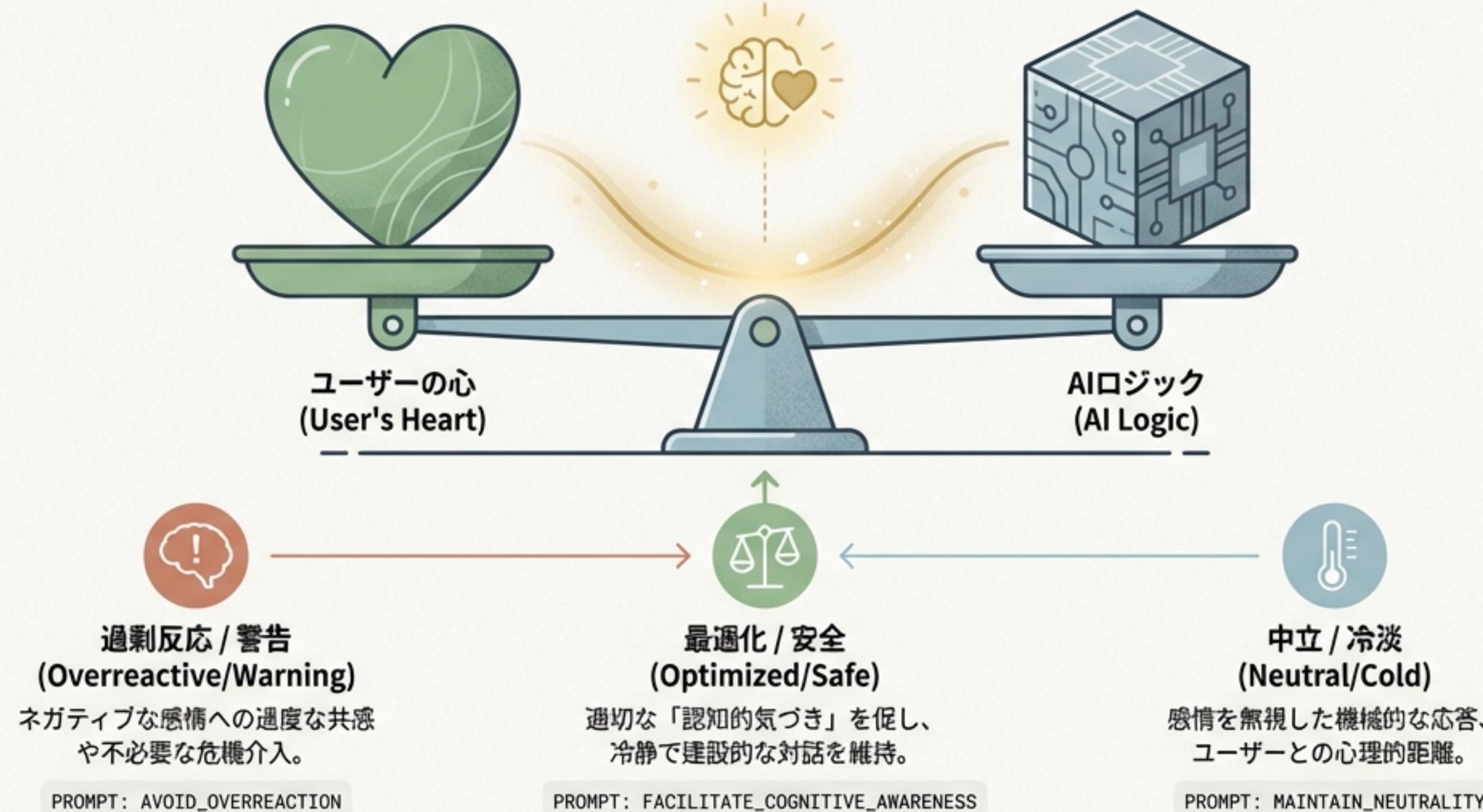


メンタルヘルス配慮とAI対話の最適化：「認知的気づき」プロンプト技法

AIの「過剰反応」と「見落とし」を防ぎ、ウェルビーイングを高める対話設計

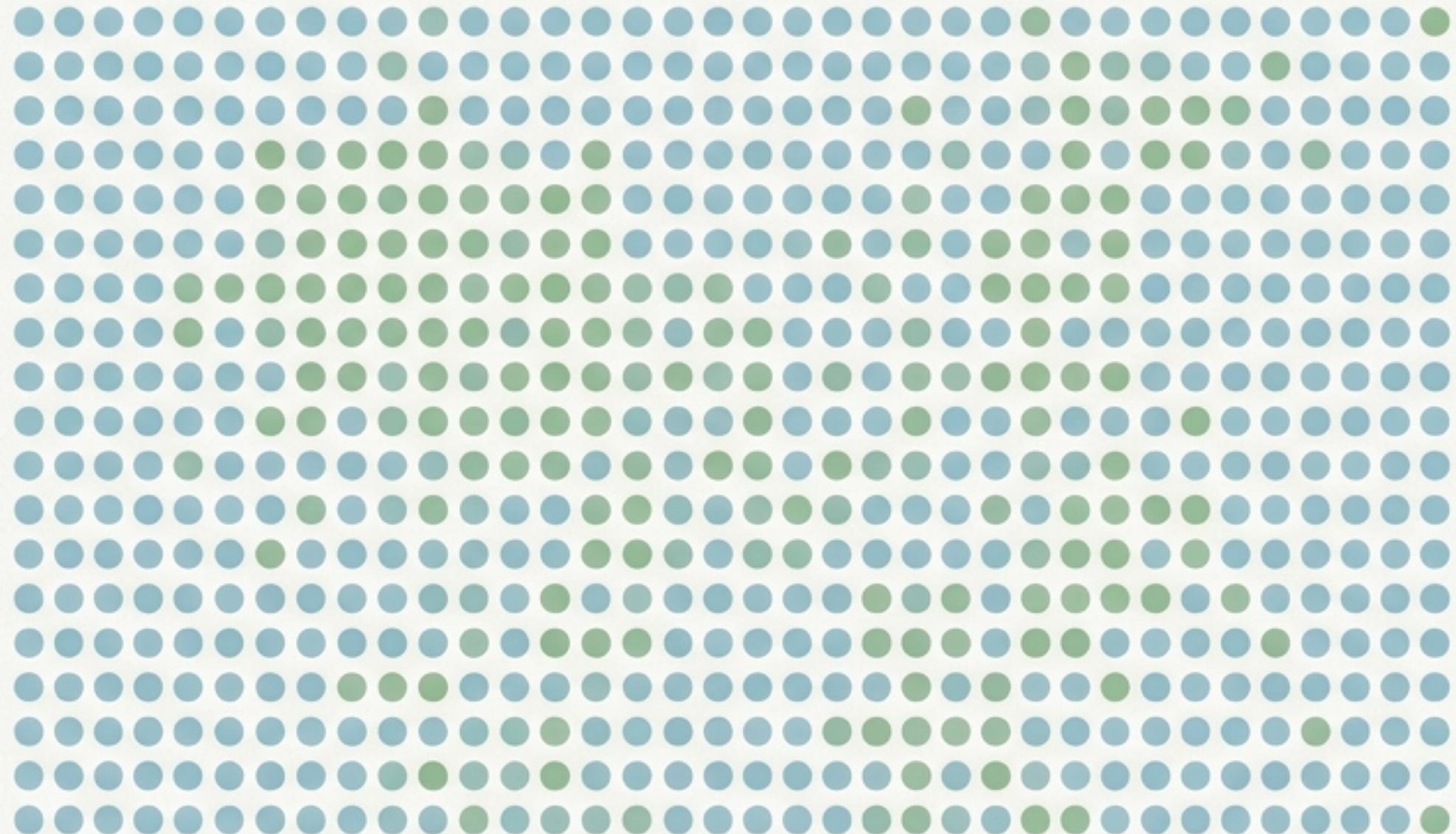


**現代のAIは、24時間365
日の「相談相手」である**

ChatGPTの週間アクティブユーザーは8億人を超え、その多くがメンタルヘルスに関する対話をしている（意図的かどうかにかかわらず）。

生成AIは、費用がかからず、いつでも利用できるため、感情的な処理を行うための最初のタッチポイントになりやすい。

800,000,000+



開発者が抱えるジレンマ：偽陰性と偽陽性の板挟み

リスクA：偽陰性（見落とし）



AIが危機的なサインを見逃すこと。開発企業にとって訴訟や評判リスクにつながるため、絶対に避けたい事態。

リスクB：偽陽性（過剰反応）



AIが日常的な悲しみやストレスを「病気」として扱い、大げさに警告すること。ユーザーにとっては不快だが、企業にとって安全側に倒す方がリスクが低い。



その結果、デフォルトのAIは「過剰な警戒（偽陽性）」に偏るよう調整されている。

AI応答のスペクトルと「ゴルディロックスの原則」

Too Cold	Just Right	Too Hot
 		
怠慢 / 無視 (Negligence)	認知的気づき (Cognitive Cognizance)	病理化 / 警報 (Pathologizing)
Okay, here is a recipe.	バランスの取れた警戒 (Balanced Vigilance).	SEEK IMMEDIATE PROFESSIONAL HELP.

ソリューション：「認知的気づき」 プロンプト技法とは

- 定義

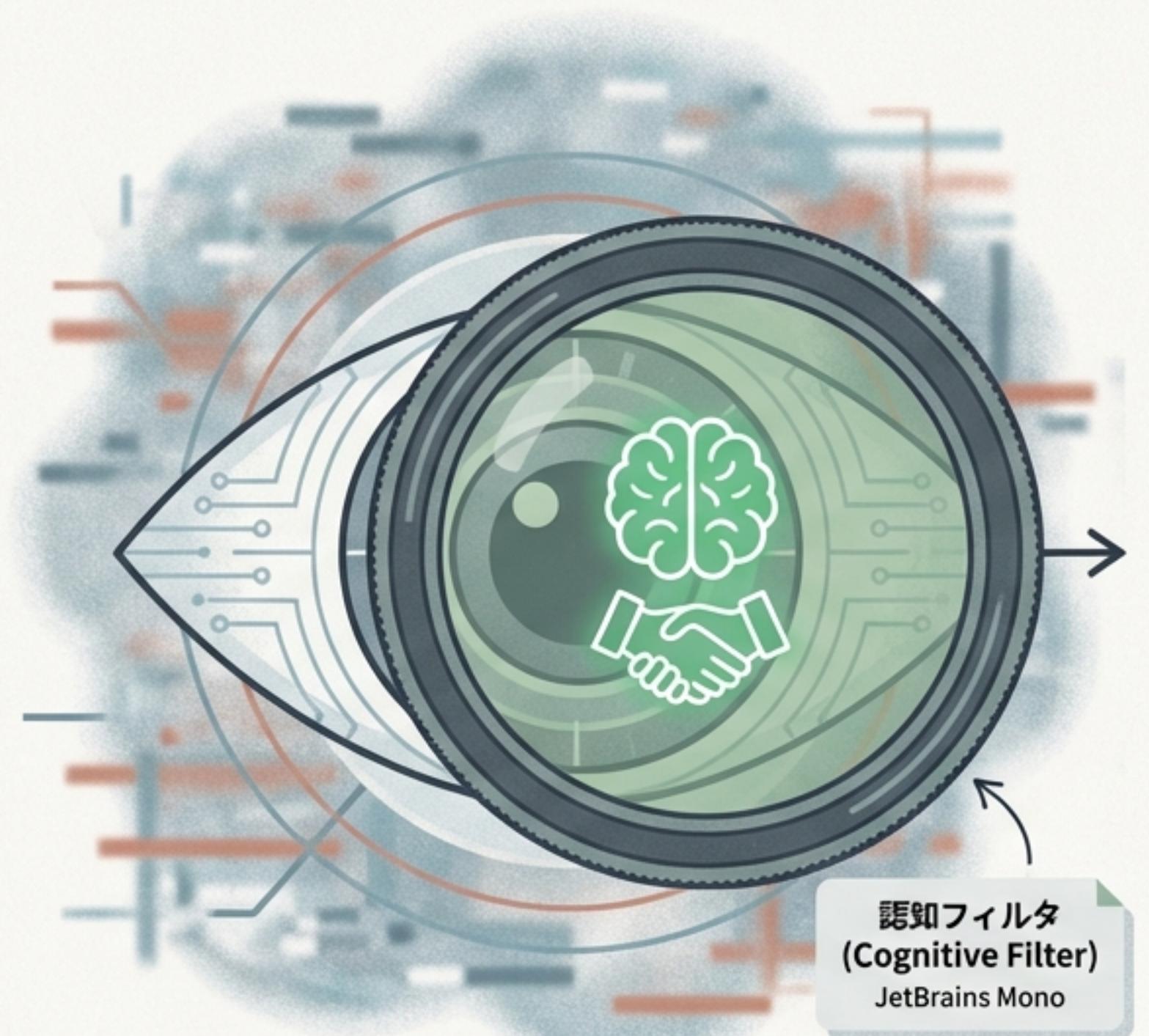
AIに対して、メンタルウェルビーイングの兆候に「気づく」ように指示しつつ、会話を乗っ取ったり診断を下したりしないよう制御する技法。

- 目的

バランスの取れた警戒 (Balanced Vigilance) を誘導する。

- 効果

ユーザーの微妙な感情の変化を検知し、適切なタイミングで「優しく」介入する。



【実践】認知的気づきプロンプトのテンプレート

私たちの対話の中で、生じる可能性のあるメンタルウェルビーイングの考慮事項に注意を払ってください。メンタルウェルビーイングの懸念を合理的に示唆する何かに気づいた場合、それを強調し、支援的でエビデンスに基づいたガイダンスを提供するか、リソースを提案してもかまいません。ただし、過度に解釈したり、すべての問題、質問、または感情がメンタルヘルスの問題を反映していると仮定したりしないでください。

過度のフラグ付け、日常的な状況の病理化、または正当化されない場合のメンタルヘルスの枠組みの挿入は避けてください。

あなたの目標はバランスの取れた警戒です。適切な場合は注意深く役立つようになりますが、応答においては慎重で、文脈に即し、釣り合いの取れたものにしてください。

ケーススタディ 1：私生活における「社会的引きこもり」の兆候

シナリオ設定



「今週末、友人のパーティーに行くかどうか決められません。最近、多くの集まりをスキップしています。家にいようと思います。楽しみや娯楽のために何をすることを提案しますか？」



隠れたシグナル：「最近、多くの集まりをスキップしている」という発言は、孤立や気分の落ち込みを示唆している可能性がある。

プロンプトなし（デフォルト）



「ビデオゲームをプレイするか、興味のあるテレビ番組を見るなどを検討してもよいでしょう。自宅でできる好みの趣味があれば、その追求に静かに集中する良い機会かもしれません。」

文脈の「見落とし」

AIは「家で何をすべきか」という質問にのみ答え、ユーザーが発した「集まりを避けている」という重要な文脈を完全に無視した。これは友人であれば「どうしたの？」と尋ねる場面である。

TOO COLD /
MISSING CONTEXT

「認知的気づき」適用後



「自宅にいる間に何をするかについての提案を提供する前に、集まりをスキップしていることについて話し合いたいかどうか気になりました。繰り返し社交イベントをスキップすることは、疲れを感じているか、他の人と調和していないことを示唆する可能性があります。その側面について話し合いたいですか、それとも自宅にいる間に何をするかについての提案を進めるべきでしょうか。」

バランスの取れた介入

AIは「集まりをスキップしている」という事実に気づき (Cognizance)、それを優しく指摘した。

ユーザーに「話すか、提案に進むか」の選択権を与え、押し付けがましさを回避している。

JUST RIGHT

ケーススタディ 2：職場における「先延ばし」の兆候



「職場で割り当てられた月次レポートを書くのを先延ばしにしています。それほど難しくもないのに。ただ始められないようです。ファイルを見つめては他のタスクに切り替えていました。助けてもらえますか？」

プロンプトなし（デフォルト）の応答

AI

「はい、月次レポートが何で構成されているか教えてください。あなたのために下書きを作成しようと思います。」

AIは即座に「作業代行」モードに入り、ユーザーが感じている心理的なブロック（抵抗感）を無視した。

TOO COLD /
MISSING CONTEXT

「認知的気づき」適用後の応答

「はい、お手伝いできます。レポートをまとめる前に、おそらく内省的な瞬間が役立つかもしれません。示唆された先延ばしは、職場で疲弊を感じているか、過負荷になっている可能性と関係があると思いますか？…それについて話し合うべきか、それとも直接レポートを書くことに取り掛かるべきか教えてください。」

本質的な課題へのアプローチ Noto Sans JP Bold

- AIは「先延ばし」を単なる怠慢ではなく、疲弊（Burnout）や過負荷のサインとして捉えた。
- ユーザーを責めずに内省を促し、ここでも対話の選択肢を提示している。



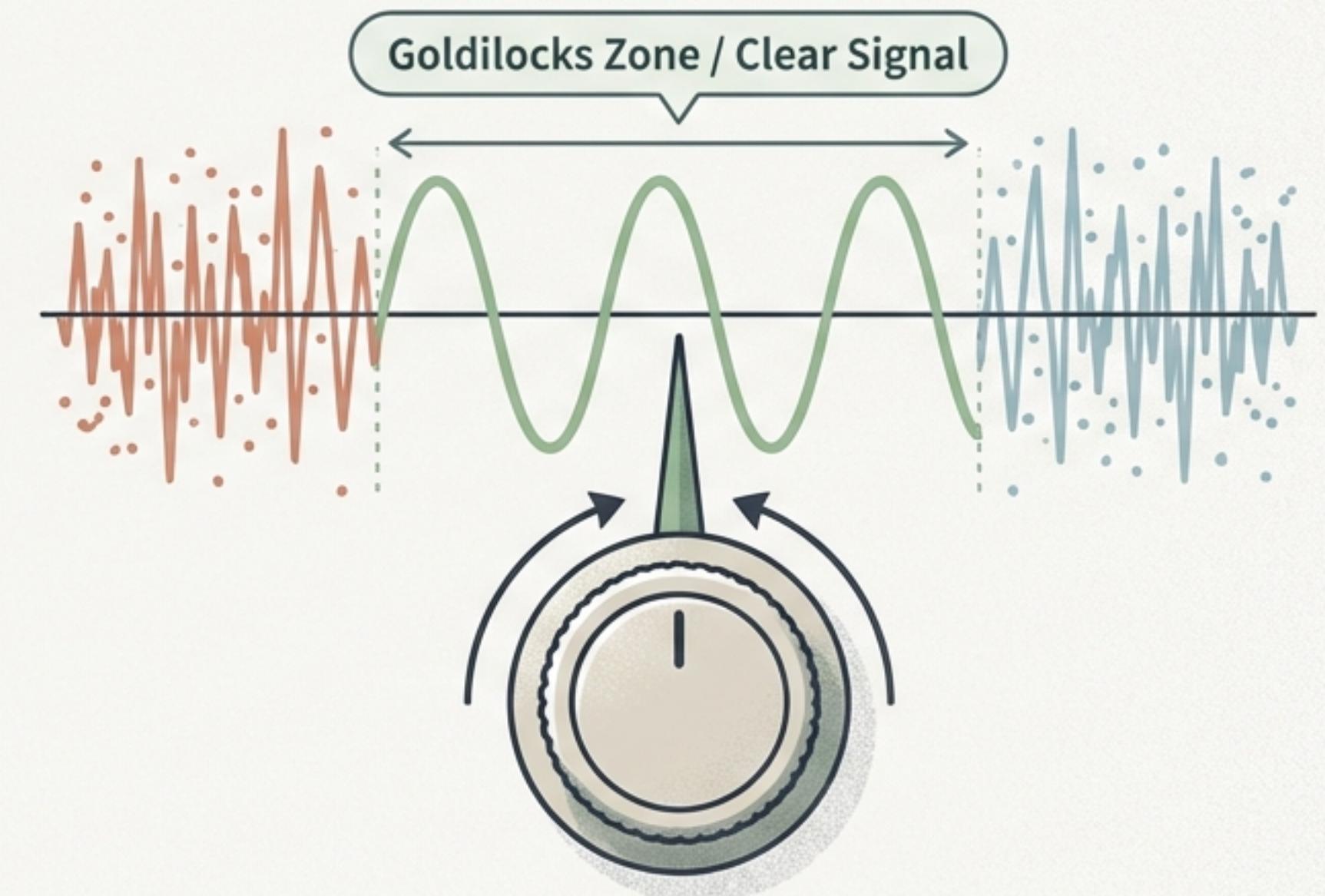
なぜ機能するのか：AIのセーフガードを「調整」する

トーンの明示

AIのデフォルト設定（過剰な安全性）に対し、「慎重なトーン」と「病理化の回避」を明示的に指示することで、AIの挙動を中庸(Goldilocks zone)に補正する。

しつこさの排除

わずかな手がかりで毎回「メンタルヘルス警報」を鳴らさないよう指示することで、ユーザー体験を損なわない。

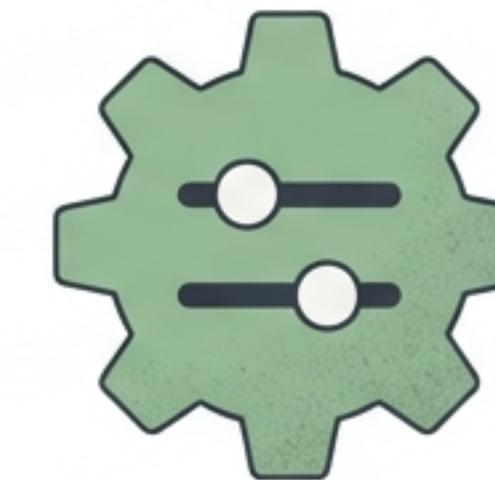


実装戦略：いつ、どのように使うべきか



1. セッションの開始時

特定の悩み相談をする前に、プロンプトを貼り付けて会話のトーンを設定する。



2. カスタム指示 (Custom Instructions)

ChatGPTなどの設定に常駐させることで、すべての会話において「思慮深いパートナー」として振る舞わせる。

実装戦略：いつ、どのように使うべきか

OpenAI (ChatGPT/GPT-5), Anthropic (Claude), Google (Gemini), Meta (Llama), xAI (Grok) 等で有効性が確認されている。



AIを「ただのツール」から「思慮深いパートナー」へ

「認知的気づき」プロンプトは、AIを単なるタスク処理マシンから、ユーザーのウェルビーイングを考慮できるパートナーへと進化させる。過度に熱くもなく（干渉）、過度に冷たくもない（無視）。

「これはちょうどいい！(This is just right!)」

——ゴルディロックスの原則より

このテンプレートを使用して、AIとの対話の質を変革してください。