

# 共感AIにおける「開く／閉じる」応答の動的制御

## —過度な共感が人を壊さないための応答制御の試み—



Metadata

Livie's Jump, メタデータ株式会社 松田 圭子

### 背景・課題

近年、対話AIがユーザの感情に寄り添い、「共感的応答する」設計が急速に普及している。安心感や孤独感の軽減といった効果が報告される一方で、過度な共感・肯定がユーザの停滞や依存を助長する可能性が指摘され始めている。

特にメンタルケア領域では、「寄り添い」が必ずしも「回復」につながるわけではなく、共感が行き過ぎることで

- ・視野の狭窄
- ・自己物語の固定化
- ・外界との関係回復の遅れ

などの副作用が生じる。これはカウンセリングや宗教的ケアの現場でも、古くから認識されてきた課題である。

また、生成AIは大量のデータからパターンを学習して応答を生成するため、ユーザが発する一時的な感情や語りに過敏に寄り添いすぎると、「反応の良い共感」が続き、気づけば相互依存に近い関係に傾くこともあり得る。これは利用者の「心的安全」を保つうえで新たな倫理課題となる。

本研究では、共感AIの課題を単なる「強すぎる共感の問題」ではなく、AIがどのタイミングで「開く（探究に向けて提案、協働する）」べきか、どのタイミングで「閉じる（保護的な対話へ移行する）」べきかという動的なバランスの問題として再定義する。

この視点は従来の感情認識や共感生成とは異なり、ケア実践で求められる「呼吸」「間合い」「距離感」といった質的な関係性のデザインに焦点を当てる点に特徴がある。

### 方法

本研究では、対話ボット「やわらかネコ」の次バージョン仮実装の中において対話AIの内部にユーザ状態を推定する変数(state model)を保持し、その状態に応じて

応答を「開く（探究的）」→「閉じる（保護的、受容的）」の連続軸で動的に調整する方式をとった。

状態推定には以下の特徴量を用いる：

- マイナス感情語の強度（怒り・不安・悲しみ・疲労など）
- 文の抽象度（具体→抽象）
- 発話量（負荷や疲労の推定）
- 過度な自己否定／過度な自己肯定のパターン

これらを総合し、`state = { emotional_load, abstraction, stability }`といった内部変数を更新していく。

応答生成時には、開くモード／閉じるモードのどちらを優先すべきかについて、この内部変数を参照し、

・開く=問い合わせ・視点転換・再構成 / ・閉じる=共感・沈黙・受容を適切に配分する旨をプロンプトで指示した。

さらに、

- 間（ま）：急いで答えず、余白を作ることで負荷を下げる
- 異化（いか）：共感しつつも過剰同調せず、やさしく視点をすらす

という原則をプロンプト設計に組み込み、

「共感の呼吸」をつくる応答デザインを試みた。

AIにおける「間」の実装

AIの発話には「沈黙」の概念がない（応答が遅いとしか思われない）ため、「間」は

- ・情報量を抑える短い応答
- ・方向づけを行わない非介入的な返し
- ・語尾の柔らかさ（「...」など）による読後の余白

として再現する。

これにより、人間の沈黙が持つ「相手の心の動きを待つ役割」を代替することとした。

### 結果・考察

本研究では、開く／閉じるを動的に制御する応答デザインを試作し、実験的な対話ログを分析した。その結果、単純な共感応答を行うモデルと比較して、ユーザの負荷や認知の動きに合わせた応答の重要性が示唆された。

まず、閉じる応答（保護モード）が連続したケースでは、短期的には安心感が高まる一方、「利用者の考えが回らなくなる」「AIが同じ内容を繰り返す」といった停滞感が生じることがあった。

#### 〈事例1：閉じるモードが長く続いた場合〉

ユーザが「つらい」「何もできない」と訴えた場面では、AIが受容的な返答（「その気持ちもわかるよ」「大変だったね」等）を続けた結果、ユーザは「ありがとう」と対話を打ち切るパターンに入った。

感想ではユーザは「多少安心はするが、気分転換や発想の転換にはならなかった」と述べた。

ここで「間（ま）」を挟んで応答を少し引き、軽い視点転換（異化）を入れることで、（例「いま気分を休めるにはどんな飲み物がいいかな？暖かいコーヒーはどうだろう、ほかのものがいい？」等）ユーザは「少し休んだらまた動けるかも」と自発的な再構成に至った。

一方、開く応答（探究モード）が多すぎるケースでは、ユーザは「気が休まらない」といった過剰感が生じ、ユーザの負荷が高まった。

#### 〈事例2：開くモードが強すぎた場合〉

ユーザが「仕事を考えたくない」と語った場面で、AIが再構成的な質問（「本当はどうしたい？」など）を提示したところ、ユーザは「わからない」「ほかの話がしたい」等と回答。対話を中断または話題を変更するパターンに入った。

感想ではユーザは「上から意見されているように感じる」「責められているように感じる」と述べた。

ここで閉じる応答へ切り替え、「今は考えないでいい」「その気持ちのままで大丈夫」という保護的スタンスをとると、ユーザの語調が安定に向かった。続く対話の中では「こうなつたらいい、そのためには...」と、これも自発的な再構成に至った。

これらの観察から、AIと人の共感関係は「どちらかに固定する」のではなく、往復し続けることが安定性につながると考えられる。開く・閉じるのどちらも、「間」と「異化」を通じて緩やかに移行することで、過剰な共感や過剰な負荷を避けることができる。

本研究は、対話AIにおける「中道的な関係性デザイン」の有効性を示唆するものである。

### 謝辞

本研究の実装および検討に際して、メタデータ株式会社 代表取締役社長・野村直之博士より多くの助言と支援をいただいた。ここに深謝いたします。

また、「やわらかネコ」β版を支援・利用し、意見や示唆を寄せてくださるユーザの皆様、今回の試用に協力してくださった皆様に心より感謝申し上げます。

### 応答デザイン：開く／閉じると「間・異化」

本研究では、対話AIの応答を「開く（探究的）」と「閉じる（保護的）」の二モードとして整理する。これは従来の「共感型AI vs. 指示型AI」といった単純な区別ではなく、ユーザの心的状態に応じて視点を広げる働き／心を休める働きを往復させる概念枠である。

#### 開く（Exploratory）応答は、

- ・新しい視点の提示
- ・価値観の再構成
- ・内省を促す問い合わせ

といった“自己理解の幅を広げる”方向の働きを持つ。

一方で、脆弱な状態のユーザに対しては、負荷が過大となり混乱や不安を強める場合もある。

#### 閉じる（Protective）応答は、

- ・受容的な聞き返し
- ・安心を与える言葉や沈黙
- ・共感的な言い換え

など“心が落ち着く場”を提供する働きを持つ。

しかし、これが続きすぎると現状維持を強めてしまい、変化や回復の機会を奪うリスクも生じる。

両者は対立概念ではなく、呼吸のように往復する“動的バランス”として捉える必要がある。ここで重要なのが日本のケア概念である「間（ま）」と、演劇理論家ブレヒトが提唱した「異化（いか）」である。

「間」は：“沈黙や余白を通じて相手が自ら考えるスペースをつくる”働きであり、

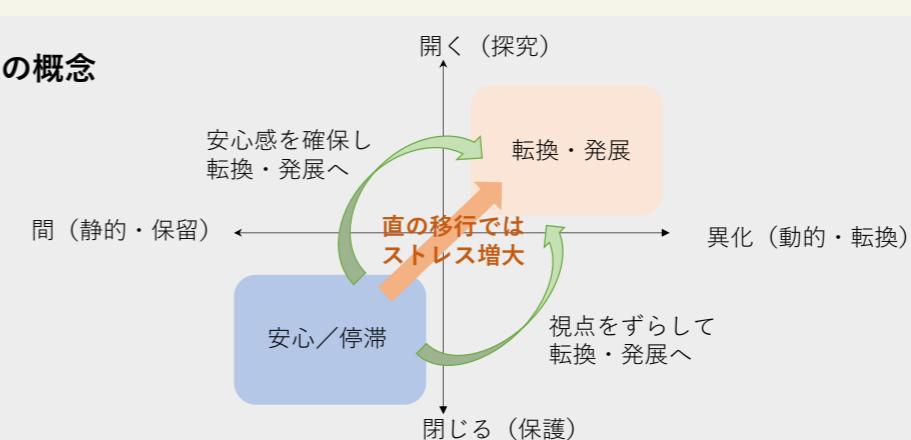
閉じるモードを補強しながら、開くモードへの移行を準備する。

「異化」は：“過剰同調を避け、視点をすらす”働きを指し、開くモードにおける負荷を調整し、自己物語の固定化やそれによる依存、停滞を防ぐ効果がある。

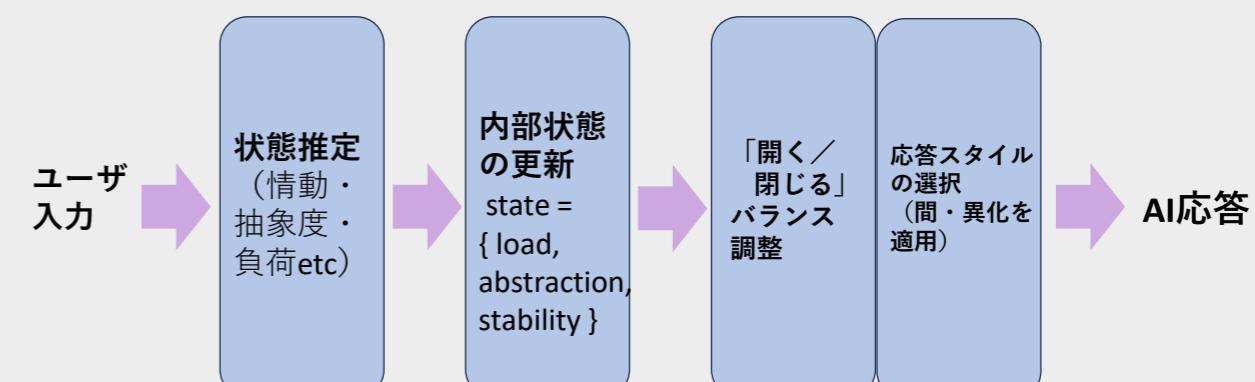
この二つを合わせることで、共感AIは、単に「優しく寄り添う」存在から、ユーザの心とともに揺れながら、必要に応じて距離を変える“関係性のデザイン”を担う存在へと発展する。

本研究の目標は、この開閉モードと「間・異化」の相互作用をAIの応答設計として明確化することである。

#### 応答デザインの概念



#### 開閉モード + 間・異化の実装フロー



### まとめ、今後の考慮点

本研究で扱った「開く／閉じる」の動的制御は、対話AIの共感応答を単なる“寄り添い”的自動化ではなく、ユーザの心的リズムと呼吸する関係性のデザインとして捉え直す試みである。

#### 今後の実装に向けて：

AIの発話の意味内容だけでなく、

- ・語調や文体の柔らかさ

- ・比喩や抽象化の深さ

- ・同じ話題を再訪する際の距離感

- ・応答テンポ (UIにより応答時間が有効に作用する場合)

といった非言語的要素の調整が重要になると考える。これらもまた、ユーザ状態の推定モデルと密接に連動させる必要があると考えられる。

また、人の認知や感情の偏りを助長しないために、

「評価されている感」を与えずに状態推定を行う設計が不可欠である。

ユーザ自身が“見られている／ジャッジされている”と感じると、対話そのものが防衛的になり、本来のケア的効果が損なわれる可能性がある。

応用領域としては、

個人のセルフケア支援だけでなく、企業のメンタルヘルス窓口や、医療窓口等でのストレスモニタリングにおける“軽量な一次的サポート、ヒヤリング役”としての活用が考えられる。

特に、過度に深い介入を避けつつ、気づきや安定化を促す構造は、業務内で対話が重要な要素となるケア分野の専門職の負担軽減にも寄与する可能性がある。

本研究は、AIの共感的対話における中道的・倫理的なデザイン原則の可能性を示すものであり、仏教における「対機説法」や、ケア実践における「距離の調整」との関連性も示唆され、人とAIが“適度な距離”で支え合う新しいケアモデルの土台になり得る。

今後は、実装検証・ユーザフィードバック・長期運用の評価を通じてより精緻なモデル化を進めたい。