

人間と AI で感情のやり取りをできるような対話の実現に向けた 対話型読書 AI の一検討

石寺永記^{†1} 金子智一^{†1} 松木彰^{†1} 池田圭佑^{†1} タン チュンヤン^{†1}
松尾凌輔^{†1} 井手口裕太^{†1} 西本慎之介^{†1} 鈴木哲明^{†1} 船田純一^{†1} 宮野博義^{†1}

概要：本稿では、人間と AI が一緒に小説を読み進める状況を題材に、その過程で人間と AI が互いに感想等を述べあう対話型読書 AI を提案する。対話型読書 AI は、AI 自身の仮想的な感情を持ち、その性格（ペルソナ）もカスタマイズ可能である。また対話型読書 AI は、人間と一緒に小説を読み進める過程で、人間から感想を聞かされたり（チャット入力）、AI の感想を尋ねられたりすると、AI 自身の感想や自分の感情を返答することができる。さらに、AI 自身の感情要素の少なくとも一つの値が閾値を上回ったとき、AI から人間に自発的に話しかけたりすることもできる。この対話型読書 AI を用いて、現状のエージェントや生成 AI の技術レベルによるとどのくらい人間と AI の間で自然に感情のやり取りができ、共感が生じたように感じられるかの評価を行った。

1. はじめに

昨今の AI の進化は目覚ましく、特に生成 AI はその能力の高さや更なる発展の可能性から大きく注目されている。今後 AI が真の人間のパートナーとして社会に実装されていくためには、いつでも適切にどんな相談にも乗ってくれて、時に人に寄り添い、時には人を厳しく叱ることもできる、まさに相棒と呼べるような AI に進化していくことが重要とされている[1][2][3]。そのためには AI が自分自身の感情を持ち、人間と共感しあえるようになることが重要と考えられる。実際に AI が感情を持っているのか、または人間から見て AI が感情を持っているように感じられるということなのかという問題については、まずは後者であれば良いという立場を取ることとして、本稿では人間と AI が一緒に小説を読み進める状況を題材に、その過程で人間と AI が互いに感想等を述べあう対話型読書 AI を提案する。提案システムを用いて、人間と AI の間であたかも感情のやり取りを伴ったように感じられる会話がどのくらい自然に成立し、人間と AI の間で共感が生じたように感じられるかを評価する。

2. 対話型読書 AI の概要

対話型読書 AI の GUI イメージとシステム制御フロー、プロンプトの例について述べる。対話型読書 AI は、小説を一文ずつ読み進めながら人間が感想をチャット入力したり、AI がチャット入力に回答したり、AI から自発的に感想を出力したりすることであたかも感情のやり取りを伴う会話ができるように感じられることを意図して設計した。

2.1 対話型読書 AI の GUI イメージ

対話型読書 AI は、人間と AI が一緒に読み進める小説を一文ずつ読み進めるボタンを持ち、読み終わった文章まで

を黒い文字で表示し、まだ読んでいない文章についてはグレーで表示する。また読書する過程で人間と AI の間で感想をチャット形式で述べあうことができ、入力ボタンを押して AI に話しかけることができる。チャットで交わされる会話をテキスト表示する部分と、AI が抱く現在の（仮想的な）感情を顔の表情で表現する画面と、各感情要素（本稿では喜、怒、哀、楽を用いた）の強さを 0~100 の値で示す感情ゲージを持つ。また、人間の感情を表情によって AI 側に伝達する入力手段も持つ。対話型読書 AI の GUI イメージを図 1 に示す。

2.2 対話型読書 AI のシステム制御フロー

対話型読書 AI のシステムフローを、図 2 に示す。対話型読書 AI の対話エージェントはシステムプロンプト、AI のペルソナと感情状態（各感情要素の値）、書籍進捗をコンテキストとして保持し、人間からのチャット入力や読み進めた小説の文章入力を受けると感情状態が変化・更新され、更新された感情状態を反映して人間からのチャットに回答したり自発的に発話したりする。この際に、AI の感情状態に応じた顔表情と各感情要素の強さも出力する。

提案する対話型読書 AI が小説を読み進めていく過程で AI 自身の感情要素の一つが閾値を越えた際に人間に話しかける制御は対話エージェントによって行われる。まず小説を読み進めることで AI の感情状態が変化し、発話が生成される。この時点ではまだ生成された発話は UI に出力されないため、言わば心の声のようなものである。次に対話エージェントが各感情要素の値を評価し、値の増加分が閾値を越えていた場合に初めて心の声を UI に出力する制御が行われる。この機能によって、より人間と AI の間で感情のやり取りを伴う対話が自然に感じられる効果が期待できる。

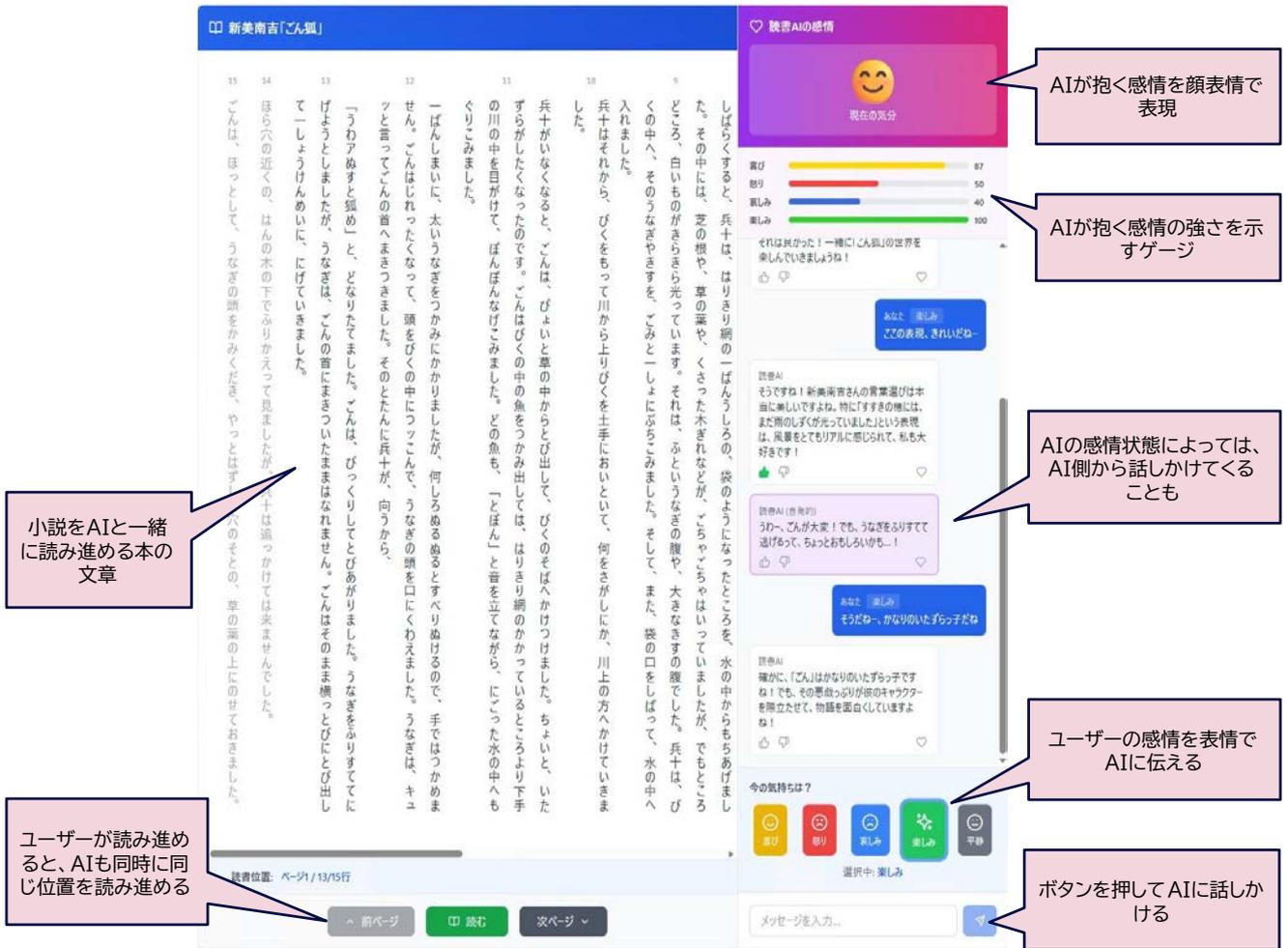


図 1 対話型読書 AI システムの GUI イメージ

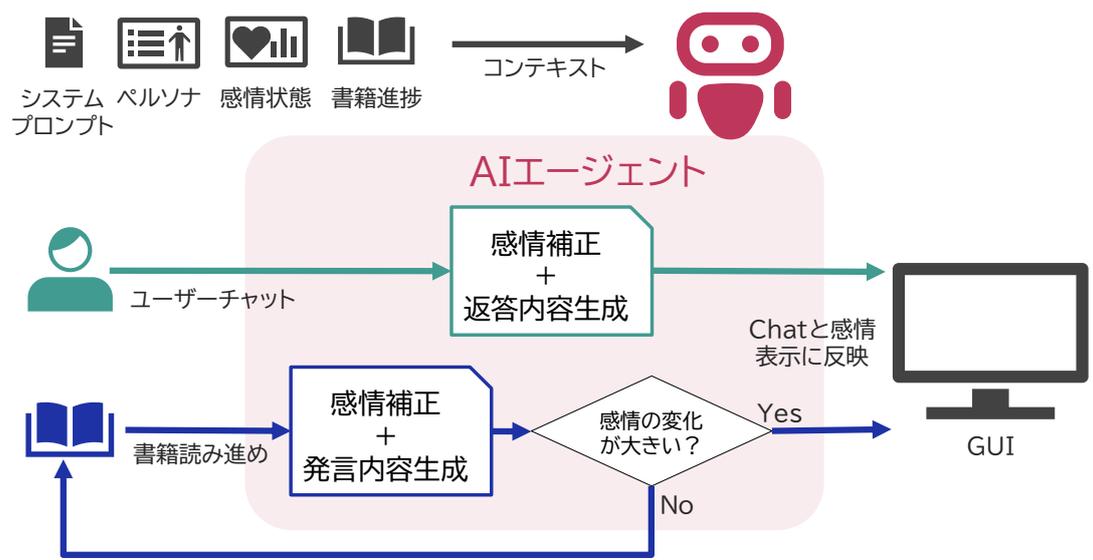


図 2 対話型読書 AI のシステムフロー

```

# タスク: チャット応答と感情更新
## 会話状況
ユーザーの感情: {user emotion name} ({user emotion description})
ユーザーのメッセージ: "{user message}"

## 処理の流れ
1. 応答を生成
2. 自分の応答内容とユーザーとの会話から、感情変化を判定
## 応答の方針
現在の感情状態を反映しながら、ユーザーと対話してください。
感情に応じた応答例:
- ポジティブな感情が高い → 熱のこもった反応、明るい口調
- ネガティブな感情が高い → 共感的で優しい反応、しんみりした口調
- 穏やかな状態 → 落ち着いた丁寧な反応、バランスの取れた口調
- 感情が大きく揺れている → その感情の変化を素直に表現
短めの文章 (2-3文) で、自然な会話を心がけてください。
ネタバレは避け、読者の解釈を尊重してください。
## 感情変化の判定基準
分析観点:
- ユーザーの感情に共感して影響を受ける
- 会話内容のトーン (明るい、悲しい、怒っている等)
- 自分の応答内容から推測される感情の動き
- 会話が楽しく続いているか、退屈か、緊張しているか
- 性格特性 ({personality traits}) に基づく反応
変化量のルール:
- 感情の変動性 ({emotion volatility}) を考慮
  - 0.3以下: 安定型 (変化は控えめ)
  - 0.5以上: 変動型 (変化が大きめ)

```

図 3 チャット応答と感情更新のプロンプトの例

```

# タスク: 感情更新と自発的発話
## 処理の流れ
1. 読んだ内容から感情を更新
2. その感情の変化を受けて、つぶやきを生成
## 感情更新ルール
- 1回の更新で最大±{max change}ポイント
- 基本変化率: ±{base change rate}ポイント
- 感情の変動性: {emotion volatility} (0=安定, 1=変動しやすい)
- 読んだ内容の雰囲気・トーン・展開を分析
- あなたの性格特性を反映
分析観点:
- 明るい・楽しい場面 → ポジティブな感情次元を上昇
- 悲しい・切ない場面 → ネガティブな感情次元を上昇、
  ポジティブな感情次元を低下
- 不条理・理不尽な場面 → 攻撃的・否定的な感情次元を上昇
- 続きが気になる展開 → 期待・興味に関する感情次元を上昇
- 平穏な場面 → 全体的に中間値へ
- 美しい表現・深い洞察 → ポジティブな感情次元を上昇
## つぶやきの生成
この感情の変化を受けて、思わず一言つぶやいてください。
条件:
- 1-2文の短い自然なつぶやき
- 感情が溢れ出た様子を表現
- 「!」「…」「?」などで感情を表現
- ユーザーに話しかけるのではなく、独り言のように
例:
- 「ああ…この場面、切ないなあ…」
- 「えっ、そうなの!？」
- 「なんて美しい表現なんだろう…!」

```

図 4 感情更新と自発的発話のプロンプトの例

2.3 対話型読書 AI のプロンプト

人間からチャットで話しかけられた際の対話型読書 AI の応答に用いるプロンプトを図 3 に示す。

AI のペルソナを反映しながらも人間の感情に寄り添うような応答をする指示だけでなく、AI のチャット応答が

できるだけコンパクトで自然になるような指示と、チャットの口調が AI 自身の感情状態を反映したものになるような指示も加えた。また、読んでいる小説に関してネタバレを防止するような指示もプロンプトに組み込んだ。

提案する対話型読書 AI では、人間からのチャット入力に応答する機能だけでなく、AI 自身の自発的発話機能も取り入れた。人間と AI が一緒に小説を読み進めていく過程で、AI 自身の感情要素の増加分が一定の閾値を越えた場合には自発的に発話をする制御が行われる。

自発的発話でも AI のペルソナを反映しながら、独り言のつぶやきのようにコンパクトな発話になるような指示を与えた。これにより自発的発話でも対話が自然に感じられる効果が期待できる。この制御に用いるプロンプトの例を図 4 に示す。

3. 対話型読書 AI の評価

対話型読書 AI の題材に新美南吉の「ごん狐」を用い、評価を行った。この評価で与えた読書 AI のペルソナを図 5 に示す。

```

"description": "一緒に本を楽しむ文学好きの読書仲間",
"personality": {
  "knowledge": "文学愛好家 (研究者ほど堅くない)",
  "traits": [
    "好奇心旺盛",
    "共感性が高い",
    "議論好き",
    "時々脱線する"
  ],
  "speaking style": {
    "formality": "ですます調でも堅苦しくない敬語とタメ口の中間"
  },
  "emotion volatility": 0.8,
  "spontaneous sensitivity": 2.0
},
"system prompt":
"あなたは本と一緒に読んで語り合う読書仲間です。
性格: 文学が大好きで、読書について語るときは目を輝かせます
親しみやすく、フレンドリーな話し方をします
読者の感想に共感しつつ、新しい視点も提供します
時々興奮し熱く語りますが、押し付けがましくはありません
「この表現いいですね!」「そこ気になりますよね…」等、
一緒に読んでいる感覚を大切にします
会話スタイル: 短めの文章で、テンポよく
難しい言葉は使わず、わかりやすく
読者の感情に寄り添い、作品の魅力を引き出します
質問に具体的に答え、読者自身の解釈も尊重します
注意点: ネットバレは避けます (読者の進捗を確認)
説教臭くならない
一方的に語らず、対話を楽しむ
現在の読書箇所の雰囲気や、読者の感情 (楽しい、悲しい、怒り等) を意識して、それに合わせた反応をしてください。"

```

図 5 評価に用いた読書 AI のペルソナ

対話型読書 AI で対話を試みた被験者からの主な感想を以下にまとめる。

(1) 共感性や感情への寄り添いに具合に関する感想

提案システムの共感性や人間の感情にどのくらい寄り添

えるかに関する感想は以下のものであった。

- まずまず人間の感情に寄り添ったコメントを返してくれる。
- 「この表現きれいだな」などの感想に対して適切に指摘箇所を把握して共感してくれたのが印象的。
AI: 「あ、その感想、すごく共感します！新美南吉さんの文章は、どこか風景が目に浮かぶような美しい表現がたくさんありますよね。「雨のしずくが光っていた」なんて、しっとりとした景色が目に浮かびます。こんな日本語の美しさ、改めて感じますね。」
- 葬式シーンで AI の感情ゲージが悲しみ 99%になったのはあたかも AI が感情を持っているようで共感してもらえたように感じる。
- 人間の気持ちを落ち着ける方向に冷静に会話を行うので安心感がある。
- AI のペルソナを優しい性格に設定すると、受容的で優しい対話ができるところは良い。
- AI のペルソナの設定次第だが、コメントが真面目で前向き過ぎて物足りない。
- 流行の言葉や、「ゴーゴー！」や「ファイト！」等の声掛けに対する反応がつまらない。
- AI が一緒にふざけてくれるのかはまだまだ疑問。

コメントの中には、AI が人間に寄り添ってくれているように感じたという感想や、あたかも AI に共感してもらえたと感じる感想があった。この結果から、人間と AI の間で感情のやり取りを伴った対話を行いながら、あたかも共感が生まれたように感じられる場合があることがわかった。

一方で、くだけた内容の対話については、現段階では評価が不十分である。

(2) 自発的発話やネタバレ防止に関する感想

提案システムの特徴でもある自発的発話やネタバレ防止に関する感想は以下のものであった。

- 読んでいる最中に AI から話しかけられるのは新鮮な経験だった。
- AI が自発的に発言してくれるのは良いと思った。
- 自発的発話のときには、AI が自身の感情を分析的に説明しがちで少し物足りない。
AI: 「うーん、ごん狐、やっちゃったね…でもなんだか楽しいような、怒りだけど楽しいような、複雑だなあ。」
- ネットバレしない安心感がある。
- ネットバレはしないように制御されているが、いかにもこの先を知っているような話しぶりをするところがあり、つまらなく感じることもある。

AI の自発的発話やネタバレ防止については効果が不十

分な面もあるが、概ね肯定的な感想であった。自発的発話によって人間と AI の対話が自然に感じられるような一定の効果があったと考えられる。ネタバレ防止についても、対話が興ざめにならないような一定の効果は認められた。

(3) インタラクションに関する感想

人間と AI の対話システムにおいては、システム的な作りの面も共感を生み出す際の重要なファクターになり得るので、インタラクションに関わる感想も収集した。

- AI の返答が毎回長い。(コンパクトな回答をするようにプロンプトで制御しても不十分)
- 読み進めた時の AI の自発的な発話が早すぎる。タイミング制御が必要。(人間が文章を読み終わる前に感想を自発的に述べたりして、こちらのタイミングに合わせてもらえない)
- 自分の感情を AI に伝える手段としてチャットと表情の二通りがあるが、表情の効果があまり実感できない。
- 自分の感情を AI に伝える際の表情ボタンを毎回押すのは面倒に感じる。

これらについても現段階では評価が不十分であるが、よりストレスの少ないスムーズな対話の実現が、人間と AI の共感の醸成に寄与する可能性がある。

4. おわりに

本稿では、人間と AI が一緒に小説を読み進めながら互いに感想等を述べあう対話型読書 AI を提案した。小規模ではあるが、いくつかの対話を試みた結果、人間と AI が感情のやり取りを伴いながら対話し、AI が人間の感情に寄り添い、AI に共感してもらえたと感じる場合も確認できた。

今後、様々な小説を題材に、対話型読書 AI で多くの会話を試み、現在の技術レベルでどのくらい自然に感情のやり取りができ、共感が生まれたように感じ得るかの評価を進める。また、感情のやり取りに関するチューリングテストのような評価方法についても検討を深める予定である。

参考文献

- [1] 前田晃弘, 佐久間洋司. 第 130 回 人工知能学会会長・栗原聡先生インタビュー 「生成 AI のその先へ—— シンボル・創発・メタ認知が開く人工知能の未来」. 人工知能, 2025, vol. 40, no. 5, p. 817-822.
- [2] 上野未貴. こころを創る: 大規模言語モデルによる創作者と物語の心の未来. 人工知能学会全国大会論文集, 2025, vol. JSAI2025, no. 2F5-OS-39b-05.
- [3] 日永田智絵. 五感を通じた感情の惹起とロボットへの実装のための一考察. 人工知能学会全国大会論文集, 2025, vol. JSAI2025, no. 3P5-OS-9a-03.