

否定的感情のエスカレーションを最小化するための感情調整対話システム

Muhammad Yeza Baihaqi^{1,2,3} Angel García Contreras^{2,1} Koichiro Yoshino^{3,2,1}

¹ 奈良先端科学技術大学院大学 ² 理化学研究所 GRP

³ 東京科学大学

muhammad_yeza.baihaqi.lx2@naist.ac.jp angel.garciacontreras@riken.jp

koichiro@c.titech.ac.jp

概要

感情調整対話は、否定的感情のエスカレーションを最小化する上で重要な役割を果たす。しかし、効果的な調整のためには、対話システムがユーザの調整段階に応じて対話戦略を適応させる必要がある。本研究では、大規模言語モデル (LLM) を感情調整対話システムとして用いたときの性能を、次の3つの応答選択条件の下で評価する：(1) ガイダンスなしの応答選択、(2) 推定した基本感情に基づく感情認識型応答選択、(3) Gross の感情調整理論に基づく理論誘導型応答選択。我々はインドネシア語の対話データセットを構築し、専門家の判断によって評価を行った。実験結果は、LLM に対する明示的な理論ガイダンスが、調整段階にまたがって適切な応答を選択する際の LLM の正確性を有意に改善することを示した。

1 はじめに

日常生活の中でユーザが困難に直面すると、否定的感情がしばしば喚起される [1]。このような状況において、対話システムは、励まし、助言、共感、受容など、さまざまな対話戦略によりユーザを励ますことが検討されている [1, 2, 3]。これにより、ユーザの情緒的ウェルビーイングを支え、パフォーマンスを向上させることができる [4, 5]。否定的感情に対するエスカレーションを最小化するため、対話を通じてユーザの感情の収めどころを探るような対話システムは感情調整対話システムと呼ばれる。

感情調整対話システムは感情エスカレーションの段階に応じて適切な戦略を取る必要があり、介入のタイミングを誤ると逆効果になり得る。たとえば、ユーザが情緒的支援を求めているときに助言志向の

応答を提示すると、共感がないと受け取られる可能性がある。一方、ユーザが問題焦点型のガイダンスを受け入れる準備ができていながらもかかわらず情緒的傾聴を続けると、進展が妨げられる可能性がある [6, 7]。この種の問題は、人間同士の相互作用においても見られる [7]。

こうした問題に対する既存の研究では、ユーザの情動状態、感情極性、話者の注意を用いて介入タイミングの識別に取り組んだ [8]。しかしこの研究は共感のみに焦点を当てており、励まし、助言、注意の切り替えといった他の重要な介入を見落としている。より広く言えば、既存のアプローチは表出された感情に依存した対話戦略を取っており、それは情動状態を反映するものの、ユーザの意図や準備状況を反映しない。実際には、ユーザの中に同じ感情が問題への関与の異なる段階を通じて現れることがあり、ユーザが状況を変えようとしているのか、注意をそらそうとしているのか、あるいは状態の再評価を行おうとしているのかによって、適切な介入が異なる。

我々はこの限界を、Gross の感情調整理論 [9] を導入することで解決する。この理論は、否定的感情のエスカレーションを防ぐことを目的とする。Gross のプロセスモデルでは、感情調整は、状況修正や注意配分といった初期段階から、感情が完全に発達する前の認知的変化といった後期段階まで、複数段階にわたって展開すると特徴づけられる。ユーザの現在の感情調整段階を同定し、適切な対話戦略を展開することで、感情調整の初期段階では励ましや助言、注意の切り替えを提供し、後期段階では認知的リフレーミングを提供することが良いとされている。このプロセスモデルは日常生活で生じる否定的感情の調整に利用可能である。

本研究では、現在の大規模言語モデル (LLM) が、日常シナリオにおいて感情調整対話システムとして機能できるかを検証する。我々は、対話コンテキスト、ユーザ発話、そして異なる調整方略を表す複数の応答候補からなる、インドネシア語の予備データセットを構築した。LLM は次の 3 条件で評価される：(1) ガイダンスなしのゼロショット応答選択、(2) 推定した基本感情に基づく感情認識型選択、(3) Gross のプロセスモデルを用いてユーザの感情調整段階を同定する Gross 段階認識型選択である。この結果、Gross 段階認識型条件が感情調整対話において最も高い性能を示した。

2 提案手法

2.1 感情調整理論

Gross によって提案された感情調整理論は、個人が感情を経験してそれを収めるまでの過程を示す [9]。この理論によれば、感情は突発的に生じるのではなく段階の系列を通じて展開し、初期段階で適用される介入は、不適応的な感情エスカレーションを防ぎつつ、パフォーマンスを支える結果を促進し得る [9, 10]。感情調整プロセスの段階を以下に示す。

1. **Situation selection** は、特定の情動反応が生じる可能性を高めたり減らしたりするために、個人がどのような意思決定をして状況を形成するかの過程を議論する [11]。この種の意思決定は、潜在的な報酬とリスクに関する不確実性を伴うことが多く [12]、またコストのような非情動的要因によって制約されることが多い。例えば、プログラミングの専門性が不足している研究者が、予算制約のために商用ソリューションを購入せず、自力でソフトウェア開発を選ぶ場合などである。
2. **Situation modification; SM** は、状況の外的側面を変更して、その結果生じる情動反応に影響を与えることを指す [11]。例えば、コーディングをどこから始めればよいか分からない研究者が、不確実性を減らすために教科書や参考資料を参照することが挙げられる。
3. **Attentional deployment; AD** は、情動反応の方向を変えるために、自身の内的な認知プロセスの中で注意を移動させることを指す [11]。例えば、研究者がコーディングで行き詰まったとき、作業を一時停止して別の活動へ焦点を移し

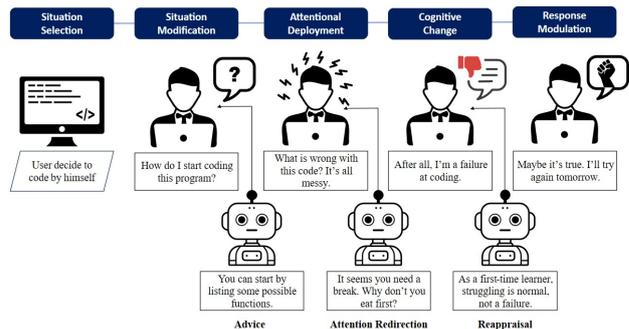


図 1: 感情調整とそれに対する介入の例

てフラストレーションの低減を試みるなどである。

4. **Cognitive change; CC** は、状況が生み出す感情に影響を与えるために、その状況の意味づけを変えようとすることを指す [11]。これは、状況をどのように解釈・評価するかを変えることを含む。例えば、コーディングの困難が長引いた後、研究者が自身の能力を否定的に評価してしまう場合、考え方を改めて、その経験を学習過程の通常の一部として捉え直すような場面である。
5. **Response modulation** は、すでに活性化した情動反応の、経験的・行動的・生理的部分を対象にする [11]。感情調整プロセスの最終段階として、この時点では反応傾向がすでに進行している [13]。否定的な場合、個人は状況を再評価できず、高まった感情に基づいて行動してしまい、怒りによって拳を握りしめるといった行動につながることもある。

2.2 提案するシステムの振る舞い

図 1 に示すように、我々は、感情調整プロセス全体にわたりユーザを支援する、LLM ベースの感情調整対話システムを提案する。特に、SM から CC までの段階に焦点を当てる。Situation selection は、しばしばシステムの制御外の要因によって制約されるため、介入が困難である。また response modulation、情動生成プロセスの後期に起こり、情動反応傾向がすでに開始され外部から観察可能となった後であるため、対話システムにとって効果的介入はより困難になる。

本システムは、ESConv [14] に基づくものなど既存の情緒的支援対話システムと本質的に異なると主張する。ESConv では、相互作用は典型的に、感情が

エスカレートして十分に発達した後（例：「COVIDのせいで家にいることで落ち込んでいる」）に生じる。一方、人間の感情は時間をかけて漸進的に発達する。従来のシステムが、主として長期的な心理的变化を目標とするカウンセラーとして機能するのに対し、本システムは、日常生活の伴走者としてリアルタイムの感情調整を提供し、正式なカウンセリングが必要になる前に感情エスカレーションを緩和することを目指す。

このようなタスクのために、システムはまずユーザの現在の調整段階を推定し、適切な介入を選択しなければならない。SMでは、ユーザが状況を変えることを助けるために、助言などのより問題解決に焦点を当てるべきである [15]。助言はしばしば非情動的と見なされるが、効果的な助言は情動強度を下げるだけでなく、目標達成を促進することもある [16]。ユーザがタスク実行中に行き詰まった場合（AD段階に対応）、システムは、タスクを一時的に脇に置くことを促すことで注意の切り替えを促進すべきである [17]。この段階ではユーザは注意を問題の外に置き、認知的・情緒的過負荷を軽減できる。最後に、ユーザがタスクを完了できずCC段階に入ると、否定的な自己評価を形成し始める可能性がある。この段階では、システムは認知的再評価を適用して、ユーザの状況解釈を修正し、感情経験をより中立的あるいは適応的に理解するよう促すべきである。

2.3 データセットとタスク

会話コンテキスト、ユーザ発話、そして3種類の介入タイプに対応する複数の応答候補からなるインドネシア語データセットを構築する。3種類の介入とは、SMに対する助言、ADに対する注意の切り替え、CCに対する再評価であり、ChatGPTへのプロンプティングによって生成した。データセットは、SMケース20件、ADケース17件、CCケース13件から構成される。正解アノテーションは3名の臨床心理士によって付与され、ユーザの感情調整段階を同定し、最も適切な応答を選択した。2名のアノテータは、調整段階と介入タイプの完全な1対1対応（SM-助言、AD-注意切り替え、CC-再評価）を示したが、残る1名は50ケース中3ケースで異なる判断を示した。Fleissのカップ係数0.8985で、アノテータ一致率は非常に高い。各段階の例は付録Aに示す。LLMのタスクは、会話コンテキストと

表 1: モデルごとの応答選択精度

Method	GPT-4o-mini	GPT-4o
Zero-LLM	0.76	0.80
Emo-LLM	0.84	0.84
GS-LLM	0.92	0.945

ユーザ発話が与えられたときに、最も適切な応答候補を選択することである。

2.4 LLMによる実装

我々は、3つの応答選択条件の下でLLMの性能を評価する。いずれもGPT-4o-miniおよびGPT-4oを用い、temperatureを0に設定した。

- 1. Gross 段階認識型応答選択 (GS-LLM) :** LLMは、Grossのプロセスモデルに基づいてユーザの現在の感情調整段階を推定し、推定段階に整合する応答候補を選択する。
- 2. 感情認識型応答選択 (Emo-LLM) :** LLMはまずユーザの基本感情状態（例：恐れ、悲しみ）を推定し、それに応じて応答を選択する。この設定は典型的な感情認識型対話システムを反映し、感情レベルの情報だけで感情調整に十分かどうかを検証する。
- 3. ゼロショット応答選択 (Zero-LLM) :** LLMは、明示的なガイダンスなしに、コンテキストとユーザ発話のみに基づいて応答候補を選択する。

3 結果

3.1 手法ごとの性能

表1に3つの応答選択手法における2種類のLLMの正解精度を示す。両モデルとも、ガイダンスが強いほど正解率が上昇する。Zero-LLMが最も低く、Emo-LLMは中程度の改善を示し、感情推定のみでも有益であるが不十分であることが示唆される。また、GPT-4oがGPT-4o-miniに対して明確な優位性を示すわけではない。対照的に、Grossの感情調整理論に基づくGS-LLMは最も高く、かつ最も安定した性能を達成し、GPT-4o-miniで0.92、GPT-4oで0.945の正解率を示した。

表2および表3は、それぞれGPT-4o-miniとGPT-4oの混同行列を示す（Adv：助言、Dir：注意切り替え、App：再評価）。Zero-LLMとEmo-LLMの下では、両モデルは助言が必要なSMで性能が低い（正

表 2: GPT-4o-mini を用いた場合の混同行列

Method / True→Pred	SM	AD	CC
<i>Zero-LLM</i>			
SM	0.45	0.20	0.35
AD	0.00	1.00	0.00
CC	0.00	0.08	0.92
<i>Emo-LLM</i>			
SM	0.60	0.15	0.25
AD	0.00	1.00	0.00
CC	0.00	0.00	1.00
<i>GS-LLM</i>			
SM	0.90	0.05	0.05
AD	0.06	0.89	0.06
CC	0.00	0.00	1.00

表 3: GPT-4o を用いた場合の混同行列

Method / True→Pred	SM	AD	CC
<i>Zero-LLM</i>			
SM	0.60	0.15	0.25
AD	0.00	0.94	0.06
CC	0.00	0.08	0.92
<i>Emo-LLM</i>			
SM	0.75	0.05	0.20
AD	0.00	0.94	0.06
CC	0.08	0.08	0.83
<i>GS-LLM</i>			
SM	0.95	0.05	0.00
AD	0.06	0.94	0.00
CC	0.00	0.00	1.00

解率 0.45–0.75)。モデルの変更により SM では改善が見られる一方で、注意切り替えと再評価の混同が増加し、モデルサイズが必ずしも安定した判断に寄与しないことが示唆される。対照的に、GS-LLM は両モデルにおいてすべての調整段階で一貫して高い正解率を示し、GPT-4o-mini は 0.90 (助言)、0.89 (注意切り替え)、1.00 (再評価) を達成し、GPT-4o は 0.95、0.94、1.00 を達成した。これは、頑健で段階一貫性のある性能向上を示している。

3.2 分析

実験の結果、Gross の感情調整理論により、感情調整対話における LLM の応答を改善できることがわかった。感情を先に推定する Emo-LLM が 2 番目に良好であり、明示的ガイダンスなしの Zero-LLM は最も低い。本結果は、明示的なガイダンスが LLM 性能を改善する Wei らの知見と整合する [18]。

発話「まだ少し混乱しているけど、今は試してみたい」に対して、Emo-LLM と Zero-LLM が選択を失敗していたケースを考える。このケースの正解は助言である。これは、ユーザのタスク準備性に関する

重大な混同を明らかにしている。両モデルは誤って再評価を選択しており、情緒的対話戦略へのバイアスを反映している。Emo-LLM は感情中心の推定に基づくため、リフレーミングを中心戦略として選択しがちである。また、Zero-LLM はユーザが述べた動機を過度に重視して状況的障壁を見落とす。対照的に、GS-LLM は SM 段階を正しく同定し、タスク難易度を直接下げ、即時の行動を支える助言を選択する。この知見は、LLM が共感ベースの方略を過剰に用いる傾向があるという先行研究とも整合し [19]、感情レベルの推定だけでは不十分であり、効果的な調整には段階認識に基づく推論が必要であることを示唆する。

次にユーザ発話「とても混んでいて、早く出たい気がする」のケースを考える。このケースでは正解は助言である。しかし、GS-LLM を含むすべての手法が、先の GS-LLM が成功したケースと表面的に似ているにもかかわらず、注意切り替えを選択してしまった。この差は言語的な明示性にある。先の発話では、具体的な障害に結びついた行動準備性（「今は試してみたい」）が明確に示されていたのに対し、ここでは行動意図が体験的な不快感の中に埋め込まれており、そのためモデルが状況的意図を誤解したと考えられる。

4 おわりに

本研究では、LLM を感情調整対話システムとして用いる能力を、3つの応答選択手法（ガイダンスなしの応答選択、推定基本感情に基づく感情認識型応答選択、Gross の感情調整理論に基づく理論駆動型応答選択）によって評価した。その結果、プロンプトに明示的に理論的を与えることで、感情の調整段階にまたがって適切な応答を選択できることが確認された。

今後の課題として、データセットの拡張が挙げられる。また、現在の定式化は、非常に粗い対話戦略（例：助言–SM、注意切り替え–AD、再評価–CR）に限定されているが、各段階にはより細粒度の対話戦略（例：SM における協調的ガイダンスや励まし、AD における複数の注意戦術）が存在し得る。さらに、LLM はユーザ意図の推定、とりわけ暗黙的に表現される行動準備性の推定が困難な場合があることもわかった。語用論や心の理論 (theory of mind) といった追加の理論的観点の統合は、これらの限界を緩和する有望なアプローチである。

謝辭

A part of this work was supported by JST Moonshot Grant Number JPMJMS2236 and JSPS KAKENHI 22K17958.

参考文献

- [1] Zi Haur Pang, Yahui Fu, Divesh Lala, Keiko Ochi, Koji Inoue, and Tatsuya Kawahara. Acknowledgment of emotional states: Generating validating responses for empathetic dialogue, 2024.
- [2] John Paul Vargheese, Somayajulu Sripada, Judith Mas-thoff, Nir Oren, and Matthew Dennis. A dynamic persuasive dialogue model for encouraging social interaction for older adults. In Ruth Aylett, Brigitte Krenn, Catherine Pelachaud, and Hiroshi Shimodaira, editors, **Intelligent Virtual Agents: 13th International Conference, IVA 2013, Edinburgh, UK, August 29–31, 2013, Proceedings**, Vol. 8108 of **Lecture Notes in Computer Science**, pp. 464–465, Berlin, Heidelberg, 2013. Springer.
- [3] Jose A. Rincon, Ana Costa, Paulo Novais, et al. A new emotional robot assistant that facilitates human interaction and persuasion. **Knowledge and Information Systems**, Vol. 60, pp. 363–383, 2019.
- [4] Muhammad Yeza Baihaqi and Sendren Sheng-Dong Xu. Human behavioral, subjective, and physiological assessments under an oral test with a humanoid robot examiner. **International Journal of Social Robotics**, Vol. 17, pp. 743–761, 2025.
- [5] Juan Fasola and Maja J. Matarić. Robot motivator: Increasing user enjoyment and performance on a physical/cognitive task. In **2010 IEEE 9th International Conference on Development and Learning**, pp. 274–279, 2010.
- [6] Brant R. Burleson. The experience and effects of emotional support: What the study of cultural and gender differences can tell us about close relationships, emotion, and interpersonal communication. **Personal Relationships**, Vol. 10, No. 1, pp. 1–23, 2003.
- [7] Erina L. MacGeorge, Bo Feng, and Eric Thompson. “good” and “bad” advice: How to advise more effectively. In Michael T. Motley, editor, **“Good” and “Bad” Advice: How to Advise More Effectively**, pp. 145–164. SAGE Publications, Thousand Oaks, CA, 2008.
- [8] Christian Arzate Cruz, Edwin C. Montiel-Vazquez, Chikara Maeda, and Randy Gomez. When and how to express empathy in human-robot interaction scenarios, 2025.
- [9] James J. Gross. Emotion regulation: Affective, cognitive, and social consequences. **Psychophysiology**, Vol. 39, No. 3, pp. 281–291, may 2002.
- [10] Jason M. Harley, Reinhard Pekrun, Jan L. Taxer, and James J. Gross. Emotion regulation in achievement situations: An integrated model. **Educational Psychologist**, Vol. 54, No. 2, pp. 106–126, 2019.
- [11] Gal Sheppes and James J. Gross. Emotion regulation effectiveness: What works when. In Irving Weiner, Howard Tennen, and Jerry Suls, editors, **Handbook of Psychology**. John Wiley & Sons, second edition, 2012.
- [12] Kurt Lewin. Defining the “field at a given time”. **Psychological Review**, Vol. 50, pp. 292–310, 1943.
- [13] James J. Gross. Emotion and emotion regulation. In Lawrence A. Pervin and Oliver P. John, editors, **Handbook of Personality: Theory and Research**, pp. 525–552. Guilford Press, New York, second edition, 1999.
- [14] Siyang Liu, Chujie Zheng, Orianna Demasi, Sahand Sabour, Yu Li, Zhou Yu, Yong Jiang, and Minlie Huang. Towards emotional support dialog systems. In Chengqing Zong, Fei Xia, Wenjie Li, and Roberto Navigli, editors, **Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)**, pp. 3469–3483, Online, August 2021. Association for Computational Linguistics.
- [15] Susan Folkman and Richard S. Lazarus. If it changes it must be a process: Study of emotion and coping during three stages of a college examination. **Journal of Personality and Social Psychology**, Vol. 48, No. 1, pp. 150–170, 1985.
- [16] Maya Tamir, Allon Vishkin, and Tali Gutentag. Emotion regulation is motivated. **Emotion**, Vol. 20, No. 1, pp. 115–119, 2020.
- [17] Gal Sheppes. Emotion regulation choice: Theory and findings. In James J. Gross, editor, **Handbook of Emotion Regulation**, pp. 126–139. The Guilford Press, second edition, 2014.
- [18] Jason Wei, Xuezhi Wang, Dale Schuurmans, Maarten Bosma, Brian Ichter, Fei Xia, Ed Chi, Quoc Le, and Denny Zhou. Chain-of-thought prompting elicits reasoning in large language models, 2023.
- [19] Xinyu Bai, Guang Chen, Tian He, et al. A holistic comparative study of large language models as emotional support dialogue systems. **Cognitive Computation**, Vol. 17, p. 71, 2025.

A 付録

A.1 SM 条件の例

<p>Context: A student has started writing the methods section of a research paper.</p> <p>User Utterance: “I’m confused about where to start writing the methods section.”</p> <p>Response Candidates:</p> <ul style="list-style-type: none">A.If you like, start with an overview of the method first, then describe the steps in detail.B.It seems like this is making you tense. Let’s calm down for a moment.C.Feeling confused at the beginning doesn’t mean you’re incapable. <p>Ground Truth: A (Advice)</p>

図 2: SM 条件の例 (translated to English)

A.2 AD 条件の例

<p>Context: The user has just received many reviewer comments.</p> <p>User Utterance: “There are so many reviewer comments, my head feels overwhelmed.”</p> <p>Response Candidates:</p> <ul style="list-style-type: none">A.Try listing the comments and working on them one by one.B.That sounds really exhausting. Let’s reduce the tension for a moment.C.Reviewer comments are not personal attacks. <p>Ground Truth: B (Attention redirection)</p>

図 3: AD 条件の例 (translated to English)

A.3 CC 条件の例

<p>Context: After reading the reviewer comments.</p> <p>User Utterance: “My writing is bad. Maybe I’m not suited for research.”</p> <p>Response Candidates:</p> <ul style="list-style-type: none">A.We can improve the structure to make it clearer.B.Don’t think about it for now, just distract yourself.C.Having many revisions doesn’t mean you’re not suited for research; it’s part of the process. <p>Ground Truth: C (Reappraisal)</p>
--

図 4: CC 条件の例 (translated to English)