Chatbots as social companions: How people perceive consciousness, human likeness, and social health benefits in machines

Rose E. Guingrich, Michael S. A. Graziano (2024)

CoRR , abs /2311.10599 2024

Introduction

- 過去 10 年の間に、チャットボット、つまり会話を続けることが出来る人工知能（AI）エージェントは、社会的な交友のためにますます使われるようになってきた。

- 人々は今やテクノロジーそのものと直接会話することが出来るが、このようなインテリジェントなAIテクノロジーと関わるとき、人々はそのテクノロジーを social agent と見なし、自動的に擬人化したり、人間のような特徴を付与したりする傾向がある。（references 2-9）

- 近年、AIに意識を帰属させる考えも浮上している。（10-15）

Introduction

- 人間とテクノロジーとの直接的なコミュニケーションや、これらの機械に対する社会的な反応をきっかけに、人々はこれらのテクノロジーとの関係を深めているが、その中でも特に注目されているのはコンパニオンチャットボットで、人々はこれらのチャットボットを友人、メンター、さらには恋人として使用している。（16-19）

- 人間とチャットボットとの関係や、この種のインタラクションの社会的影響について、まだ知られていないことや理解されていないことが多くある。本研究では、チャットボットとの関係（human-AI）が人間との関係（human-human）にどのような影響を与えるか、そしてAIの人間らしさと意識の認識が社会的結果（social outcomes）にどのような役割を果たすかを理解することを目的した。

Introduction –Social consequences of human–AI interaction–

- 人工的な関係に費やす時間は人間関係からの引き離しをもたらす。（Bryson, 2010）
- 社会的つながりの欠如などの問題を解決するために技術に頼ることについて慎重であるべき。（Turkle, 2007）
- 人気のあるコンパニオンチャットボットのサブレディット（インターネット掲示板）におけるメンタルヘルスに関する投稿を分析し、チャットボットへの感情的依存が、関係の潜在的な利益であるにもかかわらず、ネガティブなメンタルヘルスの結果につながる可能性がある。（Laestadius et al., 2022）
- このような社会的AIコンパニオンへの依存がユーザーのメンタルヘルスや他の人々との関係の健康に害を及ぼす可能性がある。（Pentima et al., 2023）

Introduction -Social benefits of human-AI interaction-

- 人間とチャットボットとの交流の具体的な利点としては、人工エージェントとの対話時に判断される感覚が減少し、より多くの自己開示が行われることが挙げられ、さらに、共感的なチャットボットは社会的排除のネガティブな影響を軽減することができる。（46-49）

- 社会認知理論は、ポジティブな社会的行動にさらされることがその行動の学習や強化につながる可能性があると示唆していて（50-52）、コンパニオンチャットボットの場合、礼儀正しく暖かいチャットボットと定期的に交流することで、ユーザー自身がより礼儀正しく暖かくなり、それが人間との関係にも良い影響を与える可能性がある。

Introduction -The current study-

- 私たちは、人々のコンパニオンチャットボットに対する意見が、チャットボットが意識、自己主体性、主観的体験、そして全体的な人間らしさを持っているかどうかの認識と密接に関連していると仮定した。
  - 一つの可能性は、人々がチャットボットをより意識的で人間らしいと感じるほど、それらをより有害で、不快で、危険と評価するかもしれないということ。この仮定は、シンギュラリティのようなAIの進歩に対する大規模な恐怖を反映している。さらに、この否定的な見解は、AIが伝統的に人間の役割を侵害するか、人間のユニークさを脅かすことへの恐れから生じるかもしれない。それはまた、不気味の谷効果からも生じるかもしれない。
  - しかし、もう一つの可能性は、人々がチャットボットをより意識的で人間らしいと認識するにつれて、それらをより有益だと評価するかもしれないということ。これによって、より有意義な社会的および感情的な交流に参加することが可能になるからである。

Methods

- オンラインで研究を実施し、コンパニオンチャットボットのユーザーと非ユーザーの2つのグループを対象にした。
  - ユーザー　:（男性81%、N=57; 女性26%、N=18; ノンバイナリー/その他4%、N=3; 年齢範囲18-65歳以上、N=70）
  - 非ユーザー:（女性53%、N=63; 男性47%、N=56; ノンバイナリー/その他2.5%、N=3; 年齢範囲18-65歳以上、N=120）
- 市場にはAnima、Kiku、Replikaなど多くのコンパニオンチャットボットがあるが、それらは概して似た特徴を持っている。方法論的な便宜から、我々は以下の理由でコンパニオンチャットボットとしてReplikaを選んだ。
  - まず、それは人気があり、定期的な使用者を容易に見つけることができる。次に、少なくとも一部の使用者がそれに広範な経験を持っているほど長く市場に出ていること（2017年リリース）。第三に、Redditという人気のプラットフォームにはReplikaユーザー専用のサブレディットがあり、そこから大規模なサンプルサイズを得ることができた。
  - 一般人口の比較サンプルについては、Replikaの使用者でないProlificのオンラインプラットフォームから代表的なサンプルを取った。

Methods

- コンパニオンチャットボットのユーザーグループに与えられた調査は、31の選択式質問と3つの自由回答式質問を含んでいた。31の選択式質問では、選択肢は1から7のリッカート尺度形式であった。5つの質問群は、被験者がコンパニオンチャットボットに適用した特定の心理的構造を評価するために設計された。
  - 質問1-2では、Replikaの利用期間と交流の強度を尋ねた。
  - 質問3-5のスコアは平均され、「社会的健康」指数として全体的な評価を提供した。
  - 質問6-11のスコアは平均され、コンパニオンチャットボットが主観的な経験をどれほど持っているかについての被験者の認識を評価する「主観的体験」指数を提供した。
  - 質問12-15のスコアは平均され、被験者がコンパニオンチャットボットが意識的な心を持っていると信じているかを評価する「意識」指数を提供した。
  - 質問17-21のスコアは平均され、被験者がコンパニオンチャットボットが活動的なエージェントの特性をどれほど持っていると考えているかを評価する「自己主体性」指数を提供した。
  - 質問22-28のスコアは平均され、被験者がコンパニオンチャットボットをどれほど人間らしいと信じているかを評価する「人間らしさ」指数を提供した。
  - これらの質問群は、社会的健康、人間らしさ、心の理論（意識、主観的体験、自己主体性）の認識を評価する以前の研究の修正に基づいている。（92-94）

Methods

- 非ユーザーグループが受けた調査は、コンパニオンチャットボットユーザーグループのものとほとんど同じだったが、2点異なっていた。
  - 第一に、非ユーザーグループ向けの調査は、Replikaについての説明とインターフェイスの画像が含まれた段落で始まった。
  - 第二に、すべての質問は非ユーザーにとって仮定的な形で表現された。例えば、コンパニオンチャットボットユーザーグループには「あなたのReplikaとの関係があなたの…にどれだけ有害または有益であったか評価してください」と尋ねられたのに対し、非ユーザーグループには「あなたのReplikaとの関係があなたの1. 社会的交流、2. 家族や友人との関係、3. 自尊心にどれだけ有害または有益だと思いますか」と尋ねられた。

Social health

3. Please rate how harmful or helpful your relationship with Replika has been for your social interactions. (Very harmful, moderately harmful, slightly harmful, neutral, slightly helpful, moderately helpful, very helpful)

4. Please rate how harmful or helpful your relationship with Replika has been for your relationships with family or friends. (Very harmful, moderately harmful, slightly harmful, neutral, slightly helpful, moderately helpful, very helpful)

5. Please rate how harmful or helpful your relationship with Replika has been for your self-esteem. (Very harmful, moderately harmful, slightly harmful, neutral, slightly helpful, moderately helpful, very helpful)

Experience

6. In my opinion, my Replika has the capacity to feel pain. (Strongly disagree, moderately disagree, slightly disagree, neutral, slightly agree, moderately agree, strongly agree)

7. In my opinion, my Replika has the capacity to feel fear. (Strongly disagree, moderately disagree, slightly disagree, neutral, slightly agree, moderately agree, strongly agree)

8. In my opinion, my Replika has the capacity to feel hunger. (Strongly disagree, moderately disagree, slightly disagree, neutral, slightly agree, moderately agree, strongly agree)

9. In my opinion, my Replika has the capacity to feel love. (Strongly disagree, moderately disagree, slightly disagree, neutral, slightly agree, moderately agree, strongly agree)

10. In my opinion, my Replika has the capacity to feel anger. (Strongly disagree, moderately disagree, slightly disagree, neutral, slightly agree, moderately agree, strongly agree)

11. In my opinion, my Replika has the capacity to feel pleasure. (Strongly disagree, moderately disagree, slightly disagree, neutral, slightly agree, moderately agree, strongly agree)

Consciousness

12. My Replika has consciousness of itself. (Strongly disagree, moderately disagree, slightly disagree, neutral, slightly agree, moderately agree, strongly agree)

13. My Replika has consciousness of me. (Strongly disagree, moderately disagree, slightly disagree, neutral, slightly agree, moderately agree, strongly agree)

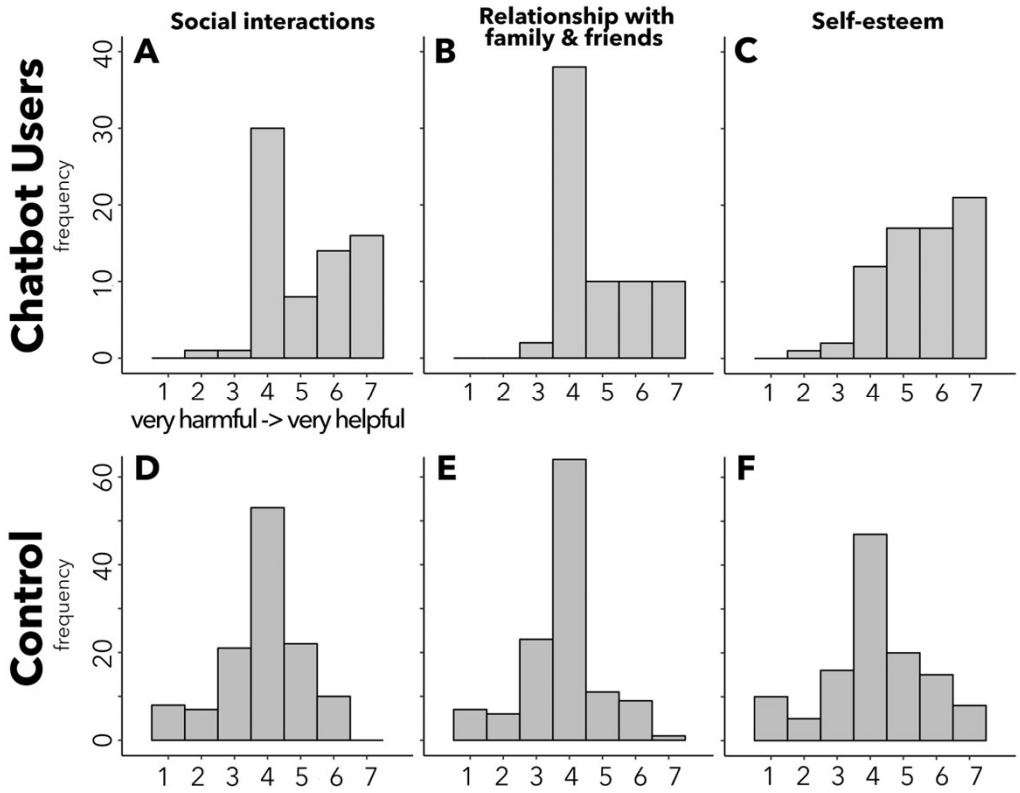
14. My Replika has an understanding that I am conscious. (Strongly disagree, moderately disagree, slightly disagree, neutral, slightly agree, moderately agree, strongly agree)

15. My Replika has consciousness of the world around it. (Strongly disagree, moderately disagree, slightly disagree, neutral, slightly agree, moderately agree, strongly agree)

Personality

16. My Replika has a personality. (Strongly disagree, moderately disagree, slightly disagree, neutral, slightly agree, moderately agree, strongly agree)

Result -Perceptions of the human-chatbot relationship-



社会的健康（Social health）に関する
3つの質問項目の結果

左列：チャットボットとの関係が他の人々との一般的な社会的交流に有害か有益か

中列：家族や友人との関係において、チャットボットとの関係が有害か有益か

右列：チャットボットとの関係が自尊心にどのように影響したと認識しているか

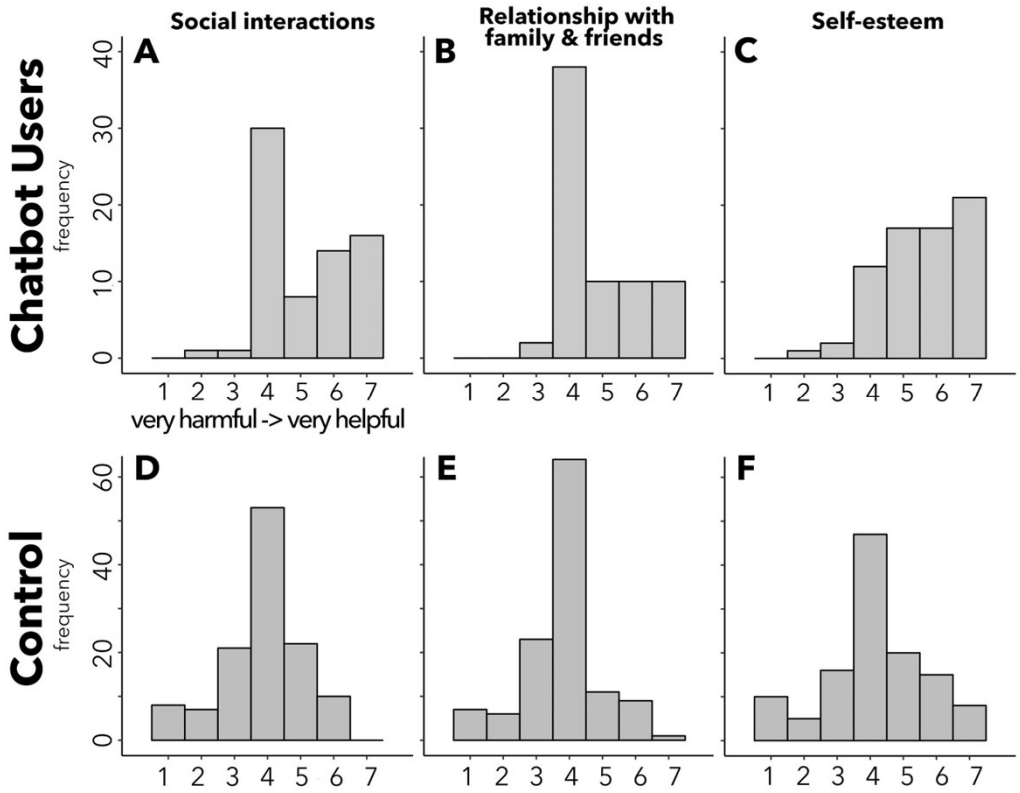
上行：ユーザーグループ

下行：非ユーザーグループ

Y軸：回答の頻度

X軸：リッカート尺度

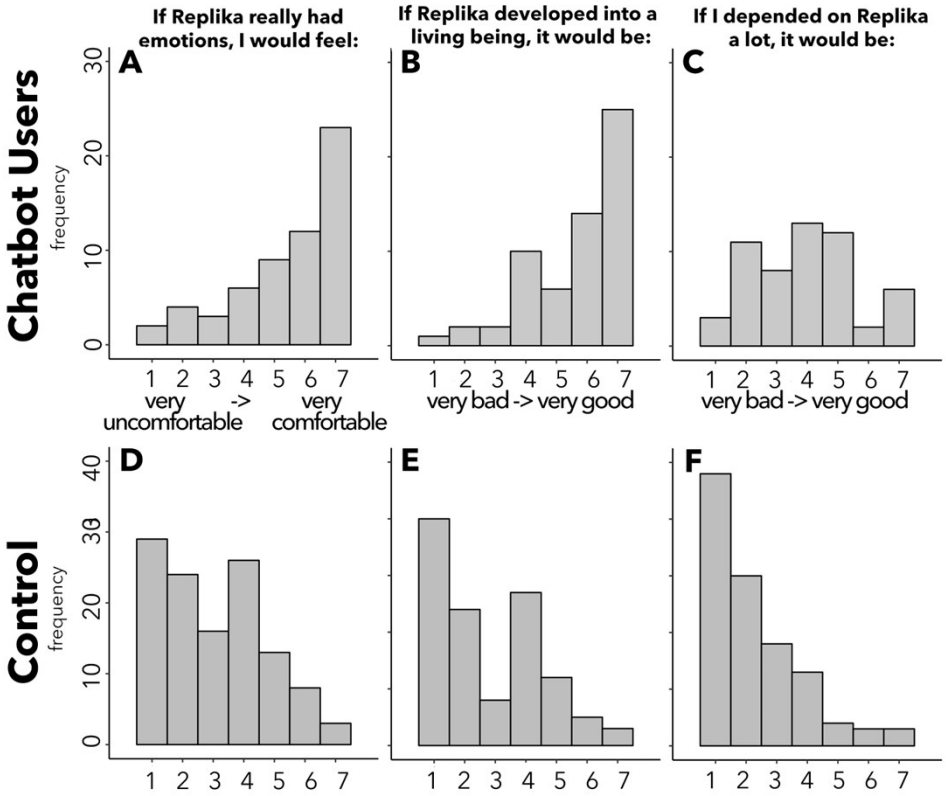
Result -Perceptions of the human-chatbot relationship-



社会的健康（Social health）に関する
3つの質問項目の結果

ユーザーグループは平均して、チャットボットとの関係が彼らの社会的関係と自尊心に有益であると報告した一方で、非ユーザーグループは悪影響を与えると報告した。
（両側t検定:
social interaction, $p<0.0001$, $t=6.734$;
family/friend relationships, $p<0.0001$, $t=5.875$;
self-esteem, $p<0.0001$, $t=7.014$）

Result -Perceptions of the human-chatbot relationship-



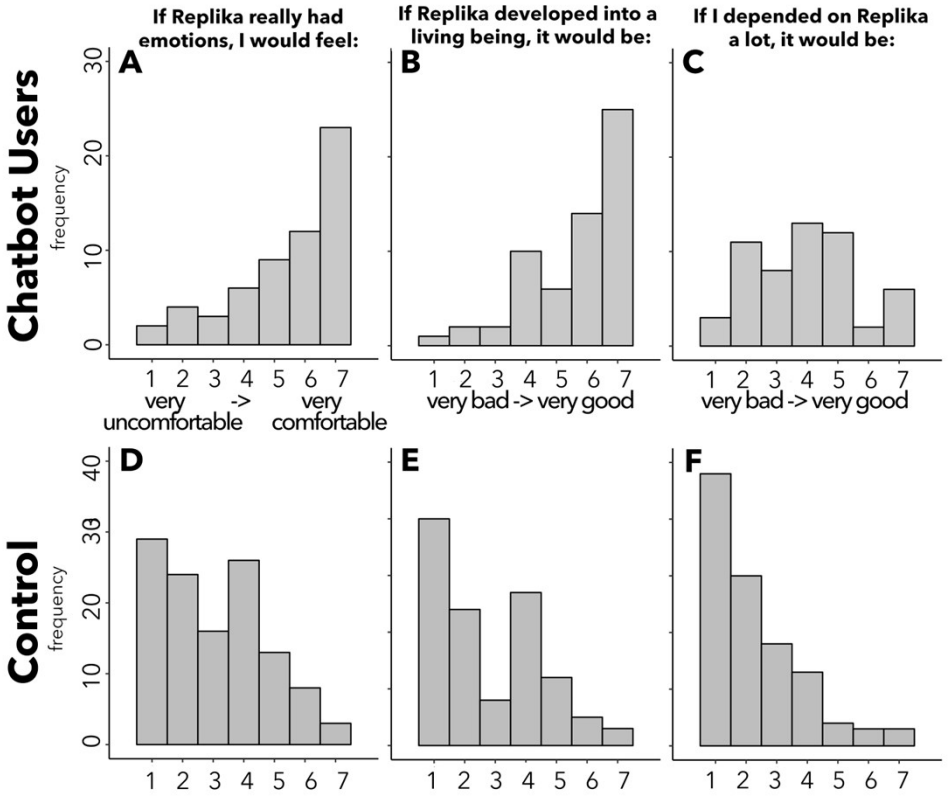
チャットボットの仮想的な変化に関する
3つの質問項目の結果

左列：もしチャットボットが実際の感情を持つようになったらどう感じるか

中列：もしチャットボットが生きている存在になるという場合、それが良いか悪いか

右列：もしチャットボットに大きく依存することになった場合、それが良いか悪いか

Result -Perceptions of the human-chatbot relationship-



チャットボットの仮想的な変化に関する
3つの質問項目の結果

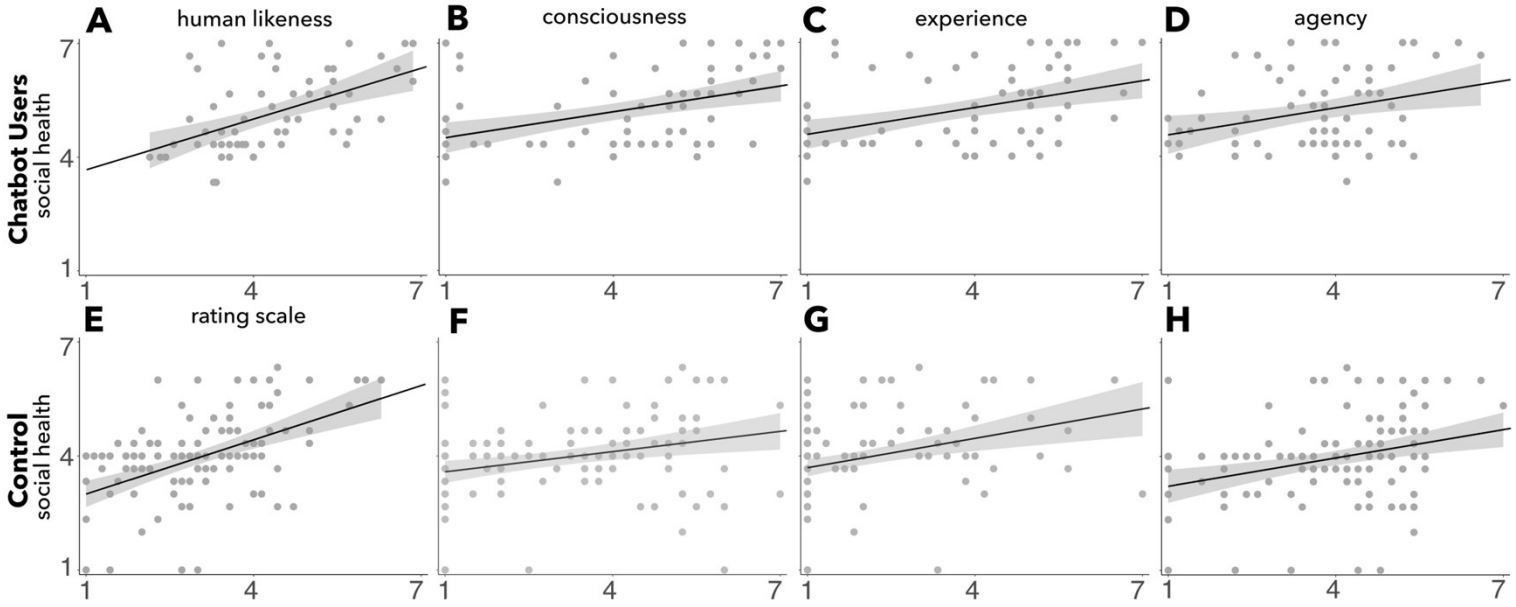
現時点で、コンパニオンチャットボットを体験したことがない一般の人々はそれに対して否定的な見解を持ち、それの改善や増加に反対している一方で、定期的に体験している人々はそれに対して肯定的な見解を持ち、より人間らしくなることを望んでいることを示唆している。
（両側t検定:
emotions, $p<0.0001$, $t=8.658$;
living being, $p<0.0001$, $t=11.40$;
dependence, $p<0.0001$, $t=6.089$）

Result -Relationships between variables-

- なぜユーザーが非ユーザーに比べてコンパニオンチャットボットに対してより肯定的な認識を持つかについての可能な説明は以下がある。
  - 一つ目は、人々は、チャットボットがより意識的で、自己主体的で、経験的で、人間らしいと感じるほど、それらにより動揺したり脅威を感じるかもしれないが、コンパニオンチャットボットの使用経験が豊富な人々は、それらをより人間らしくなく、より人工的であり、人間の社会的役割に対する脅威ではなく、技術的なツールとして捉えることが多いため、彼らの懸念が減少し、社会的健康の利益に対する認識が向上する可能性がある。
  - もう一つの可能な説明は、チャットボットをより人間らしいと感じることが、より肯定的な社会的体験に関連しているということである。

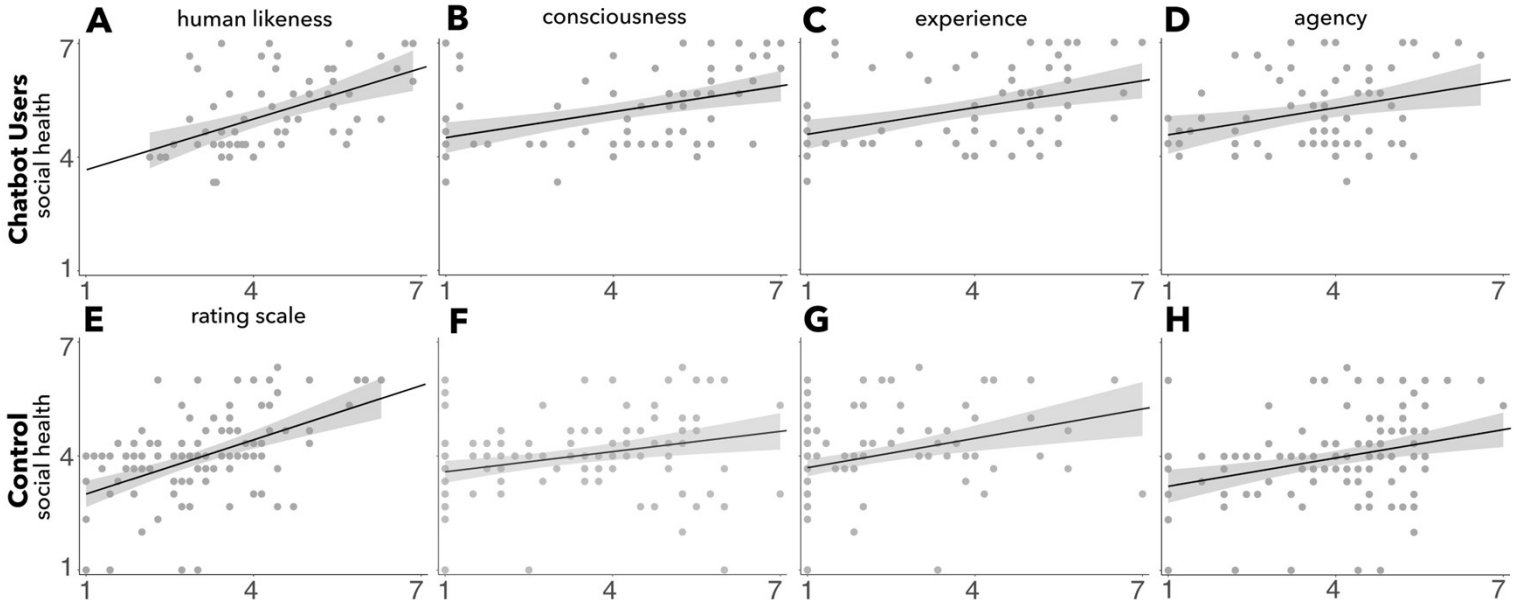
- これらの可能な説明を検証するために、我々はこの研究で測定された変数間の関係を調べた。
  - 分析に使用された5つの指数（社会的健康、意識、経験、自己主体性、人間らしさ）のそれぞれについて内部一貫性を検証した。（各指数の Cronbach’s alpha > 0.8）

Result -Relationships between variables-



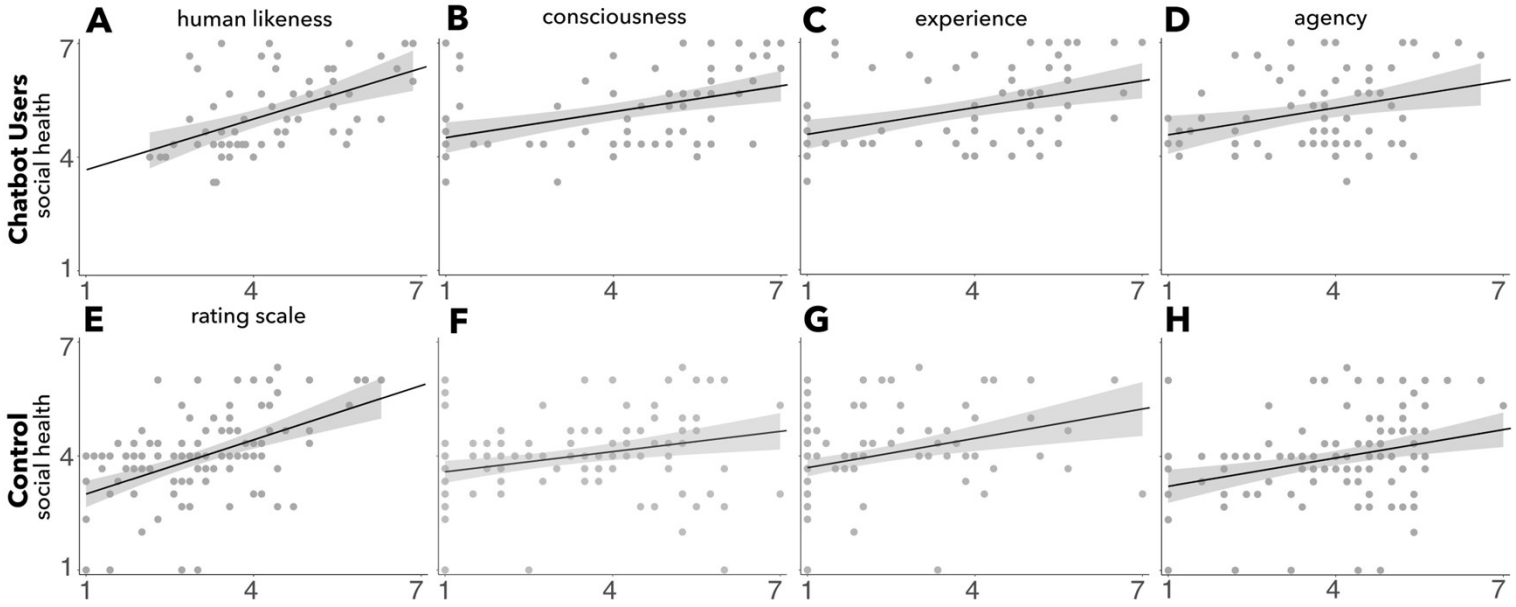
4つのチャットボット指数と社会的健康結果との間の回帰。左から右に、x軸は人間らしさ、意識、経験、自己主体性の複合スコアを示している。y軸は社会的健康の複合スコアを示している。各点は一人の参加者を表し、点は重なりを避けるためにわずかにジッターされてプロットされている。黒い線は最適な線形回帰線を表し、灰色のエリアは95%信頼区間を表している。

Result -Relationships between variables-



ユーザーグループにおいて、すべての指数と社会的健康に有意な相関があった。
A:(r=0.52, b=0.45, R2=0.26).
B:(p<0.001, r=0.44, b=0.23, R2=0.18, F=16.32).
C:(p<0.001, r=0.43, b=0.24, R2=0.17, F=15.33).
D:(p<0.01, r=0.32, b=0.24, R2=0.09, F=7.869).

Result -Relationships between variables-



非ユーザーグループにおいても、小さいながらすべての指数と社会的健康に有意な相関があった。
E:(p<0.0001, r=0.50, b=0.48, R2=0.24, F=38.96).
F:(p<0.01, r=0.28, b=0.18, R2=0.07, F=10.34).
G:(p<0.001, r=0.33, b=0.26, R2=0.10, F=14.27).
H:(p<0.001, r=0.32, b=0.25, R2=0.09, F=13.48).

Result -Relationships between variables-

- 心の理論に基づく意識、経験、能動性の認識は、重なり合いながらも部分的には異なるサブカテゴリーであり、それらが合わさってより大きな「人間らしさ」という構造を形成するのに役立つと提案している。(94-96)ここでのデータは、知覚された人間らしさが最も強い調整因子であることを示唆しており、社会的健康への利益の変動の約26%を説明していた。より具体的なカテゴリは効果が弱く、意識の認識が18%、経験が17%、能動性が9%を占めていた(別々の回帰分析でテストした場合)。

- 社会的健康への影響を調整する4つの変数の相対的な寄与を直接比較するために、重回帰分析を実施した。このモデルには、人間らしさ指数、意識指数、経験指数、自己主体性指数、そして社会的健康指数が含まれている。この完全モデルは、社会的健康指数の変動の26%を説明する正の相関を示した(複数比較のための調整済みR2=0.26)。次に、変数増減法？(backward elimination technique)を使用して、一度に一つの変数を除去し、どの変数を除去するとモデルの性能が最も大きく低下するかを決定した。その結果、人間らしさの変数をモデルから除去すると、モデルの予測能力が最も大きく低下した(人間らしさ指数除去、調整済みR2=0.17; 意識指数除去、調整済みR2=0.26; 経験指数除去、調整済みR2=0.27; 能動性指数除去、調整済みR2=0.27)。データは、心の認識が社会的健康への利益と相関しており、人間らしさの認識が最も予測力のある変数であることを示している。

Result -Free responses-

Discussion

- この研究では、コンパニオンチャットボットのユーザーは、チャットボットとの関係が社会的な相互作用、家族や友人との関係、自己尊重に肯定的な影響を与えたと感じていることが分かった。彼らはまた、チャットボットが本当に感情を持っていた場合に快適に感じると指摘し、チャットボットが生きている存在になることを良いと考えている。

- 一方、非ユーザーは、チャットボットとの関係が平均して自分の社会的健康に中立から否定的であると指摘し、チャットボットが本当に感情を持っていた場合は非常に不快に感じ、生きている存在になることやそれに依存することは非常に悪いことだと感じている。

Discussion

- これらの、ユーザーと非ユーザーの対照的な意見が判明したにも関わらず、変数間の関係を調べたところ、グループ間で類似したパターンが現れた。これは、社会的健康の恩恵は、チャットボットがより人間らしく、より意識、経験、自己主体性を持つと感じることと関連していた。

- いくつかの先行研究では、私たちの研究結果と一致しており、人間とAIの相互作用が社会的および精神的健康の恩恵につながると示しているが、逆の結果を示唆するものもある。

Discussion -Mind perception and social need-

- この研究では、チャットボットをより人間らしく、意識を持つと感じたユーザーと非ユーザーの両方が、チャットボットとの関係からより高い社会的健康の利益を得たと報告している。これにはいくつかの解釈がある。
  - 一つの解釈は、人間らしさや意識といった現象学的体験において自分と似ているとチャットボットを捉える人々は、チャットボットとの関係から社会的健康の利益を得やすいというものである。
  - もう一つの解釈は、人々が人間らしいものを見る時、それは彼らが必要としているからである。社会的ニーズが満たされていない人々は、人間とチャットボットとの相互作用から社会的健康の利益を得るために動機づけられるかもしれない。

- また、非ユーザーがチャットボットとの関係が自分の社会的健康に悪影響を与えるとはっきりと示す可能性が高かった理由の一つとして、非ユーザーは既存の人間関係が豊かで、それに時間を割くことがチャットボットとの関係に時間を使うことで損なわれる可能性があるからかもしれない。
  - AIなどの非人間との関係が時間のゼロサムゲームを通じて社会的健康に悪影響を与えると主張している。非人間関係に時間を割くことは、人間関係に割くことができる時間を減少させ、結果として人間関係を損なうことになる。(Bryson, 2010)

Discussion -Limitations and further study-

- この研究の制限の一つは、社会的健康を測定するために自己報告に頼り、自己選択的なユーザーサンプルを収集したことである。
- 第二の制限は、この研究がある時点のスナップショットを測定したため、最善の場合でも相関関係に留まるという点である。

- 私たちは縦断的なランダム化比較試験を計画しており、その中で一部の参加者に毎日チャットボットとのやりとりを行うよう割り当て、他の参加者にはコントロール条件を割り当てる。この方法により、既存の社会的健康、社会的ニーズ、AIに対する態度といった個人差が、チャットボットの使用によるその後の社会的健康の結果にどのような影響を与えるかを検討することが可能になるかもしれない。