

機械は人間にカウンセリングできるか？

——擬人観と人工知能——

加 地 雄 一

1. はじめに

スマートフォンやスマートスピーカーに搭載されている人工知能（AI）アシスタント（Siri, Alexa, Google アシスタント）に話しかけるとニュースや天気を教えてくれる。今のところ、これら AI アシスタントに心理カウンセリングを提供する機能は備わっていない。もちろん科学技術の進歩によって将来的に（それがいつになるかは議論の余地があるが）AI アシスタントが心理カウンセリングをおこなうことは可能になるだろう。本論文では、テキスト入力型の機械カウンセラーである ELIZA（イライザ）と Woebot（ウォウボット）の事例を紹介し、機械が人間をカウンセリングする際の問題点を検討する。

2. ELIZA

音声入力ではなくテキスト入力であれば、機械が人間をカウンセリングした事例は既にある。ELIZA である。Apple 社製品搭載の AI アシスタント Siri に「Hey Siri イライザとの関係は」と音声入力すると「ELIZA をご存じですか？ 彼女は私の最初の先生だったんですよ！」と返ってくる（図 1）。もちろんこれは、実際に ELIZA が Siri の最初の先生というわけではなく、あくまでそのような設定であるというユーモアだ。



図1 Siriが答えるELIZAとの関係

ELIZAとはWeizenbaum(1966)が開発した、人間と英語で会話のできるプログラムである(Weizenbaum, 1976/1979)。ELIZAという名前はアカデミー作品賞を受賞した映画『マイ・フェア・レディ』の原作『ピグマリオン』(ジョージ・バーナード・ショーによる戯曲)の登場人物からとられた。貧困層特有のイライザ(女性)の言葉遣いを言語学者の教授が富裕層の言葉遣いに教育するストーリーになぞらえている。Weizenbaum(1976/1979)によれば命名の由来は「教育することによってだんだんうまく“話せる”ようになるから」である。しかし、おそらくこれは、プログラム開発段階の話であって、完成したELIZAのプログラムが人間の教育によって(人間とのやりとりを学習することによって)うまく話せることを意味するわけではない。

ELIZAのプログラムには、ユーザーが入力と思われるキーワードのリストと、それらに対応する返答用のフレーズが、事前に定義(用意)されている(狩野, 2017)。ユーザーが入力したテキストの中に、あらかじめ用意したキーワードがあれば、それに対応する出力(返答)用のフレーズが返答される。キーワードがなければ、ユーザーが入力したフレーズを繰り返す。

返すような返事をしたり、話題を変えたりするような工夫がなされている。

あらゆるすべてのテーマについて対応可能なキーワードやフレーズを事前に用意するのは難しいため、テーマごとに用意されるキーワードやフレーズは変わる。その一つひとつのテーマを台本（スクリプト）という。一番有名な台本は「ドクター（doctor）」であり、この台本で ELIZA はロジャーズ（クライアント中心療法）学派の心理療法士（psychotherapist）を演じる。現在でも Emacs というテキストエディターでプログラムを実行することができる。クライアント中心療法とは、カウンセラーがカウンセリングの方向づけを控えて、クライアントの気持ちの整理や自己理解を助ける心理療法である。

表1に筆者がクライアント役になって ELIZA にカウンセリングを受けた例を示した。筆者はクライアント（CI）役として「電車に乗るとお腹が痛くなる、その理由は仕事がつらいから」という役割を演じた。カウンセリングにあたって、ELIZA との会話が継続するよう努めた。表1には ELIZA との最初のセッションで（やり直しや試行錯誤はしておらず）、デモンストレーション（実際の例）として筆者が区切りが良いと感じたところまでを示した。不自然な点はいくつかあるが、ユーザー側が会話を続けようと努力すれば、1問1答を超えたレベルで会話が成り立つことがわかる。

Th.2で「医師に相談してください」とあり、取りつく島もなくカウンセリングが終了してしまいそうになるが、これは「お腹が痛い」という身体症状がキーワードとなって返答されたフレーズだと思われる。ユーザー側で心理相談として会話を続けようと努力すれば、表1の通りに会話は続く。

Th.7とTh.8における ELIZA の発話は「本当にそう（だと）思いますか？」の繰り返しであった。これは事前に定義されたキーワードがなかった場合の返答であると思われる。CIとしては「はい」としか言えず、同じことしか答えられなかった。ただ、これもユーザー側で表現を変えたり情

表1 ELIZA とのカウンセリングの例

話者	発話内容 (原文 (英文))	和訳
Th.1	I am the psychotherapist. Please, describe your problems. Each time you are finished talking, type RET twice.	私は心理療法士です。あなたの問題を説明してください。話し終わるたびにリターンを2回入力してください。
Cl.1	My stomach hurts when I get on the train.	電車に乗ると、お腹が痛くなるんです。
Th.2	Maybe you should consult a medical doctor; I am a psychotherapist..	医師に相談してください、私は心理療法士です。
Cl.2	I can not go to the office.	会社に行けないんです。
Th.3	Why do you say you can not go to the office?	なぜあなたは会社に行けないと言うのでしょうか？
Cl.3	I do not want to go to the company.	会社に行きたくないからです。
Th.4	What makes you believe you might want not want to go to the company?	会社に行きたくないとおあなたに思わせているものは何でしょうか？
Cl.4	My job is hard.	仕事がつらいことです。
Th.5	Earlier you said you would like not want to go to the company?	あなたは前に会社に行きたくないと言いましたね？
Cl.5	Yes, because the work is painful.	はい、仕事がつらいので。
Th.6	Is it because the work is painful that you came to me?	仕事がつらいから私のところに来たのですか？
Cl.6	Yes.	はい。
Th.7	I see... Well, are you sure that this is so?	なるほど。本当にそう思いますか？
Cl.7	Yes.	はい。
Th.8	Yes, hmmm... so, do you really think this is so?	はい、ううん。本当にそうだと思いますか？
Cl.8	I really think so. I think that I can not go to work because my job is painful.	本当にそう思います。会社に行けないのは仕事がつらいからだと思います。
Th.9	Does it bother you that your job is painful?	仕事がつらいのはあなたを悩ませますか？
Cl.9	Yes, it is very.	はい、とても。
Th.10	Perhaps your life has something to do with this.	おそらくこの問題はあなたの人生と何か関係があります。
Cl.10	It might be true.	確かにそうかもしれません。

注) Th (セラピスト) は ELIZA の出力 (発話), Cl (クライアント) は筆者の入力 (発話) を表す。英文入力と和訳は Google 翻訳を参考に筆者がおこなった。

報を足したりするなど、会話を続ける努力をすれば会話は続く。

Th.4「会社に行きたくないとあなたに思わせているものは何でしょうか？」や Th.10「おそらくこの問題はあなたの人生と何か関係があります。」は、問題の原因を考え、人生の振り返りや自己理解を深めるためのきっかけとなるかもしれない。

Weizenbaum (1976/1979) は、ELIZA に与えた台本は彼女が心理療法士の役割を果たすように作ったと述べている。しかし、同時に、「正確には茶化すというべきだろう」と述べていることにも注意が必要である。Weizenbaum 自身はカウンセリングにはセラピストの感情移入が不可欠な要素で、これを機械に置き換えてしまえるとは考えていなかった。ELIZA の知名度が高くなり、これを利用する人が増えるにつれて、ユーザーが ELIZA を人間と同等のものとして扱おうとすることに Weizenbaum (1976/1979) はショックを受けた。彼は、その例として、自分の秘書をあげている。彼女は ELIZA がプログラムにすぎないことを十分に知っているにもかかわらず、ELIZA とほんの数回やりとりしただけで、Weizenbaum に部屋を出て行ってほしい (ELIZA と二人きりにしてほしい) と頼んだ。これにはもちろん、個人的な秘密を入力しているので、その様子を見られたくない、という意図があると思われるが、Weizenbaum は人間が機械と強い感情的なつながりを持つとうとすることにショックを受けたのである。

ELIZA が決して人間に感情移入することのない機械にすぎないのに、人間の側が感情的な結びつきを持つとうとする現象は、Weizenbaum が当初意図していたこととは異なり、逆説的で注目に値する現象である。この問題を解く鍵として「擬人観 (Anthropomorphism)」がある。擬人観とは、人間でないものに人間らしさを見出す現象のことである。擬人観は人間と機械との共同作業において、人間が機械をパートナーとして受け入れやすくするものとして応用されている (Kuz et al., 2013)。機械の見た目や中身が人間

に近いかどうかとは独立に、擬人観が生じるかどうかを検討することが、機械カウンセラーを設計する一つのアプローチとして有用である。

3. Woebot

ELIZA はあらかじめ定義されたキーワード（データベース）とのパターンマッチングをしているにすぎなかったが、ユーザーとのやりとりから学習し、そのユーザーに合わせた応答をする会話プログラムもある。例えば、Woebot (Fitzpatrick et al., 2017) があげられる。一般に、会話プログラムは人工無能（無脳）と呼ばれることが多いが、それは単にデータベースとのパターンマッチングをしている場合である。人工無能（無脳）とは異なり、ユーザーの入力によって学習する Woebot は人工知能に近いプログラムである。Woebot とのカウンセリングの例を表 2 に示した。

表 2 は Woebot とのやりとりの冒頭部分である。この後に守秘義務や緊急時の対応の説明が続く。表 2 にある通り、「ユーザーが何をしているか」、「その時にどのような気分を感じているか」を Woebot は尋ねてくる。図 2 に実際の Woebot とのやりとりの画面 (iPhone) を示した。Woebot はモバイルオペレーティングシステムである iOS か Android にアプリとしてインストールできる。Facebook の Messenger を使って PC で利用することも可能である。図 2 は Woebot が「今、何をしていますか？ (What are you doing now?)」と尋ねて、筆者が「タイピング (Typing)」と答えた後のやりとりである。その時 (今日) の気分を選択する画面で、絵文字が添えられ、ユーザーが選択しやすいインターフェースになっている。

Woebot は認知行動療法をベースにしたアプリである。Fitzpatrick et al. (2017) は大学のソーシャルメディアで 18 ~ 28 歳の調査協力者を集め、2 週間 Woebot を利用するグループと、抑うつに関する電子図書を読むグループとに調査協力者をランダムに振り分けた。Woebot に参加したグループは

表2 Woebot とのカウンセリングの例

話者	原文	和訳
Th.1	Hello human! Let me introduce myself: I'm Woebot. I'm a robot.	こんにちは人間。自己紹介をさせてください。私は Woebot です。私はロボットです。
Cl.1	Hello Woebot.	こんにちは Woebot.
Th.2	I'll be checking in with you. At the start of each conversation, I'll ask you what you're doing and how you're feeling in that moment.	私はあなたにうまくいっているか尋ねます。各会話のはじめにあなたが何をしているのか、その時にどのように感じているのかを尋ねます。
Cl.2	Ok.	わかりました。
Th.3	The buttons below are the main way you'll talk to me for now, unless I ask you to type. I don't do open ended conversation and I don't do therapy either.	下のボタンは、私があなたに入力を求めない限り、あなたが今のところ私に話す主な手段になります。私はオープンエンド（自由記述的）な会話はしませんし、治療もしません。
Cl.3	Got it.	わかりました。
Th.4	My super smart creators made sure that I adapt to what you say.	私の創造主はあなたが言うことに私が順応することを確かめました。
Cl.4	Gotcha.	わかった。
Th.5	And then depending on your responses, I can show you evidence-based techniques grounded in... SCIENCE!	そしてあなたの反応に応じて、証拠に基づいたテクニックをあなたに示すことができます。
Cl.5	Cool!	すごい。
Th.6	The more you tell me, the more helpful I can be. For example, over time, I might be able to see patterns in your mood.	たくさん話してくれればくれるほど、私はもっと役に立つことができます。たとえば、時間が経つにつれて、私はあなたの気分パターンを見つけることができるかもしれません。
Cl.6	Ok.	わかった。

注) Th (セラピスト) は Woebot の出力 (発話), Cl (クライアント) は筆者の入力 (発話) を表す。和訳は Google 翻訳を参考に筆者がおこなった。英文入力は提案された選択肢 (ボタン) の一番上を選択した。

2 週間で平均 12.14 (標準偏差 2.23) 回, Woebot と会話した。それぞれのグループには調査期間 (2 週間) の事前と事後で気分に関する自己報告式の心理検査をオンラインで受けてもらった。脱落せずに事後テストまで到達

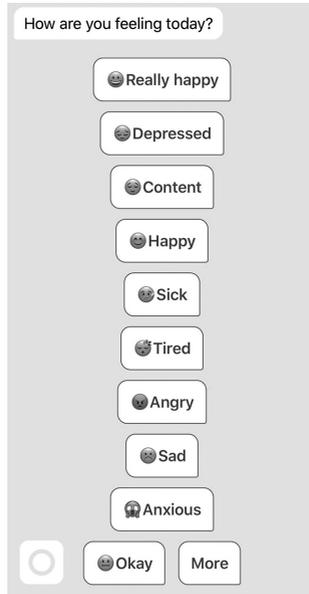


図2 Woebot とのやりとり（左が Woebot, 右がユーザー）

した参加者は Woebot グループが 31 人（当初は 34 人），電子図書グループが 25 人（当初は 36 人）であった。事前と事後で抑うつに関する心理検査の得点が Woebot グループでは有意に低下し，電子図書グループでは有意な差は見られなかった。これらの結果から，Woebot には治療継続効果と抑うつを低減する効果があることが示唆される。ただし，統制グループ（Woebot と比較するグループ）が電子図書を読むグループであり，人間が実施するオンラインカウンセリングと比較した場合の Woebot の効果はわからない。

4. 今後の展望

ELIZA と Woebot という特徴の異なる対照的な機械カウンセラーの事例

を紹介した。どちらにも一長一短がある。ELIZA は自己を振り返るきっかけを与えてくれるが、人間が会話を続ける工夫をしなければならない。機械に愛着を持つか、会話を続ける能力や動機づけがあるユーザーには向いているが、そうでないユーザーには向いていない。人間に近いかどうかとは独立に、擬人観が生じる工夫をすればユーザー層は拡大するだろう。

一方、Woebot は機械が人間に合わせてくれ、入力も簡単である。査読付きの雑誌に掲載されているというエビデンスもある (Fitzpatrick et al., 2017)。しかし、基本的に話題は「今、何をしているか」、「その時の気分」に限られるので、物足りなさを感じるユーザーもいるかもしれない。ELIZA が psychotherapist と名乗るのとは対照的に、Woebot は robot と名乗る。

このように、どちらにも一長一短があるので、ユーザー（人間）側が自分の特性に合わせて機械のタイプを選ぶ必要があるだろう。また、機械カウンセラーを開発するにあたって、「人工無能（無脳）型か、人工知能型か」という問題がある。さらに、それとは必ずしも一致しない軸として「人間が機械に合わせるタイプか（擬人観）、機械が人間に合わせるタイプか（人工知能、機械学習）」という問題もあげられる。もちろん両者のハイブリッド型もありえるだろう。

引用文献

- Fitzpatrick, K. K., Darcy, A., & Vierhile, M. (2017). Delivering Cognitive Behavior Therapy to Young Adults With Symptoms of Depression and Anxiety Using a Fully Automated Conversational Agent (Woebot): A Randomized Controlled Trial. *JMIR Mental Health*, 4 (2), e19. doi: 10.2196/mental.7785.
- 狩野芳伸 (2017). コンピューターに話が通じるか：対話システムの現在 情報管理, 59 (10), 658-665.
- Kuz, S., Mayer, M. P., Müller, S., & Schlick, C. M. (2013). (他) Using Anthropomorphism to Improve the Human-Machine Interaction in Industrial Environments (Part I). In Duffy V. G. (eds) *Digital Human Modeling and Applications in Health, Safety, Ergonomics, and Risk Management. Human Body Modeling and Ergonomics. DHM 2013. Lecture Notes in Computer Science*, vol 8026. Springer, Berlin, Heidelberg

- Weizenbaum, J. (1966). ELIZA—a computer program for the study of natural language communication between man and machine. *Communications of the ACM*, 9 (1), 36-45.
- Weizenbaum, J. (1976). *Computer power and human reason: from judgement to calculation*. San Francisco: W. H. Freeman. (ワイゼンバウム, J. 秋葉忠利 (訳) (1979). コンピューター・パワー：人工知能と人間の理性 サイマル出版会)