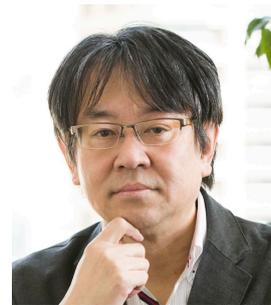


HAIヒューマンエージェントインタラクション： 人間-AI協働のための情報デザイン

大学共同利用機関法人 情報・システム研究機構
国立情報学研究所／総合研究大学院大学

教授 山田 誠二



1. オフィス・家庭に浸透していくAI

2021年現在、人工知能AIの第3次ブーム[1]は、社会導入の充実期を迎えている。今回のAIブームは始まって既に10年近く経つが、いまだに連日「AI導入」「AIサービス」に関する情報がメディアで伝えられている。このことは、AIの社会導入、社会浸透が十分迅速に行われておらず、かつまだ大きなブルーオーシャンが残っていることを意味する。既に、次のバズワードであるDX (Digital Transformation)、ヒューマンセントリック、デジタルツインなどが氾濫するなか、AIを超える、しっかりしたアルゴリズム基盤[2]をもつITテクノロジーは見当たらない。

このような状況で、今後AIの社会浸透を飛躍的に加速させるには、いくつかの課題が認識されている[1]。その1つが、『これまでAIを使ったこともない一般ユーザーがAIを使うときに生じる戸惑い』である。この一般ユーザーがAIに対して最初に感じる戸惑い、違和感をいかに取り除いて、人間とAIの協働を実現するかの方法論を開発する新しい工学が、本稿で議論するHAIヒューマンエージェントインタラクションである。

2. エージェントとは何か

では、一体本稿で扱う「エージェント (agent)」とは、何だろうか。エージェントと聞くと、映画にもなったような球団と年俵の交渉をするプロ野球選手のエージェントが思い浮かぶかもしれないが、もちろんここでのエージェントはそれではない。ここで言うエージェントとは、『私たちが「人間っぽく」感じてしまう人工物』を意味する。換言すると、「擬人化された人工物」と言ってもいいだろう。この定義は、従来のAIにおけるエージェントの定義「環境の情報を獲得して、それらを基に自分で判断し行動する」という自律性を重視したものとは趣が異なっている。このような定義になったのは、「自律性」が定義・判定しにくいことも理由であるが、かと言って「人間っぽさ」が定義しやすいのでもない。

社会心理学の方でも、『メディアの等式』[3]と呼ばれる、

人間が人工物を人や生物のように扱ってしまう現象は観測され、研究されている。このエージェントの定義は、メディアの等式に沿ったものだと言えるだろう。

ただし、対象を人間っぽく感じるのはあくまで人間自身、つまり主観であり、その意味で主観の解釈に大きく依存する定義となっており、工学にはなじみにくいものがある。

さて、メディアの等式が成り立つ人工物は、非常に多くのバリエーションがあるように、ここでいうエージェントにも多くの種類があるが、HAIで扱うエージェントは、以下の3つに大別される。

1. 擬人化 (仮想) エージェント
2. ロボット
3. 人間

擬人化エージェントは、モニタ上に現れるCGで実装された仮想エージェントであり、チャットのアバターやバーチャルYouTuberなどが典型例である。ロボットは、説明するまでもないが、擬人化エージェントとは異なり、物理的の身体を持っている特徴をもつエージェントである。そして、最後に「人間」自身も重要なエージェントである。厳密には、人間は人工物ではないので、定義から外れるがお許しいただきたい。

3. HAI: Human-Agent Interaction ヒューマンエージェントインタラクション

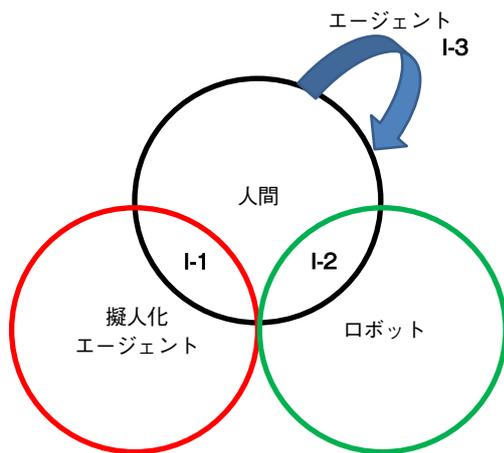
さて、前節で定義された3つのエージェントの間に生じるインタラクションを設計 (デザイン) するのが、HAIヒューマンエージェントインタラクション (human-agent interaction) という研究分野[4]である。ここで、『インタラクション (interaction)』とは、いわゆる「相互作用」と訳される無味乾燥なものではなく、2つの認知主体の間でやりとりされる言語／非言語の豊かなコミュニケーション情報すべてを意味する。

HAIは情報系としては希有 (けう) である日本オリジナルの研究分野であり、筆者と小野哲雄 (北大)、今井倫太 (慶応大)、岡夏樹 (京都工芸繊維大)、竹内勇剛 (静岡大) (敬称略) らによって開拓され、現在国際会議HAIが毎年開催

されるまでに世界的にも認知されている。

HAIの扱うデザイン対象であるエージェント間のインタラクションは、第1図に示すように以下の3つがある。

- I-1. 人間と擬人化エージェント間のインタラクション
- I-2. 人間とロボット間のインタラクション
- I-3. エージェントを介した人間-人間インタラクション



第1図 エージェント間のデザイン対象

また、これらのインタラクションデザインに関する具体的な研究トピックを下記に挙げる。

- エージェントのデザイン
 - アピアランスのデザイン
 - ◇ エージェントはどのような外見、身体を持つべきかを探求
 - エージェントの表出情報のデザイン
 - ◇ エージェントが人間に伝えるべき情報の表現、言語／非言語情報の選択と実装
 - エージェントの機能
 - ◇ 人間との円滑なインタラクションのために、エージェントに必要な機能の開発。多くの場合、AIの機械学習が有効。
 - ◇ 機能（能力）と外見との関係：後述する適応ギャップの研究など。
 - ◇ 学習アルゴリズム
 - エージェントから人間への適応。個人適応、パーソナライゼーション。
 - 人間が適応しやすいエージェントの学習アルゴリズム。XAI：説明可能なAIの研究と類似。
- 人間とエージェントの関係のデザイン
 - 人間とエージェントがどのような関係をもつべきか
 - 協調タスク／競争タスク、ゲーミフィケーション、癒（い）やし、親和性

以上のようなトピックについて、さまざまなHAI研究が行われているが、研究分野自体が文理横断的性質をもつこともあり、統一された一般的方法論はない。しかしながら、人工知能（特に、機械学習）、認知科学、社会心理学、心の理論、心理学、哲学などの研究分野が深く関連している。

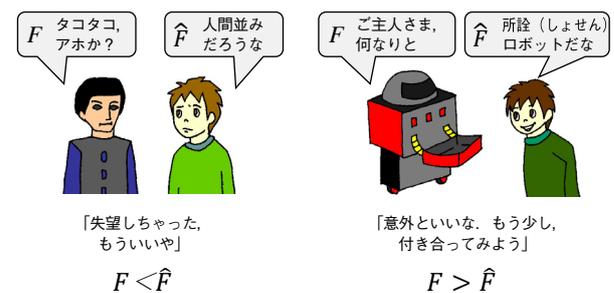
4. HAIの研究例

ここでは、HAI研究のイメージをつかんでいただくために、筆者らのグループの成果を中心に、いくつかの具体的なHAIの研究例を見ていく。

4.1 適応ギャップ

エージェントに始めて対面したとき、人間はエージェントの外見や雰囲気などから、そのエージェントのもつ能力を期待する。そして、そのエージェントとコミュニケーションをもった結果、実際のエージェントの能力が期待した能力より低い場合、がっかりきて、もうそのエージェントを相手にしなくなる。このような現象は、イベントなどに人間そっくりのアンドロイドを配置したときなどに生じる。適応ギャップとは、このエージェントに対する期待と実際の能力との差を定式化したものである[5]。

第2図は相手（エージェント）の能力の真値 F とその推定値 \hat{F} により、適応ギャップ関係を表している。適応ギャップは、 F と \hat{F} の不等式で記述される。



第2図 適応ギャップ

しかし、適応ギャップはあくまで仮説であり、実際に適応ギャップの不等式が、人間とエージェント間のインタラクションの継続に影響するかを実験的に検証する必要がある。筆者らの研究チームは、この適応ギャップ仮説を実験的に検証した[5]。AIBO^(注1)とLEGO^(注2)マインドストーム^(注2)という外見が大きく異なる2台のロボットで、2つの異なる適応ギャップを作り、その違いがインタラクション

(注1) ソニー（株）の登録商標または商標。

(注2) LEGO Juris A/Sの登録商標または商標。

継続に対して、統計的に有意に影響することを示した。

適応ギャップからは、次のようなエージェントの外見に関するデザイン指針が導かれる。人間と継続してうまくやっていくエージェントの外見は、その真の能力を下回る能力を期待させるものが望ましい。つまり、人間と同等レベルの言語コミュニケーションが可能な、いわゆるチューリングテストをパスするAIが存在しない現状においては、人間並みのコミュニケーション能力を期待させるエージェント（例えば、人間そっくりの外見をもつアンドロイドロボット）は、人間とのコミュニケーションを継続するという意味においては、適切なデザインとは言えない。

4.2 エージェントの感情表出の認知探索デザイン

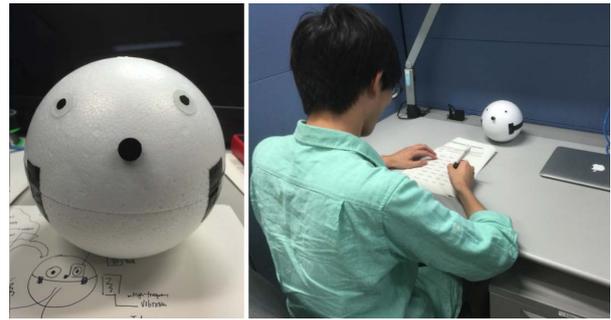
エージェントが人間とうまく協働するには、エージェントに感情をもたせることが重要であるが、内部状態である感情は人間からは直接観測できないので、エージェントが人間に感情を伝えることのできる表出（expression）をデザインすることが必要になる。しかしながら、エージェントの感情表出、特にマルチモーダルによる感情表出については十分に研究が進んでいない。

筆者らの研究チームは、認知的表出デザインという枠組みを提案し、それに基づき、マルチモーダルの感情表出を探索的にデザインする研究を行っている[6]。認知的表出デザインは、下記のような手続きである。

- | |
|--|
| <p>S1. デザイン空間の定義</p> <ul style="list-style-type: none"> ● 数種類のプリミティブ集合 <p>S2. 既知の知見で空間を枝刈り、数十の候補を設定</p> <p>S3. 参加者実験により最後の絞り込み</p> <ul style="list-style-type: none"> ● オンライン実験、絞り込み基準、アンケート |
|--|

この研究では、第3図の左図に示すような球体のロボットの感情表出を探索的にデザインした。まず、ステップS1では、複数のモダリティとしてピープ音、バイブレーション、光のプリミティブを用意し、それらを軸にした3次元空間としてデザイン空間を構成した。次に、ステップS2として、先行研究のメタ分析を行うことで得られた知見を基に、先ほどのデザイン空間を数十のデザイン候補に絞り込む。その後は、ステップS3として、実際にデザイン候補について、参加者実験による評価を実施することで、最終的な数個のデザインを得た。第3図の右図は、参加者実験の様子である。

このような方法により、膨大な候補からなるデザイン空間から感情表出にとって有効なモダリティの組み合わせを効率よく探索できる、つまりデザインすることが可能になる。



第3図 球体ロボットMARUと実験風景

この認知的表出デザインの枠組みは、ドメインに依存せずに広く適用可能である。また、同様の手法で別のドメインである生物発光に関する知識を用いて表出デザインを行い、枠組みの汎用性と有効性を確認している[7]。

4.3 ユーザーの信頼を誘発するバーチャル販売員のデザイン

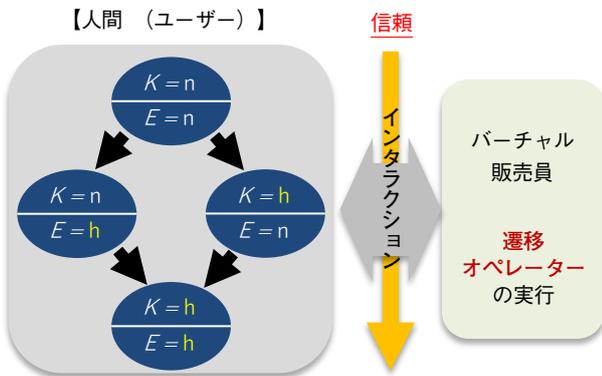
インターネットのオンラインショッピングで買い物をすることは今や日常化しているが、オンラインショッピングには実店舗のような販売員がいないことが課題となっている。そして、オンラインショッピング上に、擬人化エージェントで販売員を実装したのが、バーチャル販売員（第4図）と呼ばれている。バーチャル販売員は、リアルな人間の販売員と同じく、商品説明などを行いつつ、顧客の信頼を得ることで、お客の購買意欲を向上させることが期待される。しかし、バーチャル販売員がお客の信頼を得るには、どうすればいいのだろうか。



第4図 バーチャル販売員

この疑問に対する1つの解決を与える研究を筆者らの研究チームが行っている[8]。社会心理学の知見をベースに、お客の信頼モデル（状態遷移図）をもっているバーチャル販売員が、商品説明に合わせて、お客に種々の刺激を与えることで、お客のバーチャル販売員に対する信頼感を向上させるAIエージェントを構築する。

その信頼モデルは、第5図のようなもので、下位状態の感情状態 E （顧客の気持ち）と知識状態 K （エージェントの商品知識）の論理積で表現される。この状態変数 E と K は、 n （ニュートラル、平常値）と h （高い）のバイナリ値をとる。



第5図 信頼の状態遷移モデル

そして、 $K=E=h$ のとき（第5図の一番下の状態）、つまり説明を受けているお客の気持ちが高揚し、かつお客が「このバーチャル販売員、商品についてよく知っているなー」と感心すると、購買意欲（買う気）がアップするというモデルとなっている。

この研究では、バーチャル販売員が説明することで遷移オペレータが実行され、状態 n/n から状態 n/h 、または h/n をへて h/h に遷移していくことで、最終的にお客の購買意欲が向上することを目指している。実際にこのバーチャル販売員がお客の購買意欲を向上されるのかが疑問となるが、クラウドソーシングを利用した数百人規模の参加者実験により、その効果が実験的に検証されている[8]。

5. まとめ

本稿では、これからますます社会導入、社会浸透が進むであろうAIについて、初めてAIを使う一般ユーザーがAIと対峙（たいじ）したときに生じる「AIに対する戸惑いや違和感」を解消し継続した人間-AI協働の実現を目指す、研究分野であるHAIヒューマンエージェントインタラクションについて紹介した。

続いて、HAIの簡単な歴史と意義、HAIにおけるエージェントの定義とHAIの研究対象である3つのインタラクションデザイン、そして具体的な研究トピックについて議論した。

また、具体的なHAI研究例として、適応ギャップ、エージェントの感情表出の認知探索デザイン、ユーザーの信頼を誘発するバーチャル販売員のデザインについて、その目

的と方法、結果を紹介した。

以上のようにHAIは、AIの社会導入が進む現在においてますますその重要性が高まる、日本発の研究分野である。今後は、情報系の工学の研究開発者に留（とど）まらず、若手を中心に関連する広い分野（もちろん、企業・ビジネスサイドも含む）の研究者の参加が期待される。

参考文献

- [1] 山田誠二, 本当は、ずっと愚かで、はるかに使えるAI: 近未来人工知能ロードマップ, 日刊工業新聞社, 2019.
- [2] 馬場口登, 山田誠二, 人工知能の基礎, オーム社, 2015.
- [3] バイロン・リーブス, クリフォード・ナス, 人はなぜコンピューターを人間として扱うか-「メディアの等式」の心理学, 翔泳社, 2001.
- [4] 山田誠二, 小野哲雄, マインドインタラクション: AI学者が考える「ココロ」のエージェント, 近代科学社, 2019.
- [5] Takanori Komatsu, Rie Kurosawa, Seiji Yamada, "How Does the Difference Between Users' Expectations and Perceptions about a Robotic Agent Affect Their Behavior?," International Journal of Social Robotics, vol. 4, issue 2, pp. 109-116, 2012.
- [6] Song, S. Yamada, S., "Expressing Emotions Through Color, Sound, and Vibration with an Appearance- Constrained Social Robot," Proceedings of the 2017 ACM/IEEE International Conference on Human-Robot Interaction (HRI2017), pp. 2-11, Mar. 2017.
- [7] Song, S., Yamada, S., "Bioluminescence-Inspired Human-Robot Interaction," Proceedings of the 2018 ACM/IEEE International Conference on Human-Robot Interaction (HRI 2018), pp. 224-232, Mar. 2018.
- [8] Matsui, T. Yamada, S., "Designing Trustworthy Product Recommendation Virtual Agents Operating Positive Emotion and Having Copious Amount of Knowledge," Frontiers in Psychology, vol. 10, Apr. 2019.

《プロフィール》

山田 誠二 (やまだ せいじ)

1984	大阪大学 基礎工学部卒業
1987	大阪大学 大学院 修士課程修了
1989	大阪大学 大学院 博士課程修了 工学博士
1989-1991	大阪大学 基礎工学部 助手
1991-1996	大阪大学 産業科学研究所 講師
1996-2002	東京工業大学 大学院 助教授
2002-現在	(共) 情報・システム研究機構 国立情報学研究所 教授
2002-現在	同 総合研究大学院大学 教授
2016-2018	(一社) 人工知能学会 会長

専門技術分野：

人工知能, HAIヒューマンエージェントインタラクション

主な著書：

マインドインタラクション (近代科学社, 2019)

本当は、ずっと愚かで、はるかに使えるAI (日刊工業新聞社版, 2019)

人工知能の基礎 (オーム社, 2015)

主な編書：

AIとともに生きる未来 (文溪堂, 2020)