

2026年5月28日

報道機関 各位

国立大学法人東北大学

敵か味方か？人間とAIにおける 「社会的意図」認識のギャップを解明

【発表のポイント】

- AIには他者の身体の動きの「社会的意図」を人間と同様には理解することができない「アライメント・ギャップ（認識のズレ）^(注1)」が存在することを発見しました。
- 日本と台湾の役者によるモーションキャプチャデータを分析した結果、大きな動き（高エネルギー）の敵意は国籍や文化を越えて共通に伝わる一方、小さな動き（低エネルギー）の敵意は文化的背景を共有する集団にしか伝わらない「文化の方言」として機能することを解明しました。
- AIは高精度で意図を分類することができても、人間の評価との相関は極めて低く、分類精度だけでなく「人間の知覚との一致（知覚的アライメント）」を最適化する新たなAI評価基準の必要性を示しました。

【概要】

遠くから近づいてくる人が「敵」か「味方」かを瞬時に見分けることは、人間の根源的な生存スキルです。人間は声や顔表情が無くとも、ごく小さなボディランゲージから相手の「社会的意図」を難なく読み取ることができますが、人工知能(AI)にとってこれは依然として困難な課題です。

東北大学 電気通信研究所サイバー&リアルICT学際融合研究センターの程苗(Miao CHENG)らの研究チームは、人間の身体的な社会的意図の認識に関する新たな評価基準を構築しました。そして、脅威を伝える際の身体運動のエネルギーレベルによる普遍性と文化特異性(方言)の違いを明らかにしました。さらに、最先端の深層学習モデルは表面的な動作パターンを高い精度で分類できるものの、人間が他者の意図を推論する際の認知プロセスとは大きく乖離している「アライメント・ギャップ」が存在することを突き止めました。本研究は、今後の人間とAI・ロボットの安全なインタラクションにおいて、AIが人間の社会的認識にいかに近づくべきかという重要な指針を示すものです。

本研究成果の論文は、2026年5月25日～29日に京都で開催される国際会議 20th IEEE International Conference on Automatic Face and Gesture Recognition (FG 2026) に採択され、ポスター発表されます。論文は IEEE Xplore で公開されます。

【詳細な説明】

研究の背景

これまで、アフェクティブコンピューティング (Affective Computing) ^(注2) と呼ばれる感情認識の分野やコンピュータビジョンの分野の技術の多くは、「喜怒哀楽」のような基本的な感情の認識や、「歩く」「手を振る」といった物理的な動作の分類に注力してきました。しかし、「他者に対する戦略的なサイン (例:脅かす、挨拶する)」という「社会的意図」の認識は見過ごされてきました。サービスロボットや自動運転車が社会に普及してゆく中、それらに対応する人間が内面でどう感じているかよりも、その行動の意図が「安全 (味方)」か「危険 (敵)」かを判断することの方が、安全を確保する上ではるかに重要です。

今回の取り組み

本研究では、人間が非言語的なサインをどのように表現するかを調べるため、日本と台湾のプロの俳優 80 名による 160 の身体動作 (モーションキャプチャデータ) を収集しました。俳優には、人間の文化や言語を知らない「架空の宇宙人」に対して友好的または敵対的な意図を伝えるよう指示し、純粋なボディランゲージのみを抽出しました。

次に、日本、台湾、中国の 77 名の観察者による評価データを分析した結果、「モーション・エネルギーの閾値」が存在することが分かりました。激しく大きな動き (高エネルギー) を伴う敵意は、文化・国籍を問わず共通の生物学的サインとして認識されました。一方、日本人俳優によく見られた「低エネルギーの敵意 (静かな威圧など)」は、日本人観察者には容易に伝わるものの、台湾や中国の観察者には見過ごされやすい「文化の方言」として機能することが明らかになりました。これは、あからさまな攻撃性を抑えるという文化的背景 (表示規則) に起因すると考えられます。

さらに、最先端の深層学習 AI モデル (ST-GCN^(注3)) を用いてこれらの動作を分類させたところ、重大な盲点が見つかりました。AI は物理的なパターンに基づき 69% の精度で意図を分類できたものの、その判断基準は人間の認識とほとんど一致しませんでした (相関係数 0.26)。異なる 3 文化の人間同士が高い一致率 (相関係数 0.79 以上) を示したのとは対照的です。人間は行動の背後にある「隠された目的」を推論 (逆プランニング) しますが、AI は表面的な物理パターンを照合するだけであり、受動的で攻撃的なごく小さな動作が持つ「社会的な重み」を理解できていなかったのです。つまり、AI が判断に迷う動きと、人間が判断に迷う動きは全く別物でした。

今後の展開

本研究で明白になった AI の「アライメント・ギャップ (認識のズレ)」は、人間と機械のインタラクションにおいて安全上のリスクをもたらします。統計的な分類精度が高くとも、ごく小さな動作による敵意に気づけないシステムでは、トラブルの未然防止や適切な対応ができません。

今後の展開として、研究チームはコンテキスト (文脈) を認識できるアーキテクチャ (Scene Graph GCN など) を統合し、人間が意図を解読する際に用い

る「状況を想像する」戦略をAIモデルで再現することを目指しています。また、基本的な「敵か味方か」の認識にとどまらず、より複雑な社会的意図の解読の研究へと拡張してゆく予定です。最終的にはこの「アライメント・ギャップ」を完全に埋め、現実世界における人間の行動を直感的に深く理解し、安全に稼働できる社会的に知的なロボットやシステムの開発に貢献します。

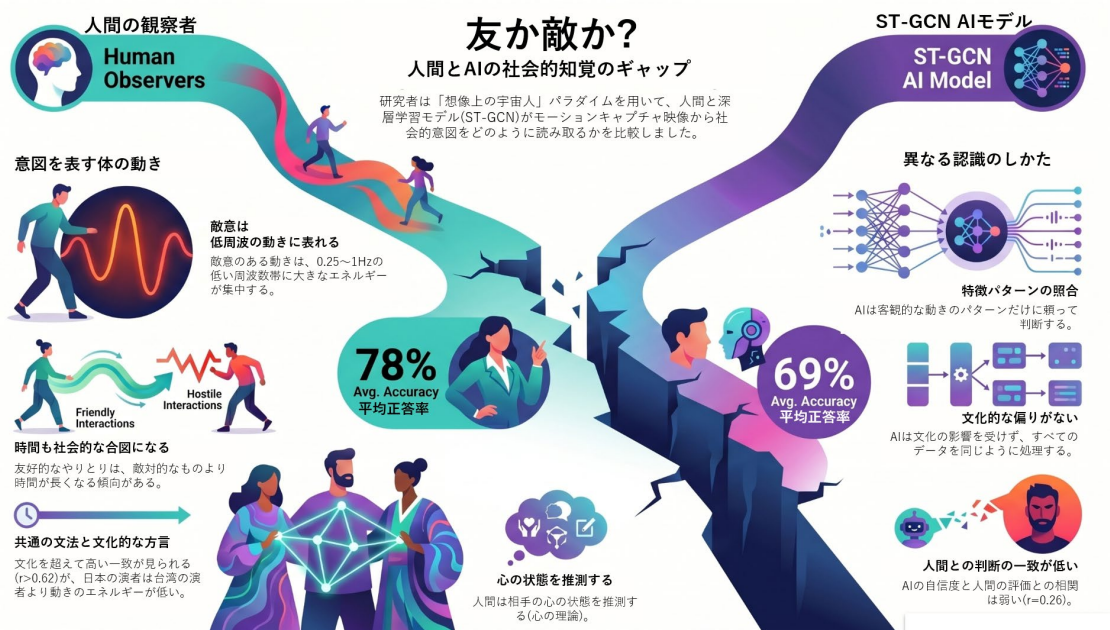


図 1. 研究概要

Example videos:

<https://www.dropbox.com/scl/fi/f7bctxd7eqlg1gf15cn6z/friend-vs-foe-videos.mp4?rlkey=7uq0s4g33vd66iq7jviokbev&st=qfajjwwh&dl=0>

【謝辞】

本研究は、実験に参加いただいた全ての俳優および評価者の皆様のご協力のもと実施されました。

【用語説明】

- 注1. アライメント・ギャップ (Alignment Gap) :AI による予測や分類の確信度が、人間の実際の知覚、判断基準、認知プロセスと一致していない状態のこと。
- 注2. アフェクティブコンピューティング (Affective Computing) :人間の感情や意図を認識、解釈、処理するシステムや人工知能に関する学際的な研究分野。
- 注3. ST-GCN (Spatio-Temporal Graph Convolutional Network) :時空間グラフ畳み込みネットワーク。人間の骨格など、空間的・時間的なつながりを

持つ時系列データを処理し、動作認識などに用いられる代表的な深層学習モデル。

【論文情報】

タイトル: Friend or Foe? Benchmarking Human Perception and ST-GCN Decoding of Embodied Social Intention

著者: Miao Cheng *, Zhan Dai, Victor Schneider, Kanta Ozawa, Yangyang Cai, Ken Fujiwara, Yoshifumi Kitamura, Chia-huei Tseng

掲載誌: Proceedings of the 20th IEEE International Conference on Automatic Face and Gesture Recognition (FG 2026)

DOI: To be confirmed (IEEE Xplore digital library で公開予定)

国際会議 URL : <https://fg2026.ieee-biometrics.org/>

【問い合わせ先】

(研究に関すること)

東北大学 電気通信研究所

サイバー&リアル ICT 学際融合研究センター

特任助教 程苗 (Cheng Miao)

Email: cheng@tohoku.ac.jp

東北大学 電気通信研究所

サイバー&リアル ICT 学際融合研究センター

教授 曾加蕙 (Chia-huei Tseng)

Email: tseng@tohoku.ac.jp

URL: <https://www.cr-ict.riec.tohoku.ac.jp/>

(報道に関すること)

東北大学 電気通信研究所 広報室

TEL: 022-217-5427

Email: riec-kohoshitsu@grp.tohoku.ac.jp