Engineering socially-interactive AI

Elisabeth André

Angaben zur Veröffentlichung / Publication details:

André, Elisabeth. 2026. "Engineering socially-interactive AI." In *EICS '26 companion: companion proceedings of the 18th ACM SIGCHI Symposium on Engineering Interactive Computing Systems, Patras, Greece, 30 June - 3 July 2026*, edited by Andreas Komninos, Carmen Santoro, and Anke Dittmar, 3–5. New York, NY: ACM. https://doi.org/10.1145/3807968.3816801.

Engineering Socially-Interactive AI

Elisabeth André
Institute of Informatics
University of Augsburg
Augsburg, Germany
andre@informatik.uni-augsburg.de

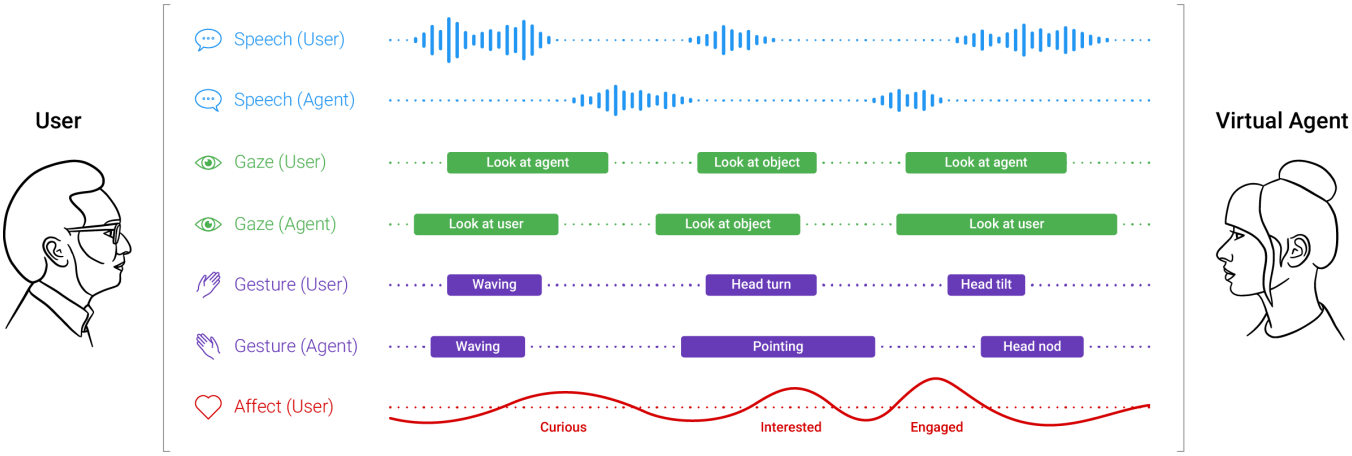

Figure 1: Verbal and non-verbal behaviors of the user and the agent emerge from continuous interaction and reciprocal behavioral adaptation. (Illustration for this paper created by Michael Dietz)

Abstract

In recent years, considerable efforts have been made to improve the expressive behaviors of artificial interactive agents, ranging from animated virtual characters and anthropomorphic robots to digitally enhanced everyday objects. Experience shows, however, that realistic appearance and animation alone do not suffice to maintain long-term engagement beyond an initial novelty effect. Motivated by the goal of enabling smooth, natural interactions, this talk discusses the challenges of engineering socially interactive AI. Particular emphasis is placed on temporal alignment, including the concurrent generation of appropriate system responses, such as backchannels, during ongoing interaction with the human interlocutor. The talk highlights the need for tightly integrated real-time systems that coordinate Multimodal Social Behavior Analysis and Generation under dynamically evolving interaction conditions.

CCS Concepts

• **Human-centered computing** → **Virtual reality**; **Mixed / augmented reality**; **Natural language interfaces**; • **Computer systems organization** → **Robotics**; • **Computing methodologies** → **Intelligent agents**.

EICS Companion '26, Patras, Greece

ACM ISBN 979-8-4007-2669-9/26/06
https://doi.org/10.1145/3807968.3816801

Keywords

Socially-Interactive Agents, Multimodal Behavior Analysis and Generation

ACM Reference Format:
Elisabeth André. 2026. Engineering Socially-Interactive AI. In *Companion Proceedings of the 18th ACM SIGCHI Symposium on Engineering Interactive Computing Systems (EICS Companion '26), June 30–July 03, 2026, Patras, Greece.* ACM, New York, NY, USA, 3 pages. https://doi.org/10.1145/3807968.3816801

1 Introduction

Human communication is inherently multimodal: speech is interwoven with gestures, facial expressions, posture shifts, and subtle affective cues. Motivated by the goal of enabling more natural and intuitive communication with machines, increasing research efforts have focused on socially-interactive agents, including virtual characters and anthropomorphic robots [4].

Recent advances in multimodal representation learning, large language models, and generative models have substantially expanded the behavioral repertoire of socially-interactive systems. Yet socially competent interaction requires more than generating isolated behaviors. Interactive agents must couple social behavior analysis and multimodal behavior generation while continuously reacting to evolving user behavior (see Fig. 1).

While generative AI has substantially improved the realism of artificial agents, the field is now aiming toward closely integrated real-time systems in which verbal and non-verbal behaviors are

shaped by the interactional context as well as by the personality and social role an agent is intended to portray [2].

2 Multimodal Social Behavior Analysis

Multimodal behavior analysis has undergone rapid change over the past decade, driven by advances in deep learning and large-scale multimodal datasets. While robustness has increased, social behavior analysis remains challenging because social cues are inherently ambiguous and depend on the interactional context, and user-specific traits. Modern systems increasingly rely on data-driven multimodal representations that integrate audio, visual, and linguistic information to capture dependencies and complementarities across modalities.

Recent community efforts, such as the MultiMediate Grand Challenge, investigate core challenges in analyzing social and conversational behaviors, including the recognition of backchannels and agreement [5]. Benchmark results from MultiMediate '25 [1] highlight the difficulty of robust engagement estimation in a conversation across cultures and domains. Beyond recognizing social signals, interactive systems must also anticipate conversational behaviors. The first edition of the MultiMediate Grand Challenge, therefore, included next-speaker prediction [6] as an initial step toward timely conversational interventions by an agent.

3 Socially-Aware Behavior Generation

Behavior generation transforms communicative intent into multimodal behavior across speech, gaze, gesture, facial expression, and posture. Natural interaction depends not only on the realism of individual modalities, but also on their semantic and expressive coherence. Inconsistencies between modalities or poorly timed reactions can negatively affect the perception of the interaction with an agent.

Earlier approaches to behavior generation often relied on handcrafted symbolic AI models and manually specified behavior rules. In contrast, modern neural architectures learn the semantics, expressive characteristics, stylistic variations, and temporal organization of human behavior directly from large multimodal datasets. Recent systems increasingly combine large language models for content generation with multimodal generative architectures integrating audio, visual, linguistic, and contextual representations.

Community initiatives such as the GENEA [3] and REACT [7] challenges have advanced benchmarking for multimodal behavior generation through shared datasets and standardized evaluations.

4 Integrated Perspective and Outlook

Human interaction is fundamentally incremental and overlapping: conversational partners interrupt each other, produce backchannels while listening, and adapt to the evolving flow of interaction. Engineering socially-interactive AI, therefore, requires moving from modular, sequential architectures toward integrated, real-time systems that support continuous, reciprocal behavior adaptation between a human and an agent.

Emerging multimodal foundation models provide new opportunities to unify analysis and generation within a single architecture. However, their deployment in interactive systems raises major engineering challenges, including latency, controllability, interpretability, validation, and real-time coordination.

Despite advances, many systems still heavily rely on statistical regularities that only partially capture the complexity of social interactions. Such systems may produce socially plausible responses, but remain only weakly grounded in the conversational setting and fail to adapt across diverse real-world settings.

Evaluating such adaptive behaviors remains a major challenge. Most evaluations still focus on offline analyses of recorded interactions rather than on agents engaged in free conversation.

Beyond technical robustness, simulating socially expressive behavior may raise ethical concerns. Embodied agents may reproduce stereotypes, reinforce social biases, or convey inappropriate social signals in sensitive domains. Generative approaches additionally introduce challenges regarding hallucinations, behavioral inconsistency, and failures to maintain a coherent social role.

Addressing these challenges requires combining advances in machine learning with principles from human-centered design and interactive systems engineering.

References

[1] Daksitha Senel Withanage Don, Marius Funk, Michal Balazia, Huajian Qiu, Shogo Okada, François Brémond, Jan Alexandersson, Andreas Bulling, Elisabeth André, and Philipp Müller. 2025. MultiMediate '25: Cross-cultural Multi-domain Engagement Estimation. In *Proceedings of the 33rd ACM International Conference on Multimedia, MM 2025, Dublin, Ireland, October 27-31, 2025*, Cathal Gurrin, Klaus Schoeffmann, Min Zhang, Luca Rossetto, Stevan Rudinac, Duc-Tien Dang-Nguyen, Wen-Huang Cheng, Phoebe Chen, and Jenny Benois-Pineau (Eds.). ACM, 14150–14155. doi:10.1145/3746027.3762076

[2] Kathrin Janowski, Hannes Ritschel, and Elisabeth André. 2022. Adaptive Artificial Personalities. In *The Handbook on Socially Interactive Agents:20 years of Research on Embodied Conversational Agents, Intelligent Virtual Agents, and Social Robotics Volume 2: Interactivity, Platforms, Application*, Birgit Lugrin, Catherine Pelachaud, and David Traum (Eds.). ACM / Morgan & Claypool, 155–194. doi:10.1145/3563659.3563666

[3] Taras Kucherenko, Alice Delbosc, Rajmund Nagy, Laura B. Hensel, Youngwoo Yoon, Oya Çeliktutan, and Gustav Eje Henter. 2025. GENEA Workshop 2025: The 6th Workshop on Generation and Evaluation of Non-verbal Behaviour for Embodied Agents. In *Proceedings of the 33rd ACM International Conference on Multimedia, MM 2025, Dublin, Ireland, October 27-31, 2025*, Cathal Gurrin, Klaus Schoeffmann, Min Zhang, Luca Rossetto, Stevan Rudinac, Duc-Tien Dang-Nguyen, Wen-Huang Cheng, Phoebe Chen, and Jenny Benois-Pineau (Eds.). ACM, 14305–14307. doi:10.1145/3746027.3762105

[4] Birgit Lugrin. 2021. Introduction to Socially Interactive Agents. In *The Handbook on Socially Interactive Agents: 20 Years of Research on Embodied Conversational Agents, Intelligent Virtual Agents, and Social Robotics Volume 1: Methods, Behavior, Cognition*, Birgit Lugrin, Catherine Pelachaud, and David R. Traum (Eds.). ACM / Morgan & Claypool, 1–20. doi:10.1145/3477322.3477324

[5] Philipp Müller, Michael Dietz, Dominik Schiller, Dominike Thomas, Hali Lindsay, Patrick Gebhard, Elisabeth André, and Andreas Bulling. 2022. MultiMediate'22: Backchannel Detection and Agreement Estimation in Group Interactions. In *MM '22: The 30th ACM International Conference on Multimedia, Lisboa, Portugal, October 10 - 14, 2022*, João Magalhães, Alberto Del Bimbo, Shin'ichi Satoh, Nicu Sebe, Xavier Alameda-Pineda, Qin Jin, Vincent Oria, and Laura Toni (Eds.). ACM, 7109–7114. doi:10.1145/3503161.3551589

[6] Philipp Müller, Michael Dietz, Dominik Schiller, Dominike Thomas, Guanhua Zhang, Patrick Gebhard, Elisabeth André, and Andreas Bulling. 2021. MultiMediate: Multi-modal Group Behaviour Analysis for Artificial Mediation. In *MM '21: ACM Multimedia Conference, Virtual Event, China, October 20 - 24, 2021*, Heng Tao Shen, Yueting Zhuang, John R. Smith, Yang Yang, Pablo César, Florian Metze, and Balakrishnan Prabhakaran (Eds.). ACM, 4878–4882. doi:10.1145/3474085.3479219

[7] Siyang Song, Micol Spitale, Xiangyu Kong, Hengde Zhu, Cheng Luo, Cristina Palmero, Germán Barquero, Sergio Escalera, Michel F. Valstar, Mohamed Daoudi, Tobias Baur, Fabien Ringeval, Andrew Howes, Elisabeth André, and Hatice Gunes. 2025. REACT 2025: the Third Multiple Appropriate Facial Reaction Generation Challenge. In *Proceedings of the 33rd ACM International Conference on Multimedia, MM 2025, Dublin, Ireland, October 27-31, 2025*, Cathal Gurrin, Klaus Schoeffmann, Min Zhang, Luca Rossetto, Stevan Rudinac, Duc-Tien Dang-Nguyen, Wen-Huang

Cheng, Phoebe Chen, and Jenny Benois-Pineau (Eds.). ACM, 13979–13984. doi:10. 1145/3746027.3762244