Supporting Mutual Theory of Mind through AR-Enhanced Conversational User Interfaces in Veterinary Care

Tobias Münch
to.muench@muench-its.de
Münch Ges. für IT-Solutions mbH
Lohne (Oldenburg), Germany
Chemnitz University of Technology
Chemnitz, Germany

Thomas Kosch
thomas.kosch@hu-berlin.de
HU Berlin
Berlin, Germany

Sebastian Heil
sebastian.heil@informatik.tu-chemnitz.de
Chemnitz University of Technology
Chemnitz, Germany

Martin Gaedke
gaedke@informatik.tu-chemnitz.de
Chemnitz University of Technology
Chemnitz, Germany

Abstract

Augmented Reality (AR) glasses offer significant potential to advance conversational user interfaces (CUIs), particularly in challenging, hands-free settings such as veterinary care in animal barns. Effective communication in such scenarios relies heavily on mutual Theory of Mind (ToM)—the ability to understand intentions and coordinate joint actions. However, dynamic, noisy, and cognitively demanding veterinary environments pose substantial obstacles to the formation and maintenance of a mutual ToM, and traditional interfaces often fail to support these essential cognitive processes adequately. This position paper argues that integrating AR glasses equipped with cameras capable of filming and interpreting environmental contexts, along with synchronized audio-visual feedback, can profoundly influence mutual ToM among veterinary professionals. By applying concepts from cognitive psychology and the Human-Agent Speech Interaction (HASI) model, our discussion examines how Augmented Reality (AR) can enhance cognitive processes in real-world settings. AR glasses can improve aspects of conversation, including recognizing intentions, interpreting visual context, and adapting communication strategies. AR with mutual ToM can enhance usability, reduce cognitive load and improve understanding among communication partners. We establish a foundation and emphasize the necessity of empirical research to explore and validate these concepts.

Keywords

Conversational User Interfaces, Theory of Mind, Augmented Reality Glasses, Hands-Free Veterinary Collaboration

ToMinHAI '25, Waterloo, ON, Canada

ACM ISBN 978-x-xxxx-xxxx-x/YYYY/MM
https://doi.org/10.1145/nnnnnnn.nnnnnnn

ACM Reference Format:
Tobias Münch, Thomas Kosch, Sebastian Heil, and Martin Gaedke. 2025. Supporting Mutual Theory of Mind through AR-Enhanced Conversational User Interfaces in Veterinary Care. In *Proceedings of Theory of Mind in Human-CUI Interaction Workshop @ ACM CUI 2025 (ToMinHAI '25).* ACM, New York, NY, USA, 3 pages. https://doi.org/10.1145/nnnnnnn.nnnnnnn

1 Introduction

AR glasses are currently being trialled as a promising platform for advancing CUIs in hands-free professional contexts [1, 9, 16]. In veterinary medicine, particularly in animal barns, veterinarians operate in dynamic, noisy, and high-pressure environments where efficient collaboration and real-time information exchange are crucial. Effective teamwork in these settings depends on a shared understanding of intentions and rapid adaptation to changing scenarios—an ability grounded in the mutual formation of Theory of Mind (ToM) [3, 15]. However, maintaining ToM in such challenging environments is non-trivial. Traditional user interfaces—even conversational systems—often fail to adequately support the cognitive processes required for seamless communication and action coordination, particularly under constraints of mobility, hygiene, and environmental complexity [9].

A new generation of AR glasses, such as Meta Glasses, now combine hands-free visual feedback, real-time audio interaction, and environmental sensing through integrated cameras [9]. These devices are not limited to passive information display; they can actively capture and interpret visual context, offering a new design space for multimodal conversational support. By linking audio-visual signals with environmental context, AR-based CUIs can lighten cognitive load, boost situational awareness, and help veterinary teams develop a shared understanding of each other's mental states [1, 16].

This position paper addresses a central question: **How can AR glasses equipped with conversational interfaces and environmental perception capabilities reshape mutual Theory of Mind in the context of hands-free, high-stakes veterinary work?** We argue that the integration of AR glasses with synchronized audio-visual feedback, supported by cognitive models such as the HASI model [16] and established principles from cognitive psychology [3], opens new opportunities for augmenting key aspects of team communication.

2 Background

Advancing hands-free CUIs in complex environments requires a multidisciplinary understanding of cognitive psychology, interaction models, and the unique constraints of wearable augmented reality technologies.

Theory of Mind in Human-CUI Interaction: ToM describes the ability to infer, model, and anticipate the intentions and mental states of others [3, 13]. In the context of Human-CUI interaction, ToM is not only a fundamental cognitive faculty for natural conversation, but also a key design challenge for artificial agents [11, 15]. Recent work has emphasized the importance of Mutual Theory of Mind in Human-AI and Human-Agent Interaction, arguing that effective collaboration with intelligent systems requires both humans and machines to iteratively shape, update, and communicate their interpretations of each other's mental states [15]. Designing CUIs capable of supporting and reflecting user intentions through dialogue, contextual cues, and adaptive strategies remains an open and critical problem in human-centered AI (HCAI) [7, 16].

Cognitive Models in Conversations: To create effective CUIs, we need to understand how people recognize, process and generate speech. Conversation models (e.g. the interactive alignment model [12] and Norman's action cycle [10]) highlight the back-and-forth nature of dialogue. This involves setting goals, taking action, providing feedback, and assessing responses. More recently, cognitive architectures for speech-based systems incorporate dynamic context, multimodal signals, and adaptation to user traits [6, 16]. Integrating these cognitive models with AR-based CUIs enables more adaptive, and contextually aware conversational experiences, allowing systems to support intention recognition better and reduce communication breakdowns [1, 4].

Veterinarians and Hands-Free Challenges: Veterinarians working in barns and animal care settings face challenging operational constraints: hands are most of the time occupied, environments are noisy, and strict hygiene protocols limit interaction with shared devices [8, 9, 14]. Traditional interfaces (e.g., touchscreens or handheld devices) are often impractical, leading to workflow disruptions or communication delays [2, 5]. There is a need for hands-free, context-aware technologies that integrate with veterinarians' workflows, support real-time communication, and are robust in the face of environmental challenges [9, 15]. Recent studies in medical and industrial contexts have demonstrated that AR glasses, when combined with multimodal CUIs, can address many of these issues; however, domain-specific research for veterinary applications remains limited [2, 14].

3 Enhancing Mutual ToM with AR Glasses

Recent advances in AR glasses create opportunities to actively support the formation and maintenance of mutual ToM in hands-free, high-stakes environments such as veterinary medicine. By integrating environmental perception, multimodal feedback, and speech-centric interaction models, AR-based CUIs can address long-standing cognitive and practical challenges in collaborative work [9, 15, 16].

Environmental Perception via Camera Input: Currently developed AR glasses, such as Meta Glasses, contain front-facing cameras that can capture the visual context of the user's environment [9]. In contrast to authentic AR glasses, Meta glasses do not have a digital overlay [9]. This enables context-aware CUIs to recognize ongoing activities, identify relevant objects, and adapt system responses in real time. For veterinarians, environmental perception means the system can distinguish between routine and emergency procedures, track the presence and behavior of animals, and support hygiene protocols by minimizing the need for physical interaction [2, 5]. By embedding camera-driven contextual awareness into dialogue management, AR CUIs become better partners in collaborative activity, proactively anticipating needs and reducing misunderstandings—a key element of mutual ToM [15].

Audio-Visual Multimodal Interaction: AR glasses facilitate hands-free interaction not only through audio but also via visual feedback and overlays [1, 14]. Synchronized audio-visual outputs support a richer, more robust channel for communication—especially in noisy or unpredictable environments. Visual prompts (e.g., step-by-step checklists, contextual alerts, and images) can complement spoken guidance, reinforcing understanding and supporting rapid, coordinated decision-making [4, 16]. This multimodal approach is especially important for maintaining mutual ToM: the system can both clarify its intentions and confirm its recognition of the user's goals, enabling smoother and more effective turn-taking and repair strategies [6, 7].

Hands-Free Veterinary CUIs with HASI: The HASI model provides a structured framework for designing CUIs that account for the dynamic interplay between user actions, system responses, individual traits, and broader context [16]. Applying HASI to hands-free veterinary CUIs emphasizes three critical layers: (1) interaction (user and system actions, timing, phrasing); (2) traits (user expertise, emotional state, system reliability); and (3) context (ongoing activities, environmental challenges, team roles). By operationalizing HASI in AR glasses, designers can ensure that system feedback is both relevant and adaptive—supporting veterinarians as they navigate complex, high-cognitive-load workflows [2, 9]. This structured interaction promotes mutual ToM by making system states, intentions, and limitations transparent, enabling human users to build accurate models of the CUI and vice versa [15, 16].

4 Model and Implications

Building on the HASI model and contemporary cognitive frameworks, we envision a model for AR-glasses-supported conversational interaction in hands-free veterinary environments.

Interaction Flow and Cognitive Processes: The interaction flow in our model is cyclical and adaptive: veterinarians initiate dialogue with the AR glasses via speech, gestures, or contextual triggers, while the system leverages camera input to interpret environmental cues and user actions [2, 9]. The CUI responds with tailored audio and visual feedback, supporting key cognitive steps—such as goal formation, action planning, execution, and evaluation—aligned with established models like Norman's action cycle [10, 16]. By bridging the gap between user intent and system understanding, this flow seeks to reduce cognitive load and encourage smooth, error-tolerant interactions in high-stakes environments [3].

Communication Strategies: Our model supports a range of communication strategies for both routine and critical veterinary

tasks. Proactive system prompts, context-aware clarifications, and multimodal confirmation cues help resolve ambiguities and breakdowns before they disrupt workflow [6, 7]. The AR system can dynamically switch between modalities—emphasizing visual overlays in high-noise situations or audio in visually complex contexts—to optimize message delivery and comprehension [1, 4]. Feedback is iterative: both human and CUI partners update their mental models based on ongoing interactions, supporting flexible adaptation and rapid repair.

Mutual ToM Formation and Maintenance: Central to the hypothetical model is the continuous, bidirectional formation of mutual ToM. The system infers the user's goals, emotional state, and situational needs through verbal, nonverbal, and contextual signals, while transparently exposing its own state, limitations, and decision rationale [15, 16]. This recursive interpretation-feedback loop enables users to calibrate their expectations, fostering trust and efficient division of labor [11]. The AR-enabled CUI becomes a cognitive partner for complex decision-making and facilitating shared understanding even under the unpredictable demands of veterinary practice. Future empirical studies are needed to validate the effectiveness of this mutual ToM model in real-world deployments.

5 Discussion

While AR-glasses-enabled CUIs present clear potential for enhancing mutual Theory of Mind in veterinary practice, several important limitations and challenges must be addressed before deployment in real-world settings. Technological constraints—including internet connectivity, robustness of speech recognition in noisy barns, and accuracy of environmental perception—may limit effectiveness [2, 9]. Using cameras in veterinary settings requires consideration of ethical principles and privacy for maintaining the veterinary staff and animal owners. Additionally, becoming too dependent on technology is a challenge because it can lead to a decline in human skills [5, 8]. To foster trust in these high-stakes environments, it is crucial to prioritize transparency, provide clear explanations, and ensure that users retain control over the technology. Additional, It is crucial to address negative biases and provide users with practical training to make an usage in a veterinary setting possible. Additionally, adapting interaction models and cognitive support to diverse user skill levels and workflows remains an ongoing research challenge.

6 Conclusion and Future Work

AR glasses with conversational interfaces and environmental perception can enhance hands-free, context-aware support in veterinary medicine. By establishing a mutual ToM through synchronized audio-visual cues, these systems can reduce cognitive load, enhance team communication, and improve workflow safety and efficiency. Despite ongoing research and development challenges, this approach could be a starting point for the next generation of HCAI in busy professional settings.

Future research should focus on critically reflecting on the proposed model through field studies and controlled experiments in veterinary settings. Key areas of exploration include the effect of AR-supported mutual Theory of Mind on communication efficiency, error rates, cognitive load, and user satisfaction. Collaborating with veterinary professionals during the design and evaluation process is crucial to ensure that solutions meet their specific needs and requirements. Additionally, ethical guidelines and best practices for camera-enabled, AI-driven CUIs should be developed with input from domain experts and stakeholders [8, 14].

Acknowledgments

This work is supported by the European Union's HORIZON Research and Innovation Programme under grant agreement No 101120657, project ENFIELD (European Lighthouse to Manifest Trustworthy and Green AI).

References

[1] Benett Axtell and Cosmin Munteanu. 2021. Tea, Earl Grey, Hot: Designing Speech Interactions from the Imagined Ideal of Star Trek. In *CHI Conference on Human Factors in Computing Systems (CHI '21)*. Yokohama, Japan, 1–14. doi:10.1145/3411764.3445640

[2] Kieran Brophy, Samuel Davies, Selin Olenik, Yasin Çotur, Damien Ming, Nejra Van Zalk, Danny O'Hare, F Guder, and Ali K Yetisen. 2021. The future of wearable technologies. *Imperial College London: London, UK* (2021).

[3] Michael W. Eysenck and Mark T. Keane. 2015. *Cognitive Psychology: A Student's Handbook* (7th ed.). Psychology Press.

[4] Kerstin Fischer. 2015. Conversation, construction grammar, and cognition. *Language and Cognition* 7, 4 (2015), 563–588.

[5] Cesar Herrero, D Villar Onrubia, J Cosgrove, S Kluzer, J Centeno, S Romero Rodríguez, C Moreno, L Morilla, A Arroyo Sagasta, A Zubizarreta, et al. 2024. Digital Transformation of VET. (2024).

[6] Stephen C Levinson. 2016. Turn-taking in human communication–origins and implications for language processing. *Trends in cognitive sciences* 20, 1 (2016), 6–14.

[7] Ewa Luger and Abigail Sellen. 2016. "Like Having a Really Bad PA": The Gulf Between User Expectation and Experience of Conversational Agents. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*. ACM, 5286–5297. doi:10.1145/2858036.2858288

[8] Ashley Mitek, Dylan Jones, Andrew Newell, and Samantha Vitale. 2022. Wearable devices in veterinary health care. *Veterinary Clinics: Small Animal Practice* 52, 5 (2022), 1087–1098.

[9] Tobias Münch, Sebastian Heil, and Martin Gaedke. 2025. Building Conversational User Interfaces: An Architectural Exploration with Meta Glasses for Developers and Researchers. In *ACM Conversational User Interfaces 2025 (CUI '25)*. Waterloo, Canada.

[10] Don Norman. 2013. *The Design of Everyday Things: Revised and Expanded Edition*. Basic Books.

[11] Candida C Peterson and Michael Siegal. 1995. Deafness, conversation and theory of mind. *Journal of child Psychology and Psychiatry* 36, 3 (1995), 459–474.

[12] Martin J. Pickering and Simon Garrod. 2004. Toward a mechanistic psychology of dialogue. *Behavioral and Brain Sciences* 27, 2 (2004), 169–190. doi:10.1017/S0140525X04000056

[13] David Premack and Guy Woodruff. 1978. Does the chimpanzee have a theory of mind? *Behavioral and Brain Sciences* 1, 4 (1978), 515–526. doi:10.1017/S0140525X00076512

[14] Petros Spachos, Stefano Gregori, and M Jamal Deen. 2022. Voice activated IoT devices for healthcare: Design challenges and emerging applications. *IEEE Transactions on Circuits and Systems II: Express Briefs* 69, 7 (2022), 3101–3107.

[15] Qiaosi Wang and Ashok K. Goel. 2024. Mutual Theory of Mind for Human-AI Communication. In *Workshop on Theory of Mind in Human-AI Interaction at CHI 2024 (ToMinHAI at CHI 2024)*. https://arxiv.org/abs/2210.03842

[16] Nima Zargham, Vino Avanesi, Thomas Mildner, Kamyar Javanmardi, Robert Porzel, and Rainer Malaka. 2024. HASI: A Model for Human-Agent Speech Interaction. In *ACM Conversational User Interfaces 2024 (CUI '24)*. Luxembourg, Luxembourg, 1–8. doi:10.1145/3640794.3665885