Think Like a Team: Graph-Based Representation of Shared Mental Models in Human-Agent Collaboration

Moinam Chatterjee[1], *Behnam Ensan*[1], *Andreas Wendemuth*[1], *Ayoub Al-Hamadi*[1]

[1]*Otto-von-Guericke-Universität, 39106 Magdeburg, Germany*
moinam.chatterjee@ovgu.de; behnam.ensan@ovgu.de; andreas.wendemuth@ovgu.de; ayoub.al-hamadi@ovgu.de

Abstract: We introduce a conceptual temporal Shared Mental Model (SMM) that maps changing human-human and human-agent team dynamics, allowing Human-Agent Teaming Systems (HATs) to achieve shared teaming goals. We use temporal graph neural networks (TGNNs) to capture the evolving roles and tasks within the team, learnt from time-stamped team interactions. We conduct a proof-of-concept exploratory small-scale user study to observe, in real time, the evolution of team dynamics and interrelationships among team members. This study simulates collaborative tasks involving human and AI agents, enabling direct observation and measurement of teaming behaviours. The proposed model bridges the current research gap in HATs and SMMs by offering a graph-based representation of agentic teaming dynamics.

1 Introduction

Artificial Intelligence (AI) has rapidly evolved into increasingly versatile agents capable of engaging in complex human interactions. The rise of conversational systems such as ChatGPT illustrates how large language model (LLM) architectures can support active decision-making across a wide range of domains [1]. These advances have driven a paradigm shift toward Human-Agent Teaming Systems (HATs) [2], in which AI agents are conceived as collaborative teammates rather than passive tools for productivity. For effective collaboration within such teams, AI agents must understand the conversational teaming context and dynamically adapt to evolving team dynamics, including user roles, alignment states, and intentions. This perspective aligns with the principles of human-centred AI, which emphasise augmenting human capabilities and ensuring trust and transparency in decision-making [3].

With the introduction of HATs, a model is needed that can develop a shared mental model of goals, tasks, and situational context. Shared mental models (SMMs) play a critical role in Human–Agent Teaming systems (HATs) that can fulfil this role [4]. SMMs are a foundational concept in team cognition, referring to the common understanding that team members develop about tasks, roles, and processes. Studies show that integrating SMMs into team decision making improves performance, reduces communication overhead, and enhances resilience under stress [5, 6].

However, current HATs struggle to adapt dynamically to evolving goals and to perceived user roles [2]. When goals change rapidly, humans and AI agents fail to maintain their shared mental models (SMMs). This happens because AI systems often rely on predefined objectives and static training data, making it difficult for systems to be dynamic and adapt to evolving team conditions. Misalignment between shared goals and a common understanding among team members can lead to a breakdown in trust and coordination [7]. This necessitates a framework that enables AI agents to maintain shared cognition in real time, enabling them to continuously update and synchronise their mental models.

In this paper, we introduce a conceptual temporal shared mental model framework that maps changing human-human and human-agent team dynamics, allowing Human-Agent Teaming Systems (HATs) [2] to achieve shared teaming goals under changing teaming conditions. We use temporal graph neural networks (TGNNs) [8], which extend graph neural networks (GNNs) to handle graphs evolving over time. We aim to lay the foundation for research on the application of temporal graph neural networks in Shared Mental Models (SMMs) that can achieve shared teaming goals in changing teaming conditions. We believe that when embarking on such an undertaking, it is essential to understand the current state of Human-Agent Teaming Systems (HATs) and Shared Mental Models (SMMs), as well as the gaps that arise in dynamic goal adaptation and user role perception. We investigate how Shared Mental Models (SMMs) are used to maintain team cognition in Human-Agent Teaming Systems (HATs) and how they can be adapted to evolve the dynamic changes in teaming goals and user roles. We conducted a small-scale proof-of-concept exploratory user study to observe, in real time, the evolution of team dynamics and interrelationships among team members. This study simulates collaborative tasks involving human and AI agents, enabling direct observation and measurement of teaming behaviours. In the study, we test whether the temporal shared mental model shapes participants' alignment experiences and their social roles in collaboration.

2 Related Work

The importance of shared mental models (SMMs) as a foundation for effective collaboration has been heavily emphasised in the research on Human–Agent Teaming Systems (HATs) [9]. In traditional human teams, SMMs enable members to anticipate each other's actions, coordinate implicitly, and adapt to dynamic environments [7]. Researchers have been seeking to extend this concept to human–agent teams, enabling AI agents in autonomous systems to maintain alignment between their shared tasks, goals, and user roles with those of their human teammates [6].

One of the early frameworks was proposed by Scheutz et al. [10], which provided a structured approach for developing, designing, and using shared mental models (SMMs) in Human–Agent Teaming Systems (HATs). It highlighted how agents could be engineered to monitor team states and communicate intentions, ensuring their internal model remains aligned with human team members. This work bridged cognitive systems theory with practical design principles for HATs by introducing a structured, formalised mechanism that agents use to contribute to a shared understanding. While designing HATs, it is also necessary to discuss what characteristics can make these agents perceived as effective teammates by their human counterparts. This is where Wynne and Lyons [11] proposes an agent teammate-likeness (AAT) conceptual model that integrates dimensions such as predictability, adaptability, and role consistency into the development of HATs. They discuss that teammate-likeness should not only be perceived as a technical capability of the agent but also significantly affects the agent's ability to foster trust and shared cognition with humans. This integrative approach underscores the socio-cognitive aspects of HATs, situating SMMs as central to perceptions of agent reliability and collaboration quality.

Andrews et al. [4] backs the notion of teammate-likeness by highlighting how alignment of task and team mental models can improve coordination and trust between humans and intelligent agents. They propose frameworks that enable AI systems to act as equal partners in collaboration by further examining design approaches for building shared mental models between humans and AI, drawing inspiration from computer-supported cooperative work (CSCW) and user experience design. These studies extend the insights from Schelble et al. [9], indicating that one of the biggest challenges in integrating SMMs into HATs is ensuring continuous alignment and recalibration as tasks evolve beyond the system's trust and performance. Kaur et al. [12] also notes that in complex domains like healthcare and autonomous driving, current approaches struggle with misalignment in shared team dynamics. In complex domains, humans prioritise ethical or contextual considerations, whereas AI systems may optimise for efficiency, leading to misalignment. Narayanan and Feigh [13] discusses how the rigid integration of SMMs is underplayed in current frameworks, which often emphasise trust and performance outcomes. They also find that rigid structuring of SMMs can limit flexibility, making teams brittle when tasks or membership evolve. They emphasise that efforts must be made to understand the teaming activity in more complex structures, such as hierarchies. Socio features such as team dynamics, social interplay, and communication significantly affect the evolution of SMMs over time. These limitations underscore the need for adaptive mechanisms that allow SMMs to evolve alongside changing team structures, user roles, and task demands.

3 Method

We propose a conceptual temporal shared mental model integrated into HATs that provides continuous alignment and recalibration of task and team mental models. As illustrated in the figure 1, to support dynamic adaptation, we incorporate TGNs into the system design.

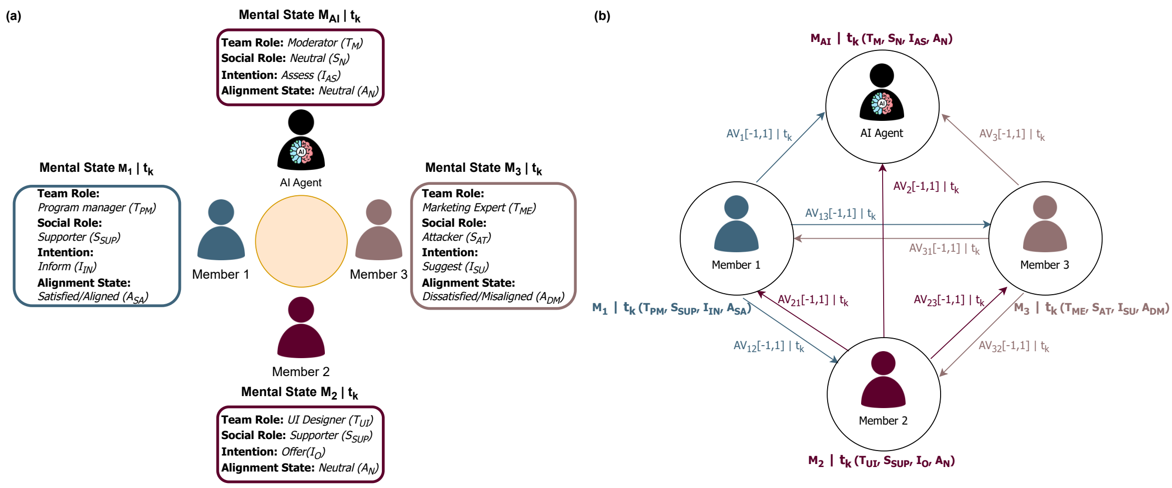

Figure 1 – Part (a) shows a team meeting scenario with four human team members and one AI agent. Each team member has a mental state defined by their perceived technical role, social role, speech intention and alignment state. Part (b) shows how the meeting scenario can be mapped using a Temporal Graph Neural Network. The edges are assigned a weight AV[-1,1] computed from equation 2.

To show how the conceptual temporal shared mental model can be applied to a meeting scenario, we take four team members comprising three human members and one AI team member, as shown in figure 1 (a). Our proposed temporal shared mental model has a mental state M for each team member, comprising four parameters: team role, social role, intention, and alignment state, that shape the cognitive model at timestamp t_k. A team role is a predefined, formal role that remains fixed throughout the meeting. This incorporates understanding the hierarchies present in the meeting, reflecting how a real-life meeting unfolds. The social roles, unlike the formal team roles, are informal roles assumed by team members that reflect their individual traits and personalities. These are not fixed and can change based on interactions with other team members and evolving team dynamics. We use the social role labels from the survey dataset annotated by Sapru and Bourlard [14] for a subset of meetings from the AMI Meeting Corpus [15] as the parameter social role for our proposed mental model. There are four social roles: Gatekeeper, Protagonist, Supporter, Neutral and Attacker. Gatekeeper(S_G) is a team member who serves as the meeting moderator, encouraging communication within the group, Protagonist (S_P) asserts their authority, providing personal perspectives and driving the conversation, Supporter (S_{SUP}) pays attention and demonstrates a cooperative attitude towards other team members, Neutral (S_N) accepts ideas from other team members in a passive manner and Attacker (S_A) attacks others and expresses disapproval. The intention parameter in the mental model corresponds to the underlying purpose or goal that the team member is pursuing in an utterance. We use the dialogue-act labels from the AMI Meeting Corpus [15], such as Inform, Assess, Suggest, and Offer, to fill the mental model's intention values. Inform (I_{IN}) states that the team member wants to provide new information, facts, or explanations. Assess (I_{AS}) displays the intention to evaluate, judge, or appraise a topic in hand. Suggest (I_{SU}) displays the intention to propose a course of action or idea, and Offer (I_{SU}) refers to a team member volunteering to perform an action. The speaker's alignment state indicates whether a team member is aligned or misaligned with the rest of the team, reflecting their level of satisfaction or dissatisfaction at that moment. Each parameter contributes distinct but interdependent influences on decision-making and team dynamics; thus, the proposed shared mental model can govern how team members interpret situations, anticipate others' behaviour, and coordinate actions. All of these parameters can be obtained in real time from both verbal behavioural data (speech audio, speech transcriptions) and non-verbal behavioural data (gaze, head, and hand movements) as provided in the AMI Meeting Corpus dataset [15].

Now, since we have a mental state for each of the team members, to demonstrate the mapping of team dynamics, we take the help of the temporal graph networks (TGNs) [8] to create a graph between the team members as shown in figure 1 (b). The nodes in the graph shown in the figure represent team members' (humans and agents) mental state M, and the edges between them represent their relationships, highlighting the team's interconnectedness and overall dynamics at time t_k. As team members interact, exchange information, and update their understanding, the graph's structure changes, mirroring the continuous updates made by TGNs on dynamic graphs. Using TGNs, the change in graph structure can be analysed to see how teams incorporate new information and maintain a shared understanding of goals over time, representing both synchronisation (alignment of understanding) and divergence (misalignment or breakdowns) among team members. Coherence (alignment) and Incoherence (misalignment) play a big part in group decision-making, as stated in the HOTCO theoretical model [16]. It is a computational model of cognition that integrates alignments into decision-making processes and builds on the coherence theory, which holds that a consistent fit among beliefs, goals, and values is achieved before a decision is made. The model treats alignment as a weight and introduces positive and negative alignment valence, with positive alignment strengthening commitment to a choice and negative alignment weakening it. To incorporate this into the team dynamics, the Edge weight AV is introduced as shown in figure 1 (b), resembling alignment valence between team members. Explanation of the equations used to calculate the alignment valence is provided below.

$$x_{\text{norm}} = 2 \cdot \frac{x - x_{\min}}{x_{\max} - x_{\min}} - 1, \quad x \in \{S, A\} \tag{1}$$

Equation 1 represents a normalisation process that maps a variable x (such as S or A) from its original range $[x_{\min}, x_{\max}]$ to a standardized range of $[-1, 1]$. This transformation ensures that different variables are expressed on a common scale, making them easier to compare or integrate into further computations.

$$AV_{AB}|t_k = \frac{w_S \cdot S_{norm} + w_A \cdot A_{norm}}{w_S + w_A} \tag{2}$$

Equation 2 defines the alignment valence AV_{AB} between two team members, A and B, at a given timestamp t_k. The valence is computed as a weighted average of two normalised components: S_{norm}, representing the social role of team members, and A_{norm}, representing their alignment state. The weights w_S and w_A determine the relative importance assigned to each factor in the calculation. A higher w_S emphasises the influence of social roles, while a higher w_A emphasises the alignment state. By combining social positioning and alignment dynamics into a single measure, the formula captures how both contribute to the overall alignment relationship between team members at a specific point in time.

4 Results and Discussion

To test whether temporal shared mental models successfully shape participants' alignment experience and social roles in collaboration, we conduct a user study to observe, in real time, the evolution of team dynamics among team members. The goal is to simulate collaborative tasks to enable direct observation and measurement of teaming behaviours. A survey[1] was published, inviting participants to rate the speaker's alignment state based on their speaking style in the provided team meeting audio clip. The audio clips for the survey are taken from a scenario-based meeting from the AMI Meeting Corpus [15] where the team is tasked with developing a new remote control. In the survey, each audio clip features a conversation between two team members during a meeting. A 5-level rating scale (1-5) is provided, ranging from dissatisfied and misaligned (1) to neutral (3) to satisfied and aligned (5). We record this rating for both speakers in the meeting and use it as a representative of team members' alignment state, which is used in the calculation of alignment valence and is influential in understanding inter-relationships between team members.

Table 1 – Simulation dataset containing the alignment states, social roles and directional conversation flow among four speakers in a team meeting scenario.

Timestamp	Attribute	Member A	Member B	Member C	Member D	Conversation Flow
t_1	Alignment State	3.25	3.25	-	-	B -> A
	Social Role	Neutral	Gatekeeper	-	-	
t_2	Alignment State	-	4	-	3.25	B -> D
	Social Role	-	Gatekeeper	-	Protagonist	
t_3	Alignment State	2.5	-	3.5	-	A -> C
	Social Role	Protagonist	-	Neutral	-	
t_4	Alignment State	-	2.75	-	3.25	B -> D
	Social Role	-	Gatekeeper	-	Supporter	
t_5	Alignment State	3	1.75	-	-	A -> B
	Social Role	Supporter	Gatekeeper	-	-	
t_6	Alignment State	-	3	2.5	-	B -> C
	Social Role	-	Gatekeeper	Supporter	-	
t_7	Alignment State	2.25	2.75	-	-	B -> A
	Social Role	Protagonist	Gatekeeper	-	-	
t_8	Alignment State	3	3	-	-	A -> B
	Social Role	Neutral	Gatekeeper	-	-	

The simulation dataset is captured for use in simulating the conceptual temporal shared mental model. The dataset contains conversations among four human team members, a sample of a meeting extracted from the AMI Meeting Corpus [15]. The idea is to simulate relationships among team members, observe changes over time, and see how they affect the overall team dynamics. The Table 1 presents the simulation dataset that contains records for eight

[1] https://befragungen.ovgu.de/ThinkTeam-ESSV2026/

timestamps, each with both alignment state values derived from survey responses and the team member's social roles. Alignment state entries quantify each team member's affective alignment and misalignment with their teammate in the conversation. At the same time, the social role row records the informal roles speakers assume, influenced by changing dynamics of group interaction. The last column, Conversation Flow, summarises the directional exchange of communication at each timestamp, indicating which speaker addressed whom (e.g., "B $\rightarrow$ A"). At each timestamp, there is a back-and-forth conversation between the speakers, meaning that when Speaker A addresses Speaker B, Speaker B also engages in the discussion, providing an answer. The survey dataset provides turn-taking data and role dynamics that evolve throughout the conversation, as shown in the simulation in Figure 2.

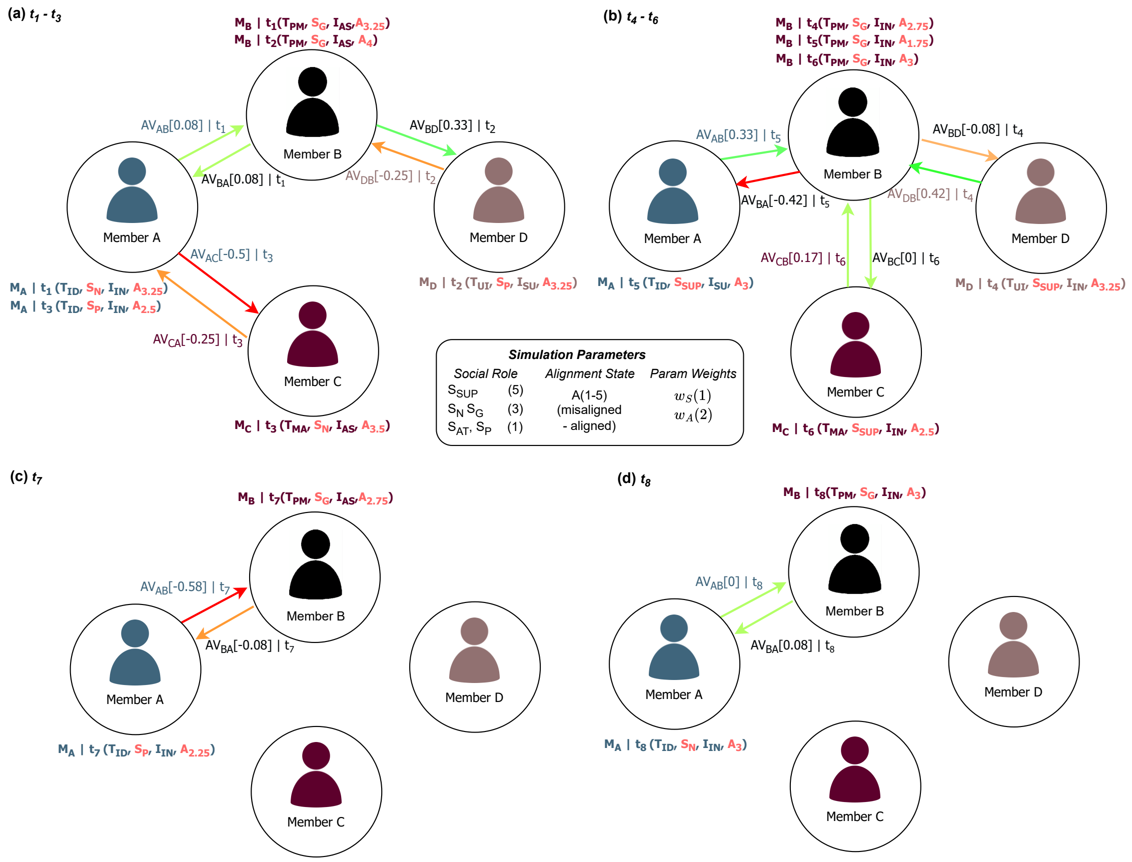

Figure 2 – The Figure demonstrates the evolution of the temporal shared mental model taking place over eight time stamps t_1 to t_8. The edges between nodes represent interpersonal dynamics, annotated with alignment valence, reflecting the affective tone of interactions between team members, ranging from -1 to +1. A positive valence indicates trust and agreement, while a negative valence indicates conflict and misalignment. The edges are assigned a weight E[-1,1] computed from equation 2.

A simulation is conducted on the data that was recorded, as shown in Figure 2. The equation 2 is used to calculate the alignment valence between team members as edge weight. The weights w_S and w_A are set to 1 and 2, respectively, to emphasise the more decisive influence of the alignment state on alignment valence. The values of the social role range from 1 to 5, where S_{SUP} is 5 since the social role of supporter signifies maximum alignment, S_G and S_N are 3 since both social roles gatekeeper and neutral, provide neutral alignment, and S_A and S_P are 1 since both social roles protagonist and attacker provide identical misalignment. The simulation, as shown in Figure 2, demonstrates the evolution of the temporal shared mental model over eight timestamps from t_1 to t_8. The Figure 2 (a) shows the evolution of team dynamics from timestamp t_1 to t_3, (b) shows the evolution from t_4 to t_6 and (c) and (d) show the specific evolution of the relationship between member A and B between timestamps t_7 and t_8. As shown in the figure, the graph updates in real time as two parameters, S (social role) and E (agent's alignment state), evolve. The dark green arrows signal a high alignment between members, the light green arrow signals neutrality, the orange arrows signal slight misalignment, and the red arrows signal high misalignment.

Graph-level outputs at timestamp t_k can be used to compute a team cohesion index, generate trust calibration signals, or trigger adaptive teaming strategies. By providing a deeper understanding of these dynamics, the system could lay the foundation for sustaining or directing the team toward positive interactions. Embedding alignment valence could enable adaptive, affect-aware teaming beyond static role assignment.

5 Conclusion

In this study, we present a conceptual temporal shared mental model in Human-Agent Teaming Systems (HATs) that leverages verbal and non-verbal behavioural data to capture evolving team dynamics and interrelationships among team members, thereby enabling real-time, active collaboration. The proposed model bridges the current research gap in HATs and SMMs by offering a graph-based representation of agentic teaming dynamics. In a small-scale proof-of-concept user study, a simulation of a team meeting among four members is conducted, and the evolution of team dynamics and interpersonal relationships is shown in a temporal graph that adapts to changes in the simulation parameters of the shared mental model.

6 Acknowledgement

This work was carried out within the projects "AI Co-Working Lab" and "Resilient human-robot collaboration in a mixed-skill environment (ENABLING)", financed by the European Regional Development Fund (ERDF), Saxony-Anhalt.

References

[1] Nowak, A., P. Lukowicz, and P. Horodecki: *Assessing artificial intelligence for humanity: Will ai be the our biggest ever advance? or the biggest threat [opinion]. IEEE Technology and Society Magazine*, 37(4), pp. 26–34, 2018.

[2] Iftikhar, R., Y.-T. Chiu, M. S. Khan, and C. Caudwell: *Human–agent team dynamics: A review and future research opportunities. IEEE Transactions on Engineering Management*, 71, pp. 10139–10154, 2023.

[3] Schmager, S., I. O. Pappas, and P. Vassilakopoulou: *Understanding human-centred ai: a review of its defining elements and a research agenda. Behaviour & Information Technology*, pp. 1–40, 2025.

[4] Andrews, R. W., J. M. Lilly, D. Srivastava, and K. M. Feigh: *The role of shared mental models in human-ai teams: a theoretical review. Theoretical Issues in Ergonomics Science*, 24(2), pp. 129–175, 2023.

[5] Cannon-Bowers, J. A., E. Salas, and S. Converse: *Shared mental models in expert team decision making. Individual and group decision making: Current issues*, 221, pp. 221–46, 1993.

[6] JONKER, C. M., M. B. VAN RIEMSDIJK, and B. VERMEULEN: *Shared mental models: A conceptual analysis*. In *International workshop on coordination, organizations, institutions, and norms in agent systems*, pp. 132–151. Springer, 2010.

[7] MATHIEU JOHN, E., F. GOODWIN GERALD, S. HEFFNER TONIA, and C.-B. J. A. SALAS EDUARDO: *The influence of shared mental models on team process and performance*. *Journal of Applied psychology*, 85(2), pp. 273–283, 2000.

[8] ROSSI, E., B. CHAMBERLAIN, F. FRASCA, D. EYNARD, F. MONTI, and M. BRONSTEIN: *Temporal graph networks for deep learning on dynamic graphs*. *arXiv preprint arXiv:2006.10637*, 2020.

[9] SCHELBLE, B. G., C. FLATHMANN, N. J. MCNEESE, G. FREEMAN, and R. MALLICK: *Let's think together! assessing shared mental models, performance, and trust in human-agent teams*. *Proceedings of the ACM on Human-Computer Interaction*, 6(GROUP), pp. 1–29, 2022.

[10] SCHEUTZ, M., S. A. DELOACH, and J. A. ADAMS: *A framework for developing and using shared mental models in human-agent teams*. *Journal of Cognitive Engineering and Decision Making*, 11(3), pp. 203–224, 2017.

[11] WYNNE, K. T. and J. B. LYONS: *An integrative model of autonomous agent teammate-likeness*. *Theoretical Issues in Ergonomics Science*, 19(3), pp. 353–374, 2018.

[12] KAUR, H., A. WILLIAMS, and W. LASECKI: *Building shared mental models between humans and ai for effective collaboration*. *CHI'19*, 2019.

[13] NARAYANAN, R. and K. FEIGH: *Influence of human-ai team structuring on shared mental models for collaborative decision making*. In *Proceedings of Workshop on Theory of Mind in Human-AI Interaction at CHI*, vol. 2024. 2024.

[14] SAPRU, A. and H. BOURLARD: *Automatic recognition of emergent social roles in small group interactions*. *IEEE Transactions on Multimedia*, 17(5), pp. 746–760, 2015.

[15] CARLETTA, J., S. ASHBY, S. BOURBAN, M. FLYNN, M. GUILLEMOT, T. HAIN, J. KADLEC, V. KARAISKOS, W. KRAAIJ, M. KRONENTHAL ET AL.: *The ami meeting corpus: A pre-announcement*. In *International workshop on machine learning for multimodal interaction*, pp. 28–39. Springer, 2005.

[16] THAGARD, P.: *Hot thought: Mechanisms and applications of emotional cognition*. MIT press, 2008.