

HEMI Tutorial at ICMI'26 Bibliography

1. Multimodal Social AI

- Vinciarelli, Alessandro, Maja Pantic, and Hervé Bourlard. "Social signal processing: Survey of an emerging domain." *Image and vision computing* (2009)
 - Link: <https://www.sciencedirect.com/science/article/abs/pii/S0262885608002485>
- Richmond, Virginia P., and James C. McCroskey. *Nonverbal Behavior in Interpersonal Relations*. Pearson/A and B, 2004.
 - Link: https://books.google.it/books/about/Nonverbal_Behavior_in_Interpersonal_Relationships.html?id=OnvsAAAAMAAJ&redir_esc=y
- Hintze, Arend. "The Four Types of AI: What You Need to Know." *World Economic Forum*, 15 Nov. 2016.
 - Link: <https://www.weforum.org/stories/emerging-technologies/the-four-types-of-ai-what-you-need-to-know/>
- Kakavá, Christina. *Negotiation of disagreement by Greeks in conversations and classroom discourse*. Georgetown University, 1993.
 - Link: <https://www.proquest.com/openview/029456cfa6e403399cdfb3b43fdca780/1?pq-origsite=gscholar&cbl=18750&diss=y>
- Poggi, Isabella, Francesca D'Errico, and Laura Vincze. "Agreement and its multimodal communication in debates: A qualitative analysis." *Cognitive Computation* (2011)
 - Link: <https://link.springer.com/article/10.1007/s12559-010-9068-x>
- Fringi, Paletz, Vinciarelli, "Speech-Based Analysis of Group Interactions", in "Computational Group and Team Dynamics: Forging an Interdisciplinary Science", Kozlowski, Hung, Lehmann-Willenbrock, Salah (eds.), to appear.
- Vinciarelli, Alessandro, Hugues Salamin, and Anna Polychroniou. "Negotiating over mobile phones: calling or being called can make the difference." *Cognitive Computation* 6.4 (2014): 677-688.
 - Link: <https://link.springer.com/article/10.1007/s12559-014-9267-y>
- Angouri, Jo, and Miriam A. Locher. "Theorising disagreement." *Journal of pragmatics* 44.12 (2012): 1549-1553.
 - Link: <https://www.sciencedirect.com/science/article/abs/pii/S0378216612001518>
- Buker, Areej, et al. "Multimodal Analysis of Disagreement in Dyadic Conversations: An Approach Based on Emotion Recognition." *ICMI* (2025).
 - Link: <https://dl.acm.org/doi/10.1145/3716553.3750754>
- Cohen, "A computerized scale for monitoring levels of agreement during a conversation", Technical Report University of Pennsylvania Working Papers in Linguistics, 2003.

- Link: https://www.researchgate.net/publication/304087399_A_computerized_scale_for_monitoring_levels_of_agreement_during_a_conversation
- Leslie, “How to Disagree: Lessons on Productive Conflict at Work and Home”, Faber & Faber, 2022.
 - Link: https://books.google.it/books/about/How_to_Disagree.html?id=OZI8EAAAQBAJ&redir_esc=y
- Walkinshaw, “Agreement and Disagreement”, The International Encyclopedia of Language and Social Interaction, 2015.
 - Link: https://www.researchgate.net/publication/282729917_Agreement_and_Disagreement
- Rosenthal, “Conducting Judgment Studies: Some Methodological Issues”, in “The New Handbook of Methods in Nonverbal Behavior Research”, Harrigan, Rosenthal and Scherer (eds.), 2008.
 - Link: <https://academic.oup.com/book/25991/chapter-abstract/193835959?redirectedFrom=fulltext>

2. Personalized human-agent interaction with MLLMs

- Alaluf, Yuval, et al. "Myvlm: Personalizing vlms for user-specific queries." ECCV, 2024.
 - Link: <https://arxiv.org/abs/2403.14599>
- Nguyen, Thao, et al. "Yo'llava: Your personalized language and vision assistant." NeurIPS (2024).
 - Link: <https://arxiv.org/abs/2406.09400>
- Hao, Haoran, et al. "Rap: Retrieval-augmented personalization for multimodal large language models." CVPR, 2025.
 - Link: <https://arxiv.org/abs/2410.13360>
- Das, Deepayan, et al. "Training-free personalization via retrieval and reasoning on fingerprints." ICCV, 2025.
 - Link: <https://arxiv.org/abs/2503.18623>
- Das, Deepayan, et al. "Personalizing MLLMs via Reinforced Multimodal Reference Game." ECCV (2026).
 - Link: <https://arxiv.org/abs/2606.28845>
- Zhao, Siyan, et al. "Do llms recognize your preferences? evaluating personalized preference following in llms." ICLR (2025).
 - Link: <https://arxiv.org/abs/2502.09597>
- Wang, Huaxiaoyue, et al. "APRICOT: Active preference learning and constraint-aware task planning with LLMs." CoRL (2024).

- Link: <https://arxiv.org/abs/2410.19656>

3. Embodied perception and reasoning in human-centric environments

- Yang, Jihan, et al. "Thinking in space: How multimodal large language models see, remember, and recall spaces." CVPR (2025).
 - Link: <https://arxiv.org/abs/2412.14171>
- Yang, Rui, et al. "Embodiedbench: Comprehensive benchmarking multi-modal large language models for vision-driven embodied agents." ICML (2025).
 - Link: <https://arxiv.org/abs/2502.09560>
- Nasiriany, Soroush, et al. "Robocasa365: A large-scale simulation framework for training and benchmarking generalist robots." ICLR (2026)
 - Link: <https://arxiv.org/abs/2603.04356>.
- Rasheed, Hanoona, et al. "Hard Vision, Easy Vision: What GPT-6 Astra Reveals Across Computer Vision." arXiv preprint arXiv:2609.35718 (2026).
 - Link: <https://arxiv.org/abs/2609.35718>
- Huh, Gio, et al. "HomeBody: A Humanoid That Explores, Remembers, and Acts on Its Own." 2026,
 - Link: <https://tml.stanford.edu/homebody/>
- Taioli, Francesco, et al. "Collaborative instance object navigation: Leveraging uncertainty-awareness to minimize human-agent dialogues." ICCV (2025)
 - Link: <https://arxiv.org/abs/2412.01250>
- Jiao, Runyu, et al. "Obstruction reasoning for robotic grasping." CVPR (2026)
 - Link: <https://arxiv.org/abs/2511.23186>
- Gilles, Maximilian, et al. "Metagraspnetv2: All-in-one dataset enabling fast and reliable robotic bin picking via object relationship reasoning and dexterous grasping." IEEE Transactions on Automation Science and Engineering 21.3 (2023)
 - Link: <https://ieeexplore.ieee.org/document/10309974>
- Yang, Jianwei, et al. "Set-of-mark prompting unleashes extraordinary visual grounding in gpt-4v." arXiv preprint arXiv:2310.11441 (2023).
 - Link: <https://arxiv.org/abs/2310.11441>
- Team, Gemini Robotics, et al. "Gemini robotics: Bringing ai into the physical world." arXiv preprint arXiv:2503.20020 (2025).
 - Link: <https://arxiv.org/abs/2503.20020>
- Bai, Shuai, et al. "Qwen2.5-vl technical report." arXiv preprint arXiv:2511.21631 (2025).
 - Link: <https://arxiv.org/abs/2502.13923>