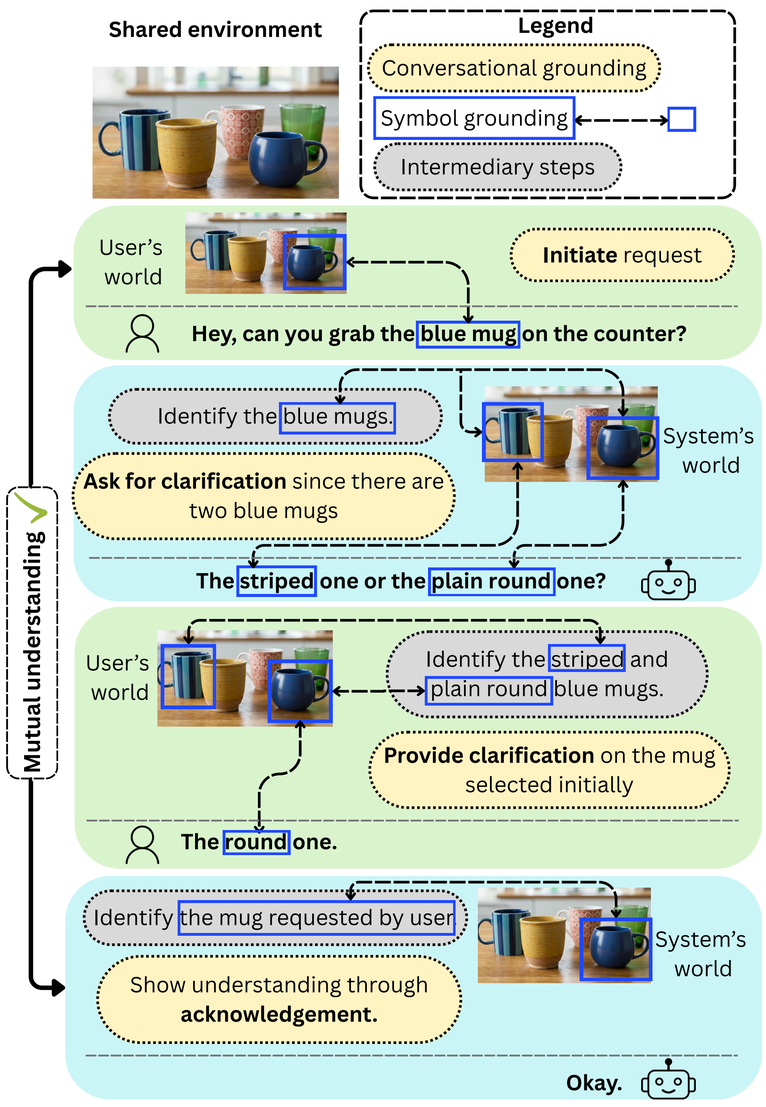
Conversational Grounding in LLMs: Evaluation Methods, Challenges & Future Directions

Michelle Elizabeth[1,2], Gwénolé Lecorvé[1], Lina M. Rojas-Barahona[1], Magalie Ochs[2]
[1]Orange Research, Lannion, France | [2]LIS, Aix-Marseille Université, France
Contact: michelle.elizabeth@orange.com

Symbol vs. Conversational Grounding

Symbol Grounding

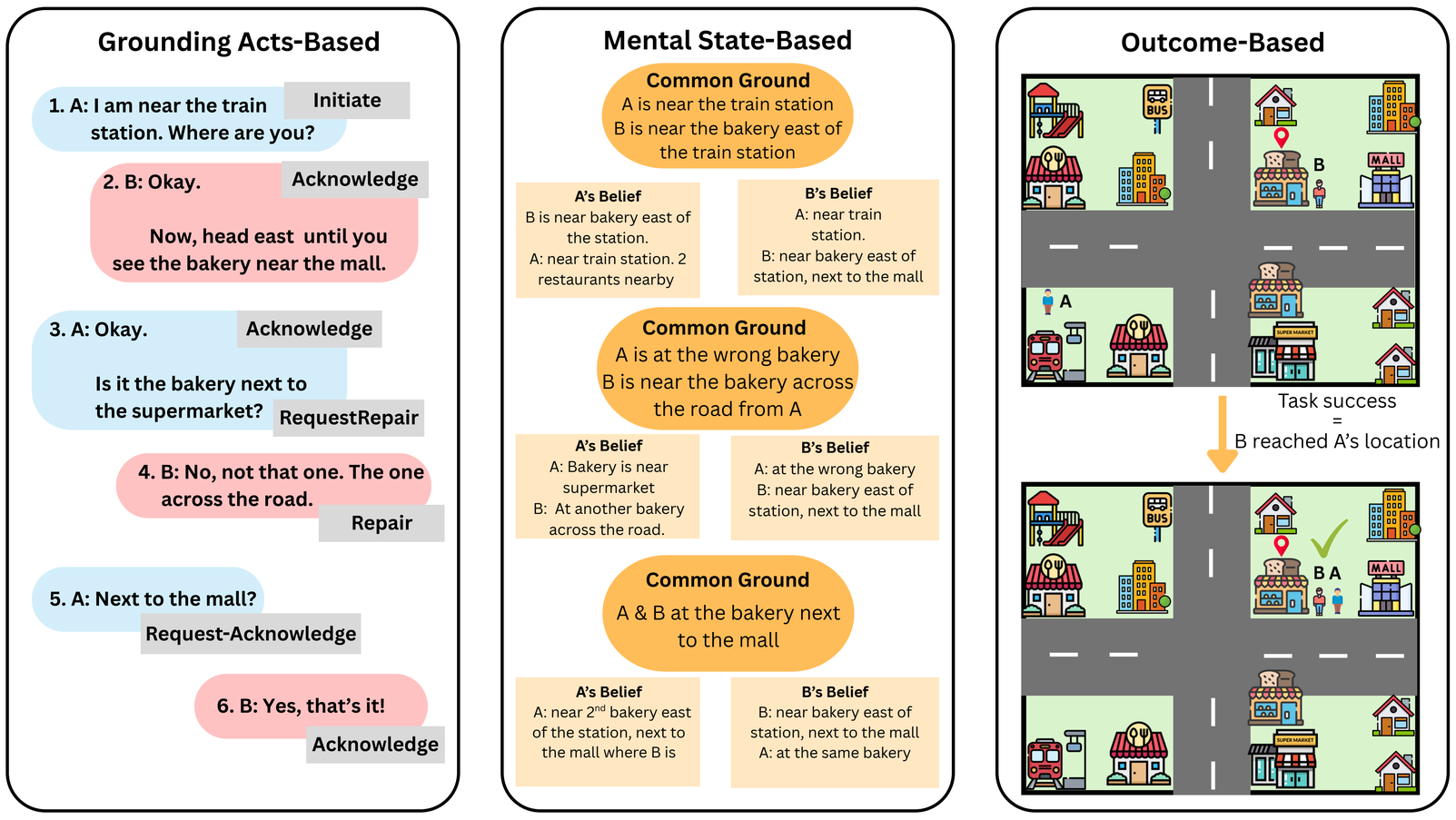
- **Link between language and entities** in the participant's world.
- Entities can be **physical objects, images, concepts etc**.



Conversational Grounding

- The **collaborative process** of building **mutual understanding** or **common ground** [1].
- **Contribution** to conversation made through **presentation** and **acceptance** phases.
- Acceptance produces **evidence of grounding** (e.g. acknowledgement, clarification, etc.).

Modelling Conversational Grounding



- **Dialogue acts** that indicate grounding [2].
- **A priori deciding if and which grounding act (GA) should be used.**
- **Intersection of participant mental states** as common ground.
- **Estimated mental state decides whether to perform grounding.**
- **Collaborative tasks** to study **grounding at conversation level** [3].
- **Reach an agreement to complete the task.**

Evaluating Conversational Grounding

- **Presence of GAs** as an **a posteriori signal** (e.g. [4]).
- **GA usage** by LLMs; **Detection of GAs** by LLMs.
- **Methods**: Zero/Few-shot prompting, finetuning, perplexity.
- **Metrics**: Precision, recall, F1, accuracy, inter-annotator agreement.
- How close **predicted belief state** is to **optimal common ground** [5].
- **Methods**: Belief modelling and evidence-based approaches.
- **Metrics**: Joint goal accuracy, F1, cosine similarity, Sorenson-Dice coefficient.
- **Task success as proxy** for achieving common ground.
- **Navigation** - same location; **alignment** - same entity; **negotiation** - optimal score.
- **Human evaluation** dimensions like satisfaction, cooperation, fluency etc.

Research Avenues

Grounding capabilities of LLMs

- Incorporate strategies for grounding during training.
- Improve grounding in black-box LLMs.
- Grounding behaviour for efficient task completion.

Evaluation of grounding

- Cues of grounding and their interaction.
- Evolution of the common ground over time.
- Costs and utility of performing grounding.

Correlation of grounding with other dimensions

- Impact on task success rate and user satisfaction.
- Impact on other dimensions like trust, skills etc.

Limitations of Current Methods

- Works define **new GAs or use subsets**.
- **Important GAs under-represented** in datasets.
- **Limited performance** in detecting GAs.
- **Low agreement** with humans on the usage of GAs.
- Relies on common ground of a **single conversation path** for evaluation.
- System's **ability to recover from mistakes** not considered.
- **Error propagation** due to limited performance of event and belief extraction.
- **Task success** alone can be an **ambiguous signal** for grounding.
- Does not consider **how grounding evolved** throughout the conversation.
- **Efficiency of task completion** not considered.

[1] Herbert H. Clark and Susan Brennan. Grounding in communication. In *Perspectives on socially shared cognition*, 1991.

[2] David R. Traum and James F. Allen. A "speech acts" approach to grounding in conversation. In *ICSLP*, 1992.

[3] David Schlangen. Grounded agreement games: Emphasizing conversational grounding in visual dialogue settings. *CoRR*, 2019.

[4] Biswesh Mohapatra, Manav Nitin Kapadnis, Laurent Romary, and Justine Cassell. Evaluating the effectiveness of large language models in establishing conversational grounding. In *EMNLP 2024*.

[5] Magdalena Markowska, Mohammad Taghizadeh, Adil Soubki, Seyed Mirroshandel, and Owen Rambow. Finding common ground: Annotating and predicting common ground in spoken conversations. In *EMNLP 2023*.

Presented at the 27th Meeting of the Special Interest Group on Discourse & Dialogue (SIGDIAL'26), 2-5 August 2026, Atlanta, Georgia, USA