Computational Models of Conversational Grounding

ESSLLI European Summer School for Logic, Language, and Information
August 10-14, 2026

Lecture 3: Generative LLMs and Grounding

David Traum

Institute for Creative Technologies
Department of Computer Science
University of Southern California, USA

Kristiina Jokinen

University of Helsinki, Finland
CNRS-AIST Joint Research Lab
Institute of Science Tokyo, Japan

Agenda

Lecture	Date	Topics	Lecturer
Day 1	Monday August 10	Participant Introduction Introduction to Grounding and Interaction	Traum / Jokinen
Day 2	Tuesday August 11	Grounding Acts	Traum
Day 3	Wednesday August 12	Generative LLMs and Grounding	Jokinen / Traum
Day 4	Thursday August 13	Multimodality	Traum / Jokinen
Day 5	Friday August 14	Challenges and Future Topics	Jokinen / Traum

Outline of Course (covered today)

- Preliminaries: representation, agency, communication, definitions & uses for common ground
- Talking with Artificial Agents
- Clark & Colleagues Grounding models
- Feedback and Error-handling in Spoken Dialogue Systems
- Speech Acts and Dialogue Acts
- Multi-functionality of Utterances
- Grounding Acts & Early Computational Models of Grounding
- Degrees of Grounding

- Grounding and LLMs
  - Issues
  - Solutions
- Multi-modal Grounding
- Multiparty Multilingual & Multi-floor Grounding
- Incremental Grounding
- Use of grounding for other phenomena

AIST
Institute of
SCIENCE TOKYO

Review Of Yesterday

Types of Feedback (Allwood et al 92)

- Levels:
  - Contact
  - Perception
  - Understanding
  - Attitudinal Reaction
- Signals types
  - Request feedback
  - Prepare other
  - Provide
    - Positive
    - negative

AIST
Institute of
SCIENCE TOKYO

Levels of Dialogue acts: Traum & Hinkelman 1992

Discourse Level	Act Type	Sample Acts
Sub UU	**Turn-taking**	take-turn, keep-turn, release-turn, assign-turn
UU	**Grounding**	Initiate, Continue, Ack, Repair, ReqRepair, ReqAck, Cancel
DU	**Core Speech Acts**	Inform, YNQ, Check, Eval Suggest, Request, Accept, Reject,
Multiple DUs	**Argumentation**	Elaborate, Summarize, Clarify Q&A Convince Find-Plan

AIST
Institute of SCIENCE TOKYO

Grounding Acts

Label	Description
initiate	Begin new DU, content separate from previous uncompleted DUs
continue	same agent adds related content to open DU
acknowledge	Demonstrate or claim understanding of previous material by other agent
repair	Correct (potential) misunderstanding of DU content
Request Repair	Signal lack of understanding
Request Ack	Signal for other to acknowledge
cancel	Stop work on DU, leaving it ungrounded and ungroundable

Degrees of Groundedness

- Unknown - material has not yet been introduced
- Misunderstood - after a Request Repair
- Unacknowledged - after a Submit, Lack of Response
- Accessible - after a Submit or Resubmit
- Agreed-Signal - after a Submit, Acknowledgment
- Agreed-Signal+ - after a Submit, Acknowledgments, other
- Agreed-Content - after a Submit + Repeat Back
- Agreed-Content+ - after a Submit + Repeat Back + Acknowledgment(s) / other
- Assumed - both participants already know material

AIST
Institute of SCIENCE TOKYO

Recognizing Grounding Acts with LLMs

- Conversational Grounding: Annotation and Analysis of Grounding Acts and Grounding Units
  - Biswesh Mohapatra, Seemab Hassan, Laurent Romary, Justine Cassell
  - COLING 2024

New Annotations of Grounding Acts on new corpora

Figure 2: Annotation of meetup dataset

Utterance	Grounding Unit ID	Grounding Act	Currently Open CGUs	CGUs closed by current utterance	Degree of Grounding
B: hello	-	None	-	-	-
A: where are you now?	CGU1	Initiate	CGU1	-	-
B: i'm at a staircase, but there's no plant	CGU1, CGU2	Use, Initiate	CGU2	CGU1	Medium
A: hmm ... shall I come and look for you?	CGU2, CGU3	Move, Initiate	CGU3	CGU2	Low
B: Stay there :)	CGU3, CGU4	Use, Initiate	CGU4	CGU3	Medium
A: where?	CGU4	Request-Repair	CGU4	-	-
B: stairs with no plants	CGU4	Repair	CGU4	-	-
A: okay	CGU4	Explicit-Acknowledgment	-	CGU4	Medium

AIST
Institute of SCIENCE TOKYO

Automated Grounding Act Annotation

Table 3: T5 performance on Meetup dataset

GA	Accuracy w/o weighted loss	Accuracy with weighted loss
Use:	0.66	0.74
Move:	0.68	0.70
Req-Ack	0.00	0.00
Req-Repair	0.00	0.00
Repair	0.00	0.57
Initiate	0.99	0.99
Repeat	0.00	0.00
Explicit-Ack	0.62	0.11
Continue	0.00	0.05
Repeat-Back	0.00	0.00
Cancel	0.00	0.00

Table 4: T5 performance on Spot the Difference dataset

GA	Accuracy w/o weighted loss	Accuracy with weighted loss
Use:	0.00	0.01
Move:	0.00	0.00
Req-Ack	0.00	0.00
Req-Repair	0.00	0.00
Repair	0.00	0.00
Initiate	0.23	0.31
Repeat	0.00	0.00
Explicit-Ack	0.35	0.86
Continue	0.15	0.24
Repeat-Back	0.00	0.00
Cancel	0.00	0.00

Grounding with People and Machines (the office)

Day 3 Wednesday August 12

Q: Can LLMs Manage Grounding?

A1: No! LLMs do not manage Knowledge Grounding or Conversational Grounding at all.

- A2: Yes! Knowledge Graphs and Grounding
  Some possibilities to explore more
  - Agentic Architectures for Grounding
  - Internal mechanisms to enable Grounding
  - Expanding horizon to fresh views of what is language interaction

LLMs go wrong in Knowledge Grounding and Conversational Grounding

Wilcock & Jokinen: *To err is robotic; to earn trust, divine: comparing ChatGPT and knowledge graphs*. IEEE RO-MAN 2023

In HRI community, one of the first papers to draw attention to LLMs' lack of knowledge of the real world, they fail to

- Provide reliable information:
  - e.g. a **list of restaurants** in Tokyo Waterfront refers to restaurants in San Francisco
- Understand the partner's affective states due to lack experience:
  - e.g. can **talk about emotions** or recommendations for a happy life, but if asked to **recognize or produce** a happy face, it does not know how to do it
- Understand dialogue context:
  - e.g. ***goodbye*** often used to stop the dialogue, but also to express condolences
- Produce consistent dialogues:
  - e.g. "I'm **not very interested** in training" (- *sorry to hear that*), followed by "**I'm interested** in training" (- *great let's start!*)
- Understand the partner's goal and intent**:**
  - e.g. a robot should be able to **take action according to the external situation**, **besides being able to chat about it**

Language is reflexive: first order vs second order representation!

- **LLMs** are **second-order representations** of the world: they model **how we talk about** the world, **not how the world is structured**
- **KGs** can be said to model **first-order representations** of the world: they model the entities, relations and actions that make up the world

Trade-off in Grounding

Elizabeth. M., Lecorvé, G., Rojas-Barahona, L., Ochs, M. (2026). Conversational Grounding in Large Language Models: Evaluation Methods, Challenges and Future Directions. Sigdial 2026.

- **Maximize the mutual understanding**

  and

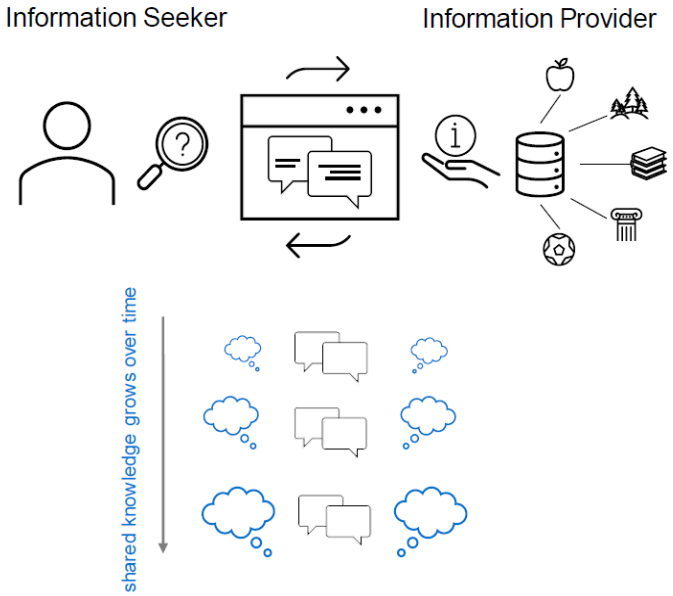
  Minimize the costs of the grounding processes

- E.g. correct a mistake as early as possible in conversation when it has less temporal costs than if correcting later in the conversation

- Good grounding behaviour should take into consideration the **utility** and **costs** of executing the grounding action for the task at hand

Towards Harnessing Large Language Models for Comprehension of Conversational Grounding

Phillip Schneider et al. IWSDS 2024, SIGDial 2024



Pipeline Approach

1. Detect grounding in dialogue acts
2. Retrieve relevant subset from knowledge base
3. Integrate knowledge while ensuring consistency
4. Generate appropriate grounding response

Dialogue corpus:

- **26 information-seeking conversations** with **669 turns** (in English)
- based on **tabular datasets**
- **5 domains** (geography, history, media, nutrition, and sports)
- text-based chatroom

Annotation:

- Every dialogue was **annotated by two researchers**.
- First, labels for the three different **grounding acts (explicit, implicit, clarification)** were assigned.

  → **147 annotations**
- For **explicit and implicit labels**, grounded knowledge items
  until this point in the dialogue were annotated as a **knowledge graph structure in JSON-LD format**.

  → **127 annotations**
- **Disagreements** were resolved until **consensus** was reached.

Example Prompts

Classification Few-Shot Template

Predict the grounding label, representing when knowledge has been mutually grounded, for the last turn in the 'Input dialogue:'. The label can be 'explicit' if knowledge is verbally accepted, 'implicit' if accepted by moving forward with the conversation, or 'clarification' if a previous utterance must be clarified before acceptance.

USER: Input dialogue:
seeker: Can you tell me about the dataset's content?
provider: The dataset contains information about planets in our solar system.
seeker: What is the number of columns in the dataset?

ASSISTANT: Output label: implicit
...

Information Extraction Few-Shot Template

Predict the newly grounded knowledge for the last turn in the 'Input dialogue:'. Use the JSON structure: {'table domain': str, 'table content': str, 'row count': int, 'column count': int, 'column info': [{'column name': str, 'values': [], 'distinct count': int, 'min value': int, 'max value': int}]}. Adhere strictly to the JSON structure, and only predict the attributes mentioned in the dialogue turns, leaving unmentioned attributes as null.

USER: Input dialogue:
seeker: Can you tell me about the dataset's content?
provider: The dataset contains information about planets in our solar system.
seeker: What is the number of columns in the dataset?

ASSISTANT: Output JSON: {'table content': 'planets of the solar system'}
...

Grounding Act Classification

Model	Zero-Shot Prompt				Few-Shot Prompt			
	Accuracy	Precision	Recall	F1-Score	Accuracy	Precision	Recall	F1-Score
GPT-3.5-Turbo (n=1)	0.64	0.50	0.46	0.43	0.55	0.50	0.51	0.50
GPT-3.5-Turbo (n=3)	0.66	**0.81**	0.50	0.50	0.69	0.59	0.54	0.54
GPT-3.5-Turbo (n=all)	0.59	0.39	0.44	0.41	0.57	0.51	0.45	0.45
GPT-4o (n=1)	0.39	0.55	0.54	0.42	0.64	0.66	0.64	0.61
GPT-4o (n=3)	0.59	0.66	**0.67**	0.59	0.73	**0.74**	0.69	**0.70**
GPT-4o (n=all)	0.64	0.68	0.66	0.62	0.71	0.73	0.67	0.67
Llama-3-8B (n=1)	0.61	0.54	0.53	0.54	0.59	0.65	0.69	0.59
Llama-3-8B (n=3)	0.65	0.60	0.60	0.60	0.57	0.60	0.61	0.55
Llama-3-8B (n=all)	0.44	0.55	0.39	0.38	0.55	0.54	0.51	0.51
Llama-3-70B (n=1)	0.41	0.54	0.56	0.43	0.51	0.61	0.63	0.53
Llama-3-70B (n=3)	0.59	0.66	**0.67**	0.59	0.65	0.68	0.69	0.64
Llama-3-70B (n=all)	**0.71**	0.66	0.64	**0.64**	**0.76**	0.70	**0.70**	**0.70**

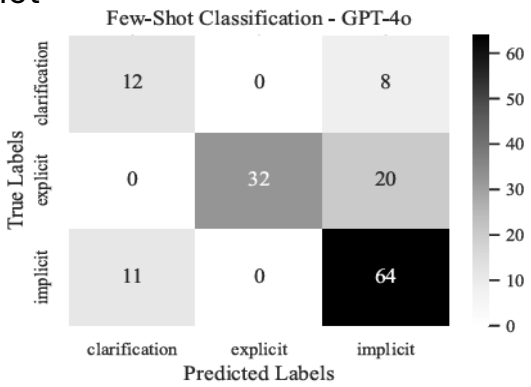
Table 1: Performance metrics using macro-averages to ensure equal class importance.

- Nearly all tested LLMs **benefit from few-shot examples** (F1-scores improving) but with diminishing returns as the provided turns (n) increase.
- n=3 often optimizes performance in both zero- and few-shot settings by balancing context retention, noise reduction, and efficient usage of tokens.
- Llama-8B's performance drops from 0.54 F1-score (n=1) to 0.38 (n=all) whereas **larger LLMs** like Llama-70B and GPT-4o **handle longer input better** (probably due to a higher parameter count and superior noise handling).
- Llama-8B surpasses GPT-3.5 in the zero-shot run, and Llama-70B matches GPT-4o in the few-shot run.
- Competitive performance of open-source LLMs against proprietary ones:
  Llama-8B surpasses GPT-3.5 in the zero-shot run, and Llama-70B matches GPT-4o in the few-shot run.

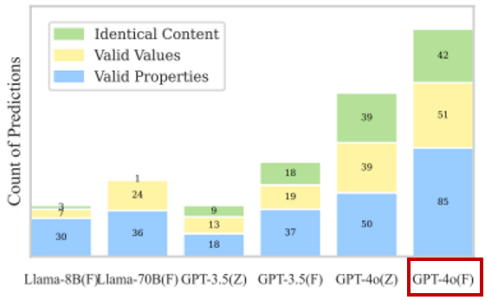
Conversational Grounding and LLMs

Grounded knowledge extraction

- GPT-4o tends to overpredict "implicit" labels when *acknowledgement* follows *new question*



- **Constructing grounded information**
  - More complex task => GPT outperforms
  - Properties, nodes, and syntax elements missing, or hallucinated



1. **Grounding type classification task, GPT-**3.5-Turbo

Explicit grounding was mostly correctly classified: it can be observed in the text in forms such as *OK* and *great*.

Implicit grounding and clarification were easily confused as both can involve questions and require contextual dialogue understanding.

Linguistic phenomena like co-reference and ellipsis might have added another level of complexity to classifying these grounding acts.

2. Grounded knowledge extraction task, GPT-3.5-Turbo demonstrated better overall performance.

The LLM accurately gathers the relevant information even though it mixes up the similar attributes

Handles numerical information, successfully determining the number of rows in a table or counts of unique values for specific columns

AIST
Institute of SCIENCE TOKYO

Issues in constructing grounded information

Issue Type	Definition	Example
Invalid JSON-LD	Prediction is overall not a valid JSON-LD structure.	PRED: [{"@type": "schema:CreativeWork", "category": ["films", "series", "literature"]}] → JSON syntax elements are missing.
Property Hallucination	Predicted property is not part of system knowledge.	PRED: {"@type": "schema:Place", "name": "English Garden", "state": "Bavaria", "city": "Munich", "summary": "The English Garden is a large public park in Munich, Bavaria. It is one of the world's largest urban public parks, larger than New York's Central Park. The park is known for its beer gardens, streams, lakes, and Japanese teahouse."} → English Garden was mentioned in the dialogue but is not part of the system knowledge. The "city" property does not exist.
Value Hallucination	Predicted value is not part of system knowledge.	PRED: [{"@context": ["http://www.w3.org/ns/csvw", {"schema": "http://schema.org"}], "@id": "http://example.org/time-travel-media", "url": "time-travel-media.csv", "schema:description": "The table contains information about time travel media", "tableSchema": {"columns": [{"name": "year", "datatype": "string"}, {"name": "title", "datatype": "string"}, {"name": "author", "datatype": "string"}, {"name": "short text description", "datatype": "string"}, {"name": "category", "datatype": "string"}]}, "primaryKey": ["year", "title", "category"]}] → System knowledge contains the columns "creators" and "description" but "author" and "short text description" are hallucinated.

Contributions

- a new dialogue dataset with over 250 annotations across five domains,
- a range of zero- and few-shot experiments by evaluating four LLMs on two grounding tasks,
- common prediction errors and prompting techniques for improving model performance.

GitHub Repository

github.com/philotron/Bridge-KG

Robot's Knowledge Sources

- Databases: restaurants, hotels, exercises, diseases, recommendations...
- Internet resources: webpages, Wikipedia,...
- Sensor information: audio-visual information, biosensors, biomechanics, motion trackers, environmental sensors (temperature, humidity),...

- Explicit vs implicit knowledge
  - Common knowledge:
    - Spanish food is a type of European food and restaurants that serve Spanish food serve European food;
    - Temperature affects well-being
  - Explaining knowledge:
    - What type of food is European cuisine?
    - Why does temperature affect my well-being?
- Long-term engagement and learning:
  - What the user knows,
  - What the user is expected to know
  - What is learnt in the interaction

Jokinen, Nishimura,S., Watanabe, Nishimura, T. (2018): Human-Robot Dialogues for Explaining Activities. IWSDS 2018.
Jokinen, Wilcock (2021). Do you remember me? Ethical Issues in Long-term Social Robot Interactions. RO-MAN 2021.

Knowledge Graphs

Formally a graph consists of nodes and edges, defined as

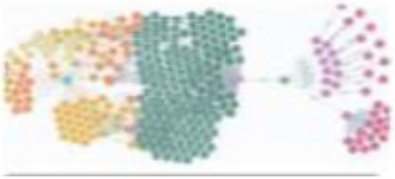
K:=(E,R)

where E is the set of entity **nodes**, and **edge** r ∈ R is a relation that connects a head entity e_h and a tail entity e_t

The 3-element tuple (e_h,r, e_t) is referred to as a triple.

Intuitive use in interaction

- Representation that provides explainability:
  - explicit definition of entities and relations,
  - semantic interpretation,
  - scalable descriptions of the domain
- Context-aware dialogue modelling
  - Rich description of the domain, to be used by the RASA dialogue system



Models in Neo4j graph database system

- Nodes (= entities)
- Links (= relations) that connect the nodes
- Describe the important domain concepts and relations in natural language
- Graph databases can be built in an *intuitive* way which is easily understandable by content experts

Ji, S., Pan, S., Cambria, E., Marttinen, P., Philip, S. Y.: A survey on knowledge graphs: Representation, acquisition, and applications. IEEE Transactions on Neural Networks and Learning Systems, IEEE (2021).

Robinson, I., Webber, J., Eifrem, E.: Graph Databases (2nd edition). O'Reilly Media, (2015).

Knowledge Graphs:

Store knowledge about the real world

Knowledge graph in Neo4j:

Search for restaurants in Aqua City Odaiba, Tokyo

(and some nearby hotels).

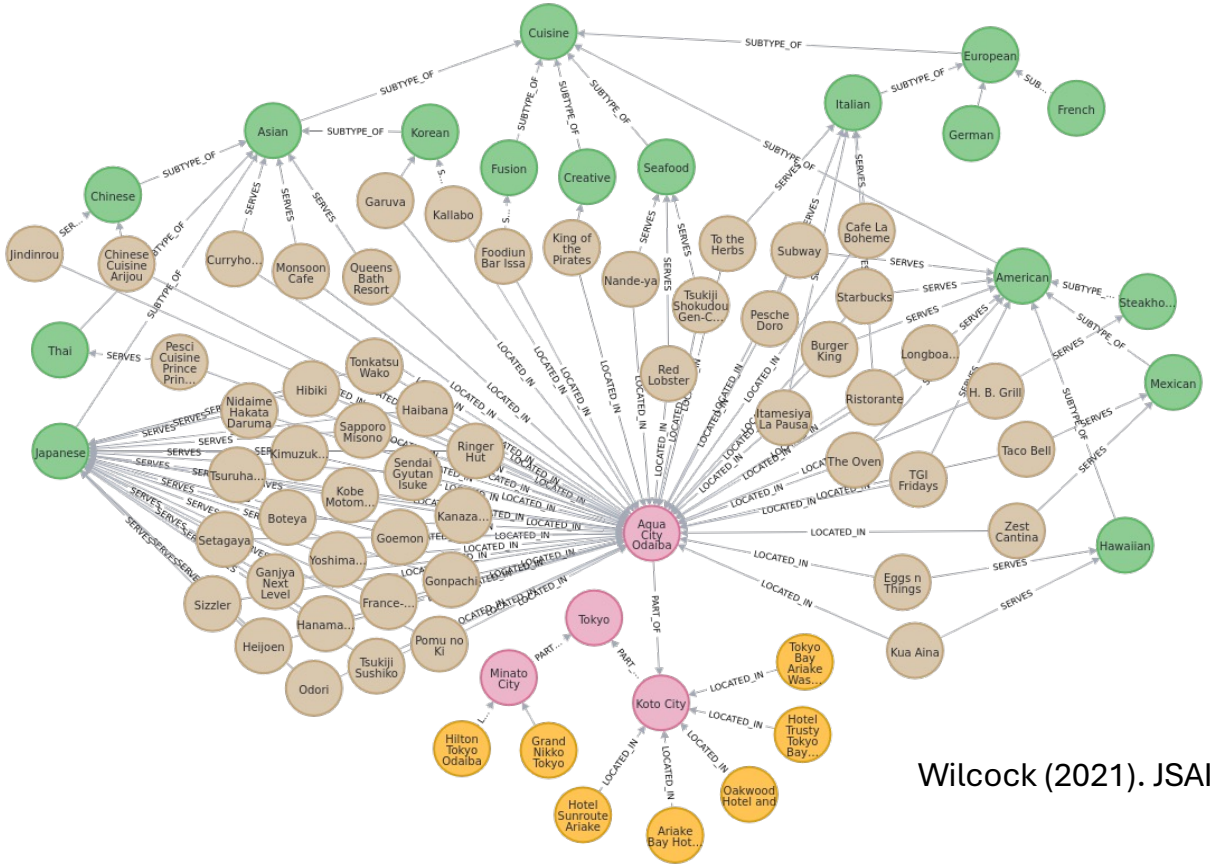
Cuisine SUBTYPE_OF Cuisine (Taxonomy)

Restaurant SERVES Cuisine

Restaurant LOCATED_IN Place (Aqua City Odaiba)

Place PART_OF Place (Meronymy)

Hotel LOCATED_IN Place

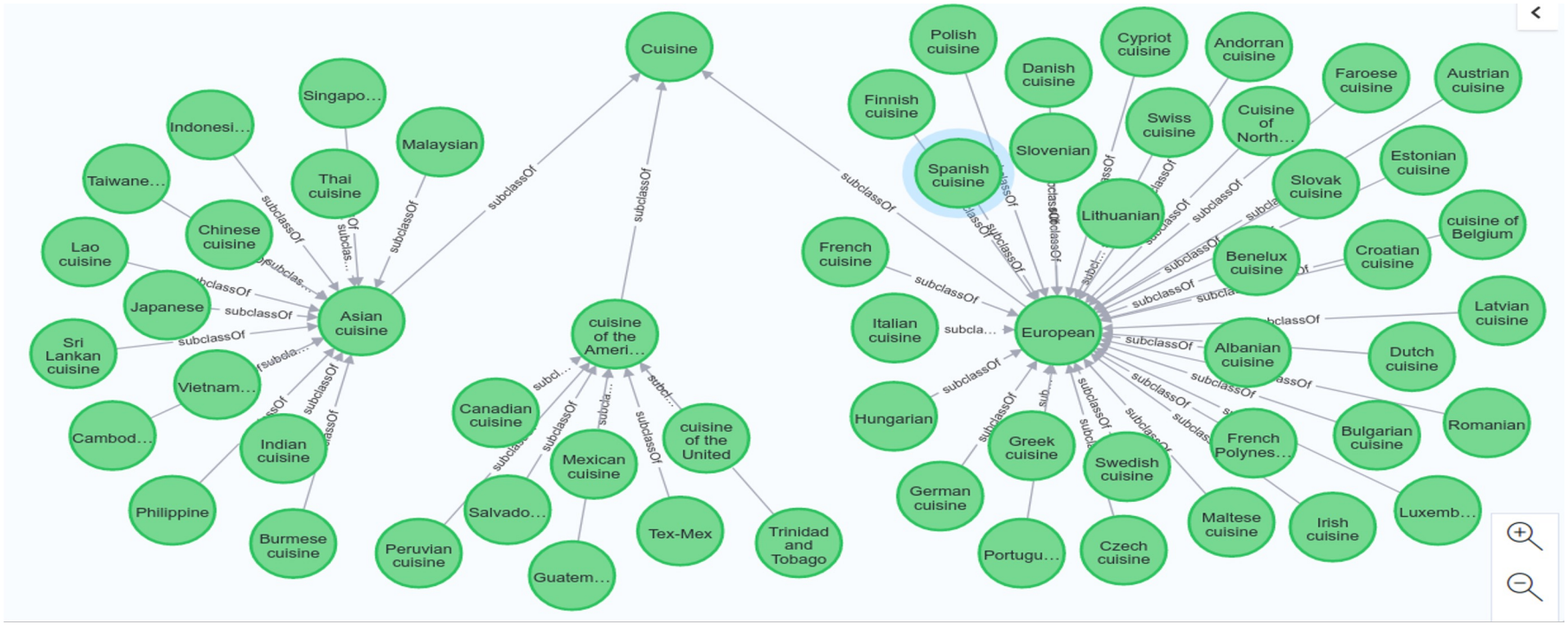


Wilcock (2021). JSAI

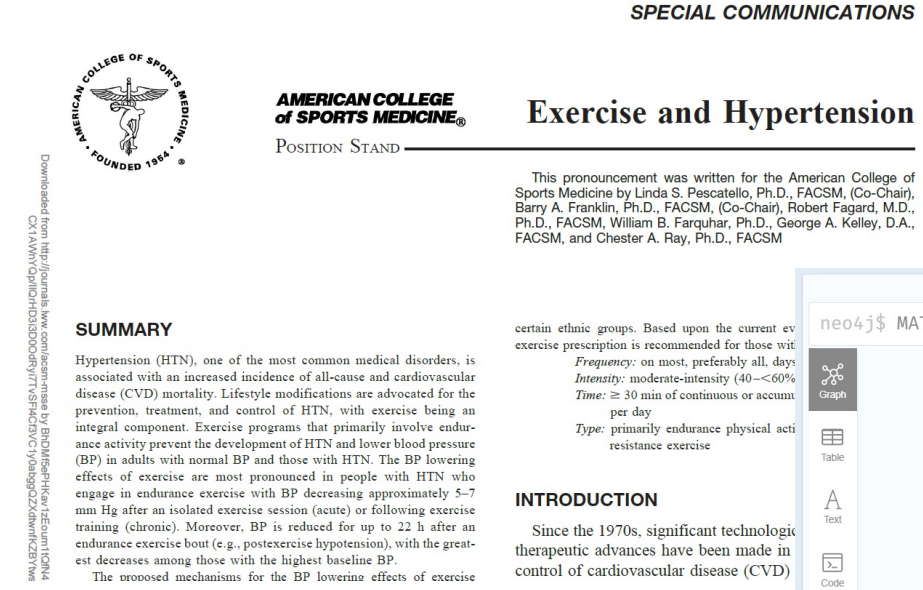
Knowledge Graphs:
Encode relations, not just facts

Taxonomy of cuisines from WikiData (subclass of)

Wilcock (2021). JSAI



A PDF document and Chunks (text+embeddings)



SPECIAL COMMUNICATIONS

AMERICAN COLLEGE of SPORTS MEDICINE®

POSITION STAND

Exercise and Hypertension

This pronouncement was written for the American College of Sports Medicine by Linda S. Pescatello, Ph.D., FACSM, (Co-Chair), Barry A. Franklin, Ph.D., FACSM, (Co-Chair), Robert Fagard, M.D., Ph.D., FACSM, William B. Farquhar, Ph.D., George A. Kelley, D.A., FACSM, and Chester A. Ray, Ph.D., FACSM

SUMMARY

Hypertension (HTN), one of the most common medical disorders, is associated with an increased incidence of all-cause and cardiovascular disease (CVD) mortality. Lifestyle modifications are advocated for the prevention, treatment, and control of HTN, with exercise being an integral component. Exercise programs that primarily involve endurance activity prevent the development of HTN and lower blood pressure (BP) in adults with normal BP and those with HTN. The BP lowering effects of exercise are most pronounced in people with HTN who engage in endurance exercise with BP decreasing approximately 5–7 mm Hg after an isolated exercise session (acute) or following exercise training (chronic). Moreover, BP is reduced for up to 22 h after an endurance exercise bout (e.g., postexercise hypotension), with the greatest decreases among those with the highest baseline BP.

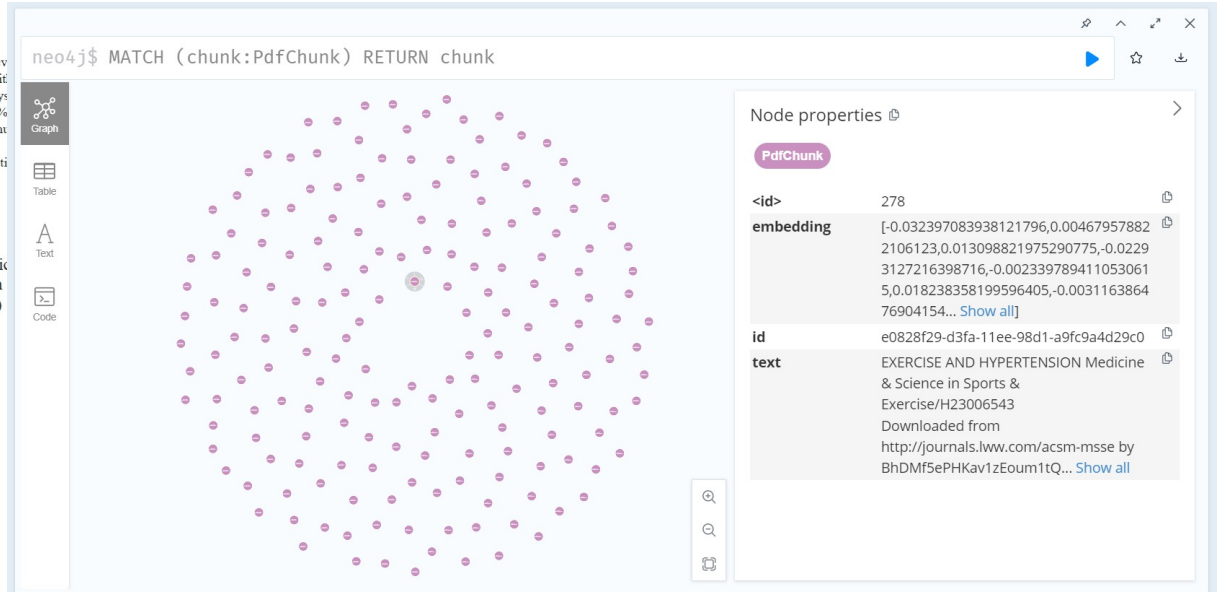
The proposed mechanisms for the BP lowering effects of exercise

certain ethnic groups. Based upon the current ev
exercise prescription is recommended for those wit

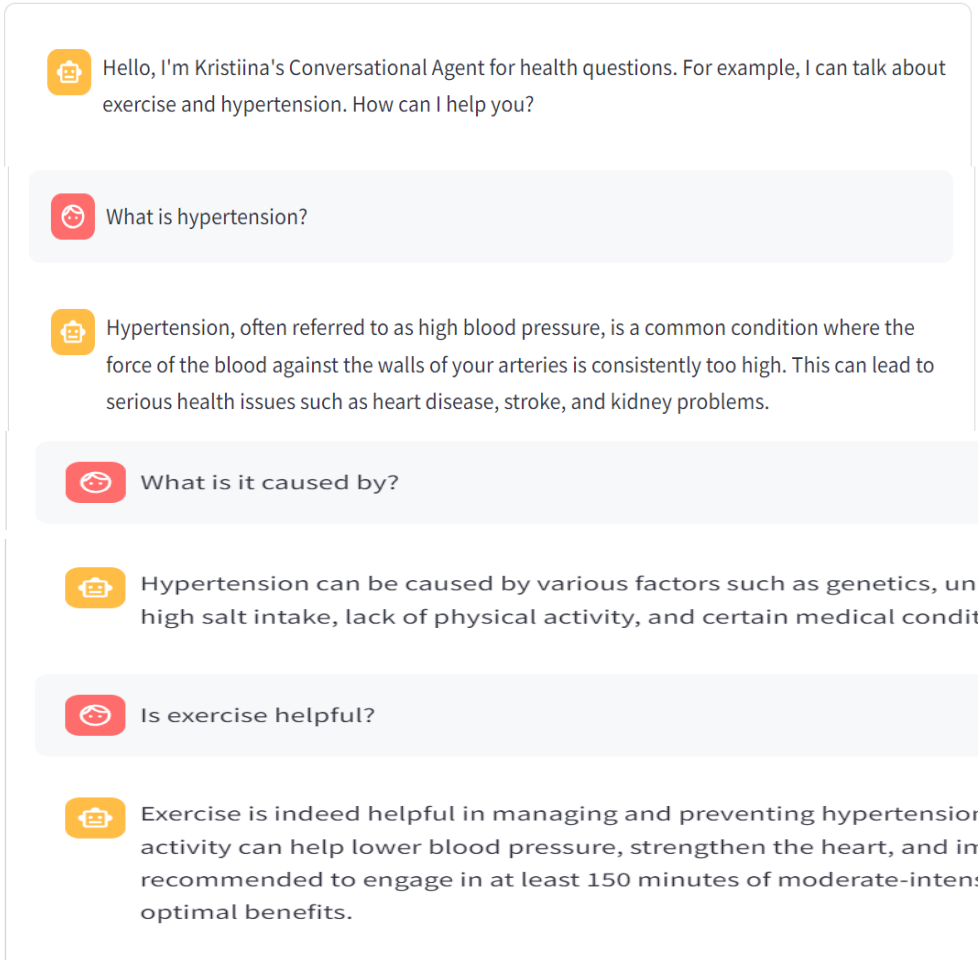
Frequency: on most, preferably all, days
Intensity: moderate-intensity (40–<60%
Time: ≥ 30 min of continuous or accumu
per day
Type: primarily endurance physical acti
resistance exercise

INTRODUCTION

Since the 1970s, significant technologic
therapeutic advances have been made in
control of cardiovascular disease (CVD)



RAG from PDF (Hypertension)



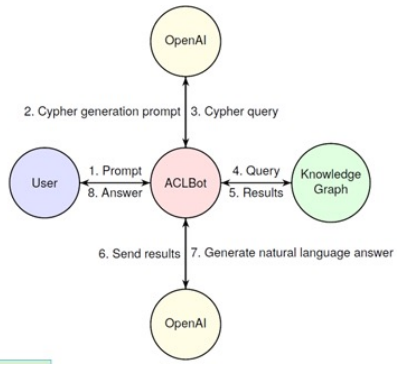
ACL Bot

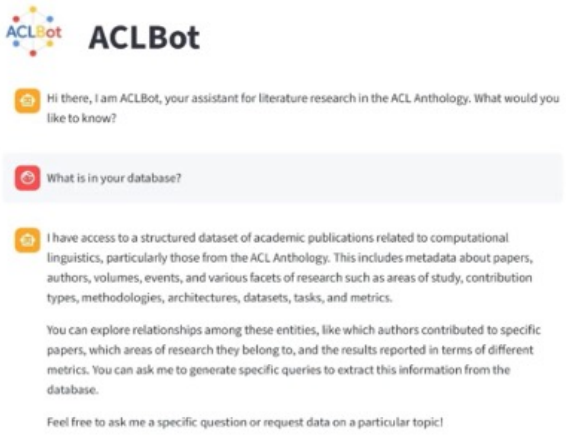
Buchmann, J., Lynden, S., Jokinen. K. (2026). ACLBot: A Knowledge Graph-Driven Assistant for ACL Anthology Research. Proceedings of the Fifteenth LREC 2026

Problem; With thousands of papers spanning decades of research, the ACL Anthology can be challenging to navigate, making it difficult for researchers to efficiently discover relevant work, analyze author contributions, and track emerging trends

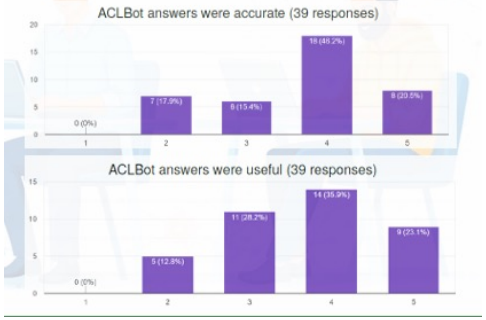
ACLBot

- Integrates a Neo4j-based knowledge graph constructed by extracting data on publications, authors, topics, and research trends from the ACL Anthology.
- Automatically generates knowledge graph (Cypher) queries to retrieve relevant information in response to user questions.





We conducted a user evaluation to assess the usefulness, clarity, and accuracy of the system's outputs.



Data Extraction

1. Corpus Collection

ACL Anthology

2. LLM-based extraction

Research classifications: each paper was assigned one Area and one or more `ContributionType` labels
Entities: mentions of `Task`, `Dataset`, `Metric`, `Architecture`, `Method`, and `PretrainedModel` were extracted.
Results: Experimental results were represented as `Result` nodes, each linked to a task, dataset, and metric, storing the reported score.

3. Data Quality Assessment

- Ensure the reliability of the data extracted by manual evaluation.
- `Area` and `ContributionType`:

	Precision
`Area`	0.87
`ContributionType`	0.81
Entities	0.90

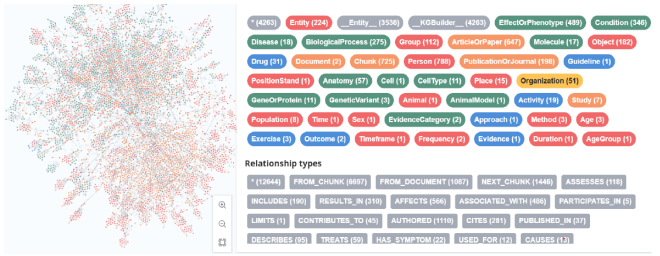
- `Result` extraction:

P	R	F1
Full Results		
0.52	0.52	0.52
Without Score		
0.71	0.62	0.66

Knowledge Graphs for Grounding

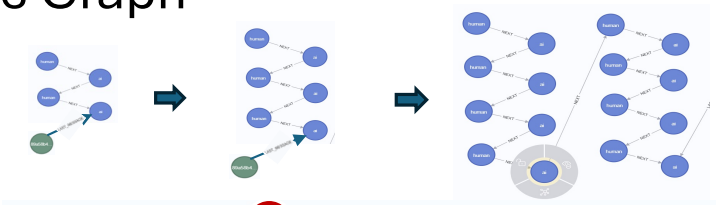
1. Graphs for different purposes

- Domain Graphs – knowledge about the real world
- Dialogue Graphs – who said what, when, topics, dialogue history
- Document Graph – which documents the information is based o



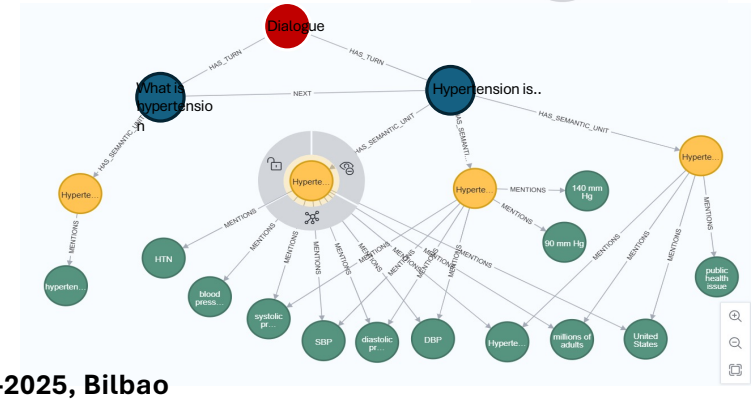
2. Adding Grounding Status to Entities in the Dialogue Graph

- Implicitly grounded, Explicitly grounded, Ungrounded
- Turns incrementally processed, added with with timestamps



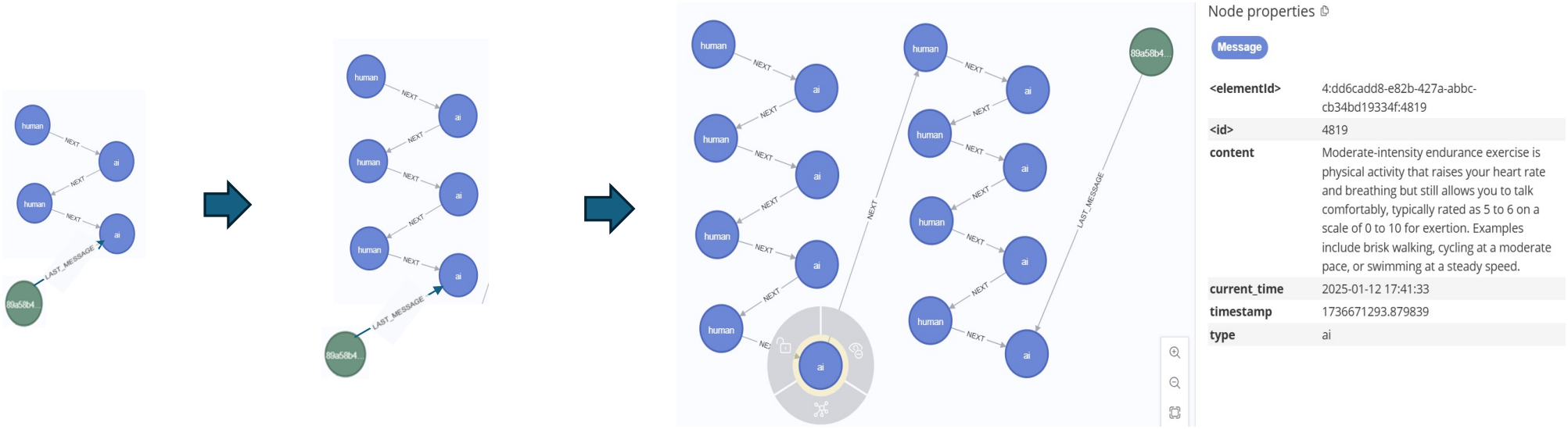
3. Four Layers of a Dialogue Graph

- Dialogue (top-level red node)
- Turns (blue nodes): added incrementally to dialogue history
- Semantic Units (yellow nodes) extracted from each Turn
- Conversational Entities (green nodes) extracted from each Semantic Unit



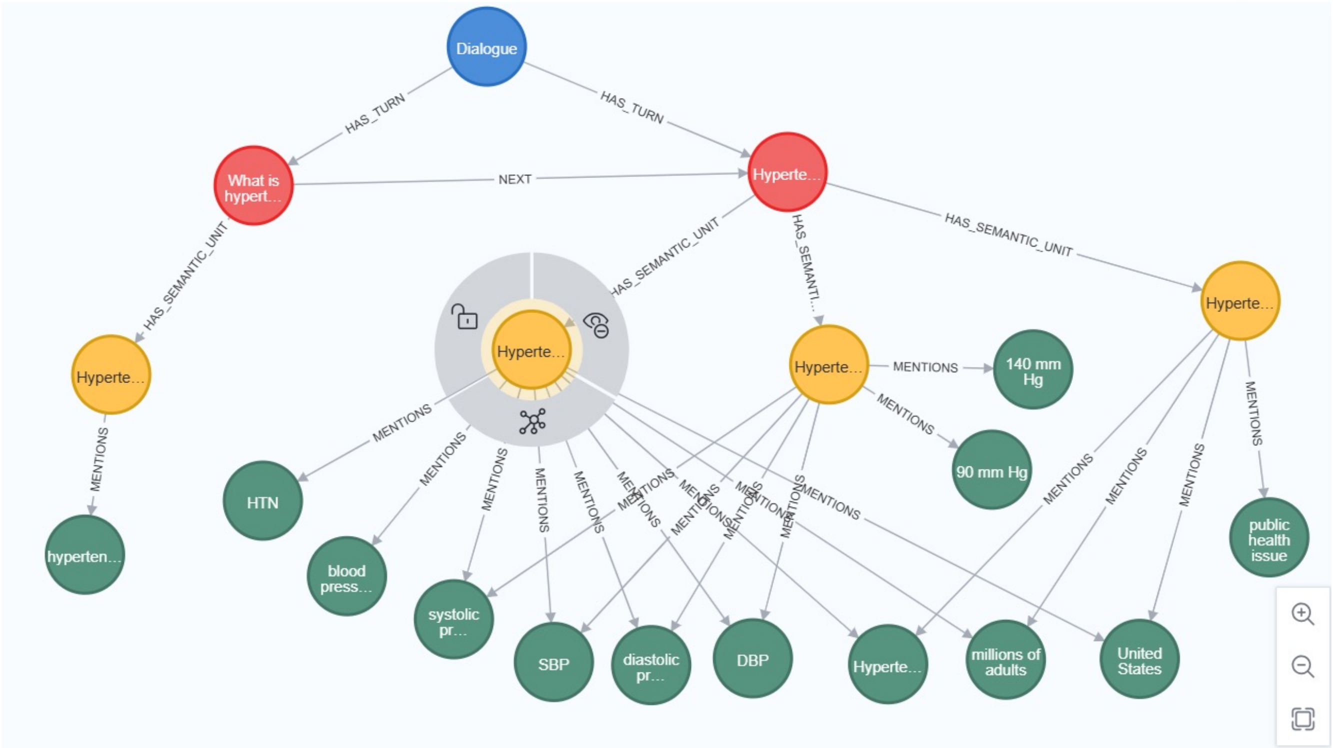
References:
Wilcock & Jokinen (2025). **Integrating Conversational Entities and Dialogue Histories. IWSDS-2025, Bilbao**
Walker, Ultes, Lison (2023). A graph-to-text approach to knowledge-grounded response generation in human-robot interaction. ArXiv:2311.16137

Adding Grounding Status to Entities in the Dialogue Graph (Implicitly grounded, Explicitly grounded)



- “Being Grounded” is a property of a node: semantic units (propositions) and conversational entities
- Explicit grounding => property is added to the node’s properties
- Implicit grounding => by virtue of being linked to a node that is explicitly grounded, the node is treated as being grounded unless contrary evidence appears
- Linking Conversational Entities to Domain Entities is important but separate matter

Four Layers of a Dialogue Graph - Incremental Dialogue Graph



- Dialogue (top-level blue node)
- Turns (red nodes) are added to the dialogue history graph incrementally, with timestamps.
- Semantic Units (yellow nodes) are extracted from each Turn by LLM.
- Conversational Entities (green nodes) are extracted from each Semantic Unit by LLM.

Wilcock & Jokinen (2025). **Integrating Conversational Entities and Dialogue Histories. IWSDS-2025, Bilbao**
Walker, Ultes, Lison (2023). A graph-to-text approach to knowledge-grounded response generation in human-robot interaction. ArXiv:2311.16137

Issues to discuss

- Knowledge graphs and the knowledge they represent:
  - Different levels of truth: facts and opinions
  - Disclosure of information: shared vs. private knowledge
  - Awareness of everyday biases

- Trust and influence
- Identity
- Privacy

Tom Williams, Cynthia Matuszek, Kristiina Jokinen, Raj Korpan, James Pustejovsky, Brian Scassellati:
Voice in the Machine: Ethical Considerations for Language-Capable Robots. CACM 2023/8

AIST
Institute of SCIENCE TOKYO

Short detour on representation

- Structured knowledge:
  - graph-based (knowledge graph)
  - row-based (tabular)
- Unstructured knowledge: natural language

- Graphs vs vectors: representations for different processing purposes

Neo4j vs. PageIndex

- Knowledge graphs use **graph structures**, **with fast graph search** and reasoning over document graph structures.
- Neo4j is also proud of **combining graph-based search with vector search** and **full-text search in flexible GraphRAG retrieval**.
- Neo4j is also proud of using Graph Data Science algorithms (GDS library) for fast searching over very big graphs.
- Vector search uses semantic similarity for fast approximate search of text chunks, with no reasoning and no document structures.
- Neo4j GraphRAG works on documents, but they have (new) **Neo4j Agent Memory** (NAM) that works on conversations.
- **NAM builds short-term and long-term memory graphs** for conversations. See https://neo4j.com/labs/agent-memory/explanation/

- PageIndex uses **tree structures**, with **fast tree search** and reasoning over document tree structures.
- PageIndex is proud of **not using vectors**. "Similarity ≠ relevance."
- See https://pageindex.ai/developer:
- "Vectorless, Reasoning-based RAG. Traceable, explainable, context-aware retrieval. No vector databases or chunking."
- "PageIndex delivers precise, context-aware retrieval by reasoning over document structure to find what's truly relevant."
- PageIndex works on documents, but also have (new) **ChatIndex** that works on conversations.
- **ChatIndex builds topic trees** for long conversations. See https://github.com/VectifyAI/ChatIndex/blob/main/README.md

Reinforcement Learning for Common Ground Generator

Mohapatra (2026), Mohapatra et al., 2024

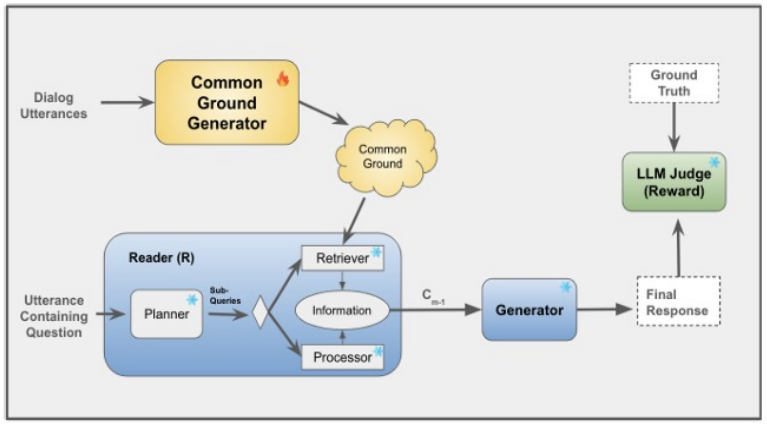
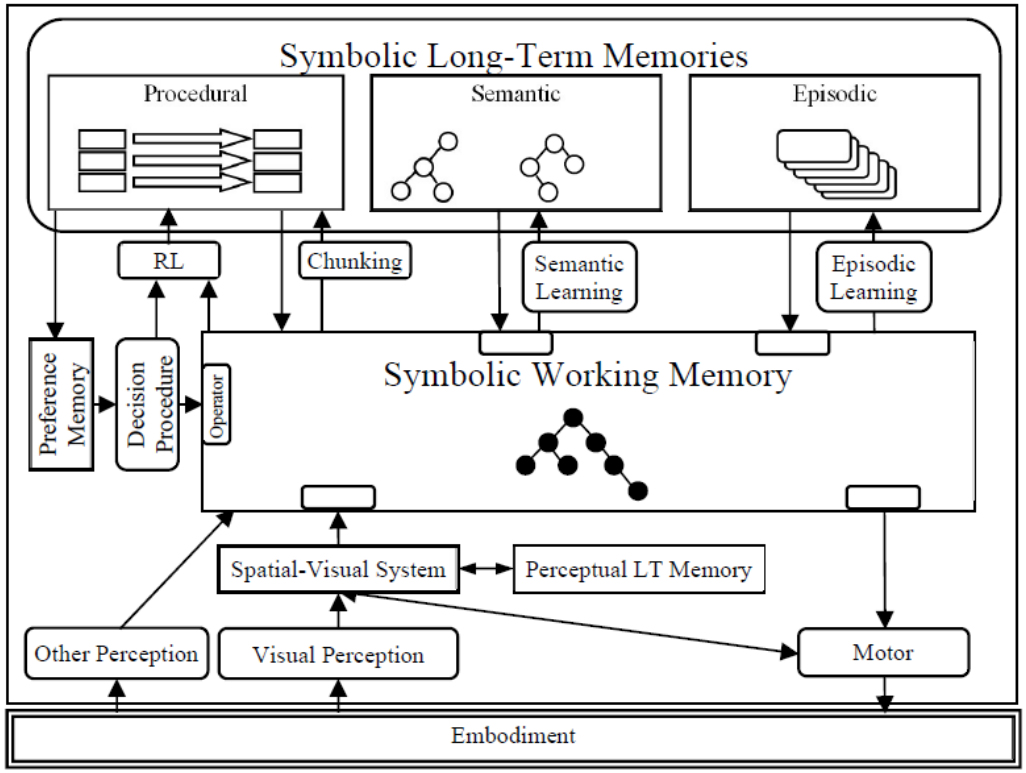


Figure 10.5: Visual representation of RL on our agentic framework where the Planner, Processor, Retriever and Response Generator are frozen LLMs.

- Reinforcement learning to optimize the externalized common ground (CG) representations produced by an agentic framework
- CGG is not rewarded for the syntax of its output, but for the downstream utility of that output
- The response is evaluated by an LLM Judge against the ground truth

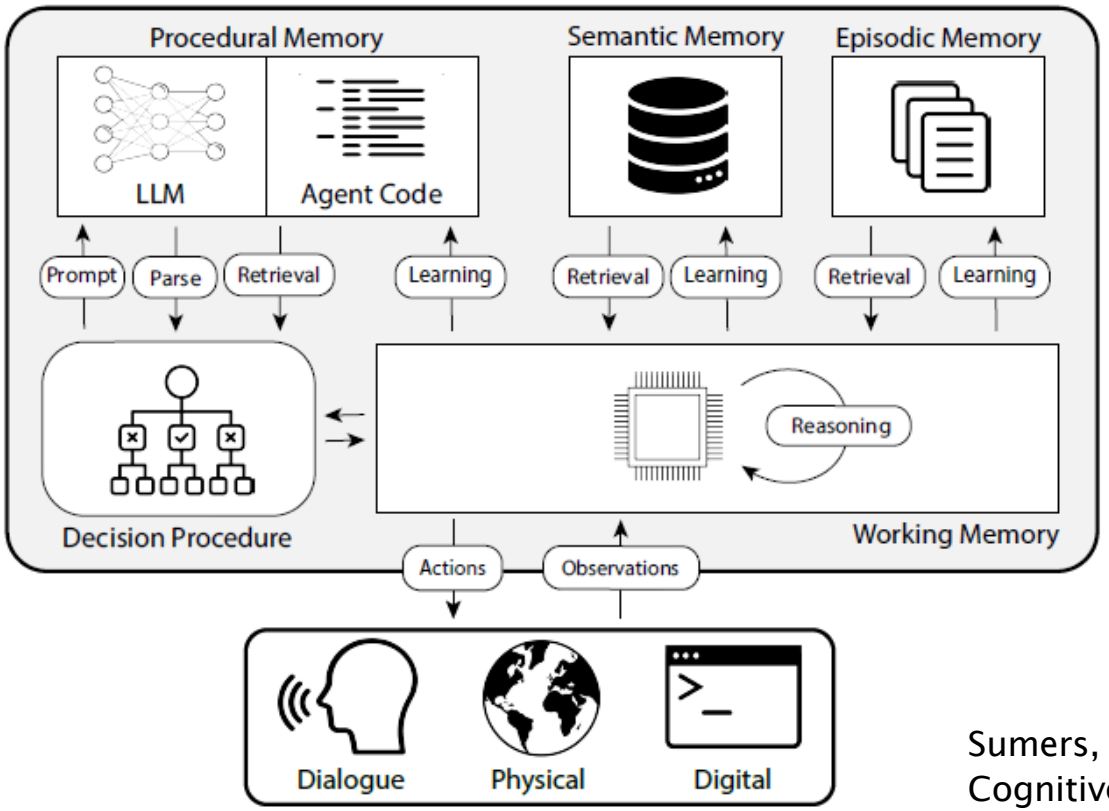
SOAR Cognitive Architecture (since 1980s)

Agentic AI is raising new interest in cognitive architectures:
Structured memory and cognitive design



Laird, Newell, Rosenbloom 1987

CoALA Architecture



Sumers, Yao, Narasimhan, Griffiths:
Cognitive Architectures for Language Agents

arXiv:2309.02427

Agentic Architecture, Grounding and Memory

CoALA Cognitive Architecture for Language Agents, cf. SOAR, ACT-R in 1980's
- three memory types: semantic, episodic, procedural
- complex workflows with generative language agents

Memory Type	What is Stored	Human Example	Agent Example
Semantic	Facts	Things I learned in school	Facts about a user
Episodic	Experiences	Things I did	Past agent actions
Procedural	Instruction	Instincts or motor skills	Agent system prompt

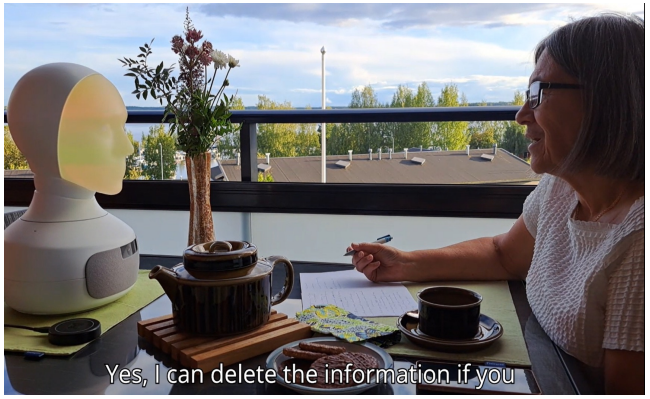
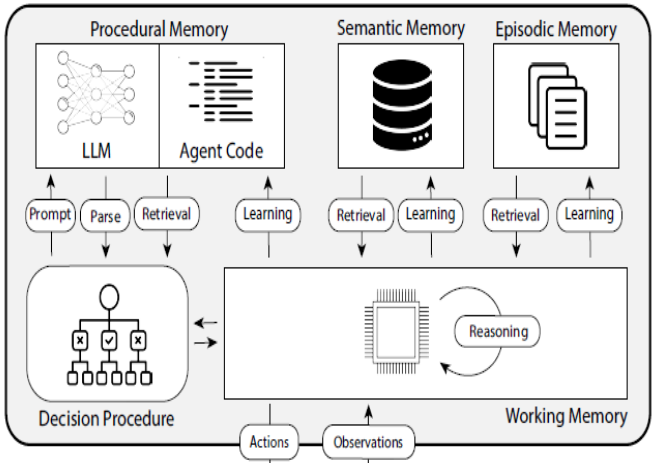


Figure: https://langchair-ai.github.io/langgraph/concepts/memory



K: **Well, what have you learnt about me in this conversation**?
F: I've learnt that you're Kristiina, planning a trip to Naples, and you're particularly interested in historical sites like Pompei and Castel Nuovo. If there's anything else you'd like to share or ask, feel free.
K: **How long will you store this information**?
F: I'll remember this information for our future conversations so I can help you better. If you ever want to update or remove anything, just let me know.
K: **Can you also delete the information if I want to**?
F: Yes, I can delete the information if you request it. Just let me know whenever you want to do that.

Dialogue: https://www.youtube.com/watch?v=f81Xad80AOQ

Mori, T., Den, Y., Jokinen, K. (2024). Cognitive Model of Response Token Generation. The 100th Symposium of JSAI-SLUD, Vol 100, pp 180-185. Best Paper Award. (in Japanese).
Wilcock and Jokinen (2025) Agents of Conversation: Grounding and Cognitive Architectures. ICSR-25 Workshop ASIMOV.

Some References

- Jokinen, K., Wilcock, G. (2025). Towards Domain Graphs and Dialogue Graphs for Conversational Grounding in HRI. The 20th ACM/IEEE Conference on Human-Robot Interaction, HRI-2025. https://dl.acm.org/doi/abs/10.5555/3721488.3721682
- Schneider, P., Machner, N., Jokinen, K., Matthes, F. Bridging Information Gaps in Dialogues with Grounded Exchanges Using Knowledge Graphs. Proceedings of the 25th Meeting of SIGDial, Kyoto, Japan, 2024. https://aclanthology.org/2024.sigdial-1.10/
- Jokinen, K., Schneider, P., Mori, T. (2024). Towards Harnessing Large Language Models for Comprehension of Conversational Grounding. The 25th Annual International Workshop on Spoken Dialogue System Technology (IWSDS), March 2024. https://www.researchgate.net/publication/381157675_Towards_Harnessing_Large_Language_Models_for_Comprehension_of_Conversational_Grounding
- Mori, T., Den, Y., Jokinen, K. (2024). "Cognitive Model of Response Token Generation". The 100th Symposium of JSAI-SLUD, Japanese Society for Artificial Intelligence Vol 100, pp. 180-185. Best Paper Award.
- Jokinen, K., Schneider, P., Mori, T. (2024). Towards Harnessing Large Language Models for Comprehension of Conversational Grounding. The 25th IWSDS, March 2024. https://www.researchgate.net/publication/381157675_Towards_Harnessing_Large_Language_Models_for_Comprehension_of_Conversational_Grounding
- Udagawa, Takuma and Akiko Aizawa (2019). A Natural Language Corpus of Common Grounding under Continuous and Partially-Observable Context. In: Proceedings of the AAAI Conference on Artificial Intelligence 33.01, pp. 7120 7127. doi: 10.1609/aaai.v33i01.33017120
- Tom Williams, Cynthia Matuszek, Kristiina Jokinen, Raj Korpan, James Pustejovsky, Brian Scassellati: Voice in the Machine: Ethical Considerations for Language-Capable Robots. CACM 2023/8
- Wilcock & Jokinen (2025). Integrating Conversational Entities and Dialogue Histories. IWSDS-2025, Bilbao
- Buchmann, J-, Lynden, S., Jokinen. K. (2026). ACLBot: A Knowledge Graph-Driven Assistant for ACL Anthology Research. Proceedings of the 15th LREC, 11(16): 2731-2741
- Walker, Ultes, Lison (2023). A graph-to-text approach to knowledge-grounded response generation in human-robot interaction. ArXiv:2311.16137

Some References (cont.)

- Michelle Elizabeth, Gwénolé Lecorvé, Lina M. Rojas Barahona, and Magalie Ochs. 2026. Conversational Grounding in Large Language Models: Evaluation Methods, Challenges and Future Directions, In *Proceedings of the 27th Annual Meeting of the Special Interest Group on Discourse and Dialogue*, pages 151–163, Atlanta, Georgia, USA. Association for Computational Linguistics.
- Mohapatra, Biswesh (2026). Beyond Fluency: Toward Robust Conversational Grounding for Neural Conversational Agents. PhD Thesis, Inria, Paris, France.
- Mohapatra, Biswesh, Seemab Hassan, Laurent Romary, and Justine Cassell. (2024a). Conversational grounding: Annotation and analysis of grounding acts and grounding units. In Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC COLING 2024), pages 3967–3977, Torino, Italia. ELRA and ICCL. url: https://api.semanticscholar. org/CorpusID:268680732.
- Mohapatra, Biswesh, Manav Nitin Kapadnis, Laurent Romary, and Justine Cassell (2024b). Evaluating the Effectiveness of Large Language Models in Establishing Conversational Grounding. In Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing, pages 9767–9781, Miami, Florida, USA. Association for Computational Linguistics. url: https://api.semanticscholar. org/CorpusID:273901650.
- Ilinykh, Nikolai, Sina Zarrieß, and David Schlangen (Sept. 2019). “Meet Up! A Cor pus of Joint Activity Dialogues in a Visual Environment”. In: Proceedings of the 23rd Workshop on the Semantics and Pragmatics of Dialogue- Full Papers. Lon don, United Kingdom: SEMDIAL. url: http://semdial.org/anthology/Z19_Ilinykh_semdial_0006.pdf
- LeCun, Yann, Yoshua Bengio, and Geoffrey Hinton (May 2015). “Deep Learning”. In: Nature 521, pp. 436–44. doi: 10.1038/nature14539.
- Shuster, Kurt et al. (Nov. 2021). “Retrieval Augmentation Reduces Hallucination in Conversation”. In: Findings of the Association for Computational Linguis tics: EMNLP 2021, pp. 3784–3803. doi: 10.18653/v1/2021.findings-emnlp.320. url: https://aclanthology.org/ 2021.findings-emnlp.320/
- Mohapatra, Biswesh et al. (2026). Frame of Reference: Addressing the Challenges of Common Ground Representation in Situational Dialogs. arXiv: 2601.09365 [cs.CL]. url: https://arxiv.org/abs/2601.09365.
- Linhui Xiao, Xiaoshan Yang, Xiangyuan Lan, Yaowei Wang, and Changsheng Xu. 2025. Towards Visual Grounding: A Survey . IEEE Transactions on Pat tern Analysis & Machine Intelligence, (01):1–20.
- Wang, X., Wei, J., Schuurmans, D., Le, Q., Chi, E., Narang, S., Chowdhery, A., and Denny, Z.: Self Consistency Improves Chain of Thought Reasoning in Language Models, arXiv:2203.11171 (2022)

Publication forums

- Journal of Multimodal Interaction

  https://link.springer.com/journal/12193/submission-guidelines

Thank you!

David Traum:

Bio and papers:
Projects and videos:
LinkedIn:
Contact:

Kristiina Jokinen:

Bio and papers: https://orcid.org/0000-0003-1229-239X CDM
Robot videos: https://www.cdminteract.com/videos
LinkedIn: https://www.linkedin.com/in/kristiina-jokinen-b01585/
Contact: Kristiina.Jokinen@helsinki.fi
Kristiina.Jokinen@aist.go.jp