



chist-era



Interactive Grounded Language Understanding ([IGLU](#))

<https://iglu-chistera.github.io/>

Jean Rouat E&C Engineering Dept. ([GEGI](#))

Neurocomputational and Intelligent Signal Processing
Research Group ([NECOTIS](#))

<https://www.gel.usherbrooke.ca/necotis/>

Virtual Projects Seminar Meeting, 28 March 2022



NECOTIS

neurosciences computationnelles et traitement
intelligent des signaux



Institut
interdisciplinaire
d'innovation
technologique



UNIVERSITÉ DE
SHERBROOKE

Who is IGLU?

- 7 research teams, across 5 different countries, it is a total effort of 325 person-months

- Experts in

Deep learning; UdeM (MILA), Québec (CS)

Reinforcement learning; Lille1 (CRISTAL), France (CS)

Neurosciences & cognitive sciences: UdeS (NECOTIS/3IT), Québec (Eng.)

Robotics: UNIZAR, Spain & INRIA, Bordeaux, France, UdeS (3IT), Québec (Rob.)

Signal Processing: UMONS, Belgium & UdeS (NECOTIS/3IT), Québec (Eng.)

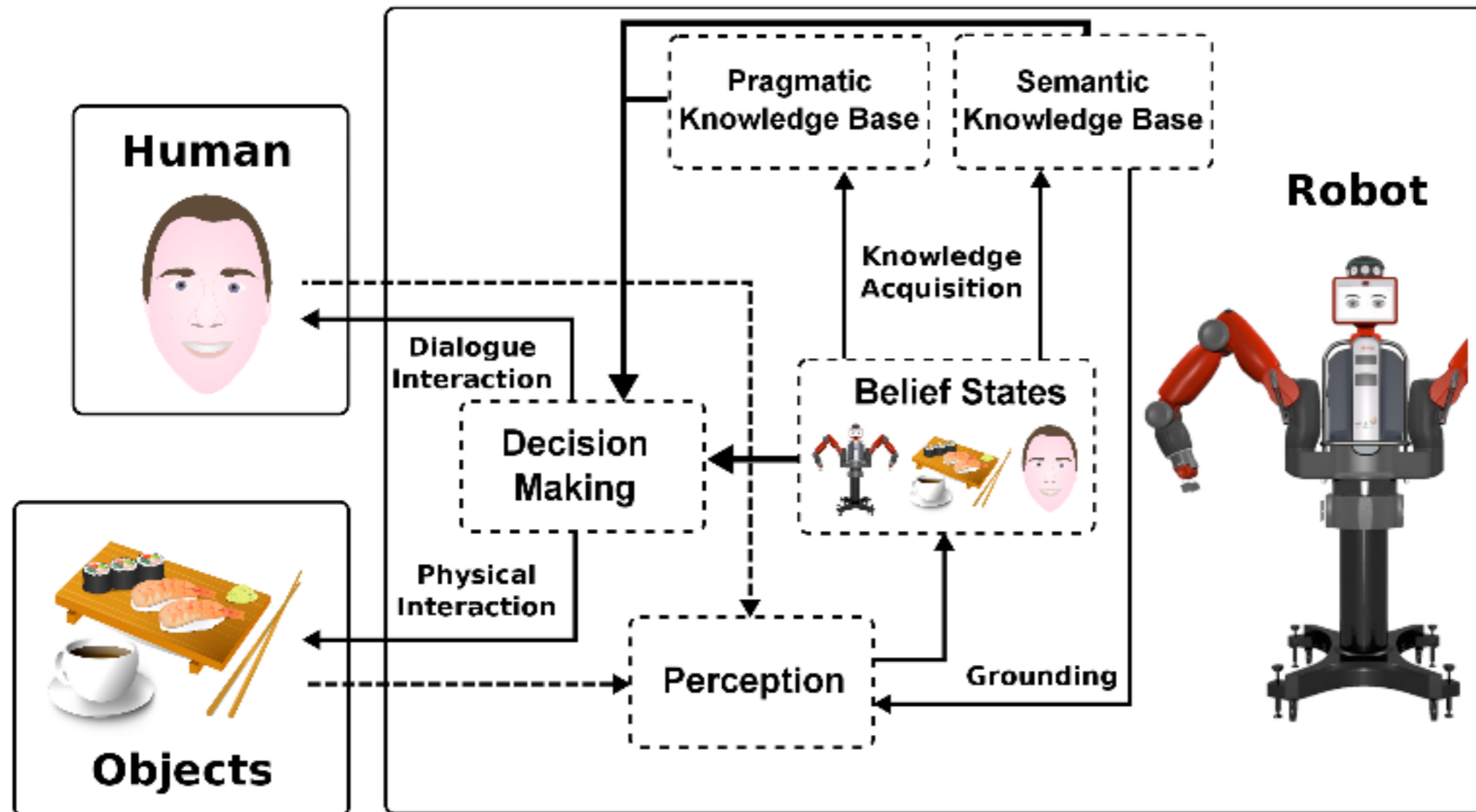
Human-machine Int.: KTH, Sweden & UMONS (Eng.)

- 1 Post-Doc., 11 PhD, 3 Msc., 10 undergraduate



J. Rouat, Virtual Projects Seminar Meeting, 28 March 2022

Initial Project (robot cooking)



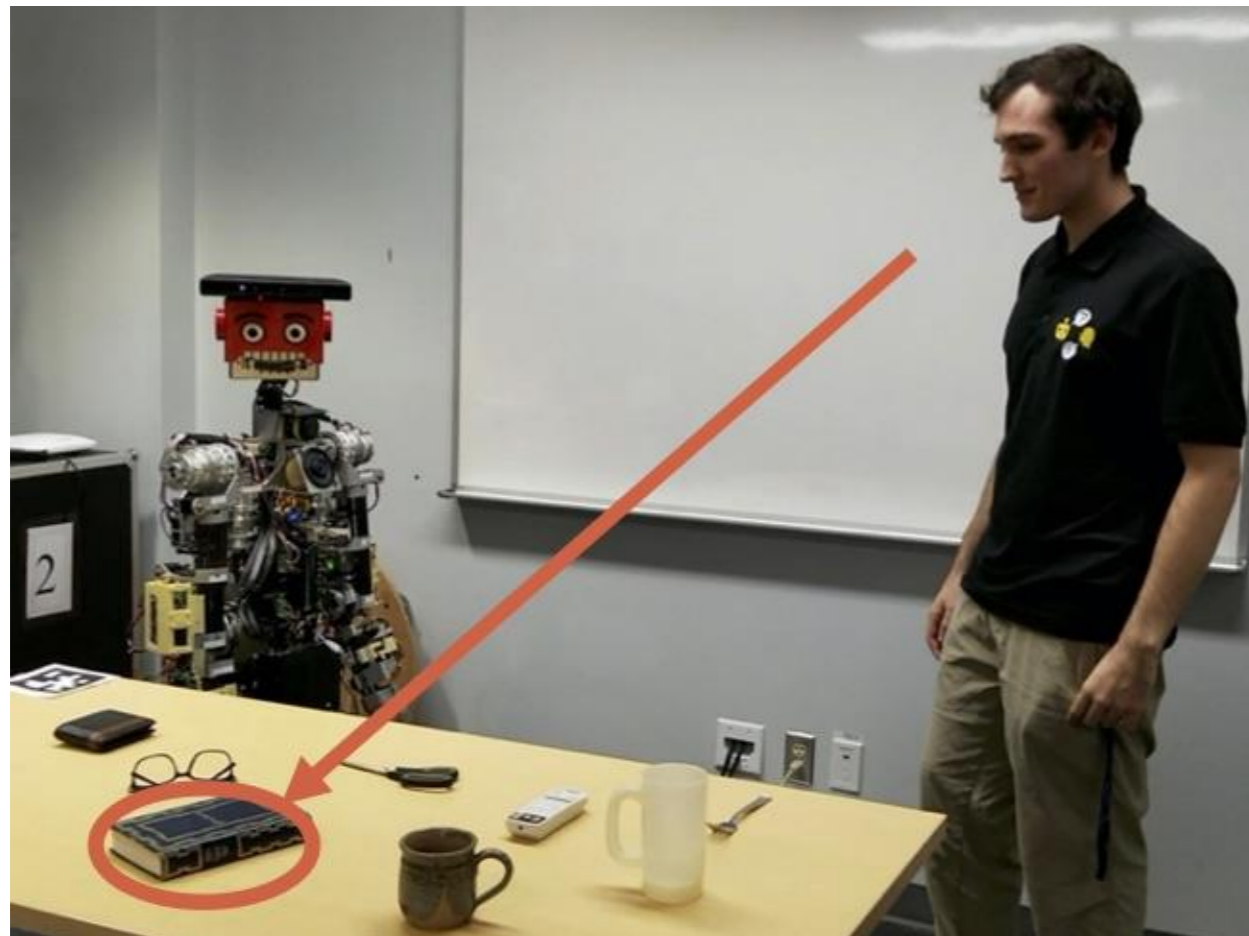
Targeted Sub Goals

- Contribute to HLU with architectures comprising:
 - Human-Agent Interactions and Features
 - Interactions/modulation/coupling;
 - Multimodal features (audio, visual, text);
 - Human-Agent dialog;
 - Learning;
 - Emotion expression;
- Real grounded system implementations (Engineering and robotic applications);
- Datasets and open source software for reproducibility and research sharing.

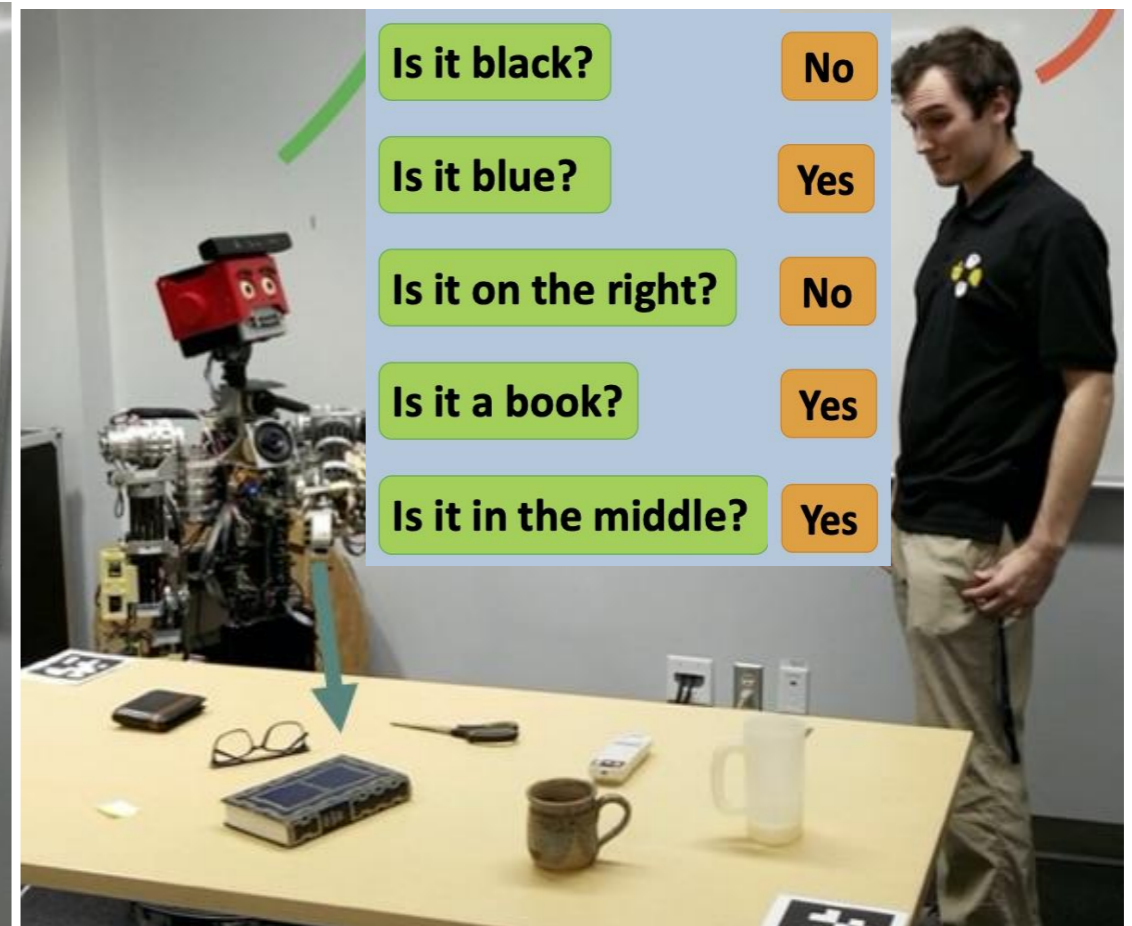


One Platform: Interactive Visual or Audio or Multimodal Question Answering

Illustration here with a Visual GuessWhat?! Game



Human thinks about the object



Agent guesses and points to the object

<https://www.youtube.com/watch?v=GexxiWIZyLA> (in French)

https://devineproject.github.io/abstract_devine.pdf (NeurIPS 2018 demo)



J. Rouat, Virtual Projects Seminar Meeting, 28 March 2022

Key Developments

See IGLU's Web page: <https://iglu-chistera.github.io>

- FiLM layers to modulate/condition one modality with another one;
- Multimodal Environments: HoME, Multimodal Human-Robot Interaction (MHRI), Audio-Visual Event Recognition, Fusion of modalities, Multimodal Translation;
- Learning from Multimodal Human-Robot Interactions;
- Visual Active Speaker detection, GCC-NMF sound source separation for the cocktail party effect;
- Generative modeling for language learning using probabilistic frameworks;
- Databases: CREATE, MHRI, Audio for HoME, CLEAR, SECLUMONS



J. Rouat, Virtual Projects Seminar Meeting, 28 March 2022

Exploitation built with IGLU realizations

- Integration of developed algorithms on multiple robotic platforms (Humanoid IRL-1, Baxter);
- GCC-NMF sound source separation implemented on a raspberry pi with very small latency (< 10ms);
- Simulation environment and open source software & databases;
- First design and implementation of neural networks for Audio Question Answering (NAQA);
- First HoME multimodal environment;
- First multimodal event detection system with features conditioning and fusion;



Valorization built with IGLU

- GuessWhat?! has been displayed for 3-5 years in the Inria Showroom and has been presented all around France as a demonstration of current AI research. A dozen of laboratories around the world (GeorgiaTech, Stanford, Cambridge, University of Amsterdam etc.) use GuessWhat?! Software.
- Many international workshops : NeurIPS ViGIL since 2017, Interspeech 2017 satellite workshop, European Reinforcement Learning workshop in 2018, eINTERFACE
- Master Class on HLU, that was held in Paris on 10 and 11 April 2018 (<https://chistera-hlu.sciencesconf.org>,).
- Many invited talks from different members of the consortium.



Impacts

- We Contributed to HLU with architectures:
 - Human-Agent Interactions and Features
 - Interactions/modulation/coupling;
 - Multimodal features (audio, visual, text);
 - Human-Agent dialog;
 - Learning;
- Real grounded system implementations (Engineering and robotic applications);
- Datasets and open source software for reproducibility and research sharing.
- A theoretical framework to assist in the choice of deep architectures, weight configurations and initializations for a given application.



Expected Impacts

- Examples of scientific and technical:
 - Modulation/coupling of modalities
 - Neuromorphic implementations of solutions proposed in IGLU
 - Others
- Long-term societal impacts
 - Bring agents in closer cooperation on daily tasks by having good interpretation and communication skills, as well as extensible knowledge.
 - Intelligent robotic agents of the future would adapt naturally to the user, based on the (perceived and simulated) experience of the user.



Conclusion

- Working with multi-disciplinary partners is difficult (culture, technical vocabulary, evaluation procedures, etc.);
- We adapted the IGLU's roadmap to take into account this situation and designed a set of sub-problems for more specific applications and algorithms that can be at first independent;
- Then, these algorithms can then be combined for the development of more general purposes Human Language Understanding Systems.

