HOL as a Lingua Franca for Argumentative Reasoning Agents

David Fuenmayor[1,2] and Christoph Benzmüller[1,2]

[1]Otto-Friedrich-Universität Bamberg, Germany and [2]Freie Universität Berlin, Germany

The proposed talk is based upon the PhD dissertation by the first author and supervised by the second author. This work explores the potential of classical higher-order logic (HOL) as a general-purpose meta-language capable of representing and formally reasoning about diverse domain-specific object languages. To demonstrate its versatility, we have developed a library composed of combinator-based modular building blocks within the Isabelle/HOL proof assistant, but easily translatable to other HOL-based systems.[1] Extending the LogiKEy [1–3] knowledge representation and reasoning methodology by a combinator-based layer, this library provides a set of fundamental functional "building blocks" from which more complex mathematical definitions, axioms, and conjectures can be automatically generated (e.g. guessed by an AI), while leveraging HOL's efficient type-checking algorithms and sound theorem provers as ground truth signal to close the feedback loop. We propose that this library exemplifies a universal mathematical language layered over HOL, sufficiently powerful to capture and represent a broad range of logical formalisms—whether classical or non-classical—as well as their underlying classes of mathematical structures. Moreover, we delve into the domain of multi-agent systems (MAS), introducing the concept of Argumentative Reasoning Agents (ARAs). This kind of agents, we suggest, are particularly well-suited to leverage the expressive power of the proposed library. In this manner, HOL emerges as an ideal common language—a lingua franca—for facilitating effective reasoning and interaction among argumentative reasoning agents in AI [4].

An important insight in the multi-agent systems (MAS) literature is that epistemic ("knows") and normative ("oughts") aspects are fundamentally intertwined in agency. On one hand, agents interpret and act upon the social norms, conventions, obligations, and permissions governing their interactions. Such norms can be represented formally (often using deontic logic) and integrated into an agent's planning or decision-making, guiding its behavior toward compliance or sanction avoidance. On the other hand, agents must reason about what they themselves know, what other agents know, and how these knowledge states change as information is exchanged or observed. This involves epistemic logic or related formalisms (dynamic logic, belief revision, etc.) for capturing and inferring statements about knowledge or belief. We refer to this kind of artificial agents as *ARAs: Argumentative Reasoning Agents*. By combining normative and epistemic reasoning, ARAs can achieve coordinated, context-aware behavior, ensuring each agent not only follows shared rules but also reasons effectively about what it and others know, so that it may act appropriately.

When it comes to contemporary "AI agents," the typical MAS researcher is initially shocked (or humbled) to discover that the current hype almost completely ignores decades of insights from 'symbolic' academic research—yet it still yields impressive results. On closer examination, one notices how mainstream deep-learning-based AI seems to be rediscovering or reinventing good old-fashioned concepts (sometimes the hard way). Thus, contemporary AI has, over the last couple of years (especially since the release of ChatGPT), become something like an '80s retro party. When we look at today's 'hot' AI topics, we see terms like "reasoning," "planning,"

[1] https://github.com/davfuenmayor/logic-bricks

"(graph-) RAG" (aka. Knowledge Representation and Reasoning), and "agents" (aka. MAS, as discussed before).

Arguably, the new era of language-capable AI systems (LLMs and the like) sets the stage for a logic-based AI renaissance. The historical "impedance mismatch" between natural and formal languages finally seems bridged. ARAs, as we conceive them, are meant to function as such 'expert' subsystems, hidden behind a ChatGPT-like "dispatcher" that handles natural-language interfacing with humans, leveraging its world knowledge to figure out what users actually want. It can then use its strong language- and code-generation capabilities to convert user requests into the ARA's controlled input language (syntactic sugar that expands into a fragment of HOL, as will be further discussed in the talk).

Below are some required capabilities which directly translate into expressivity requirements for a formal language for ARAs (we will argue that our HOL-based "universal mathematical language" addresses all of them):

1. **ARAs need to do math.** Not only evaluation of mathematical expressions or solving equations, but, importantly, also (meta-)theoretical investigations.
2. **ARAs need to understand (like) and communicate with humans.** This is nowadays greatly facilitated by LLMs, which might do the tricky step of converting between natural languages and an appropriate intermediate representation, which we conceive as a very expressive HOL-based *Controlled Natural Language* (CNL). This allows ARAs to represent and analyse sentence structure (cf. Lambek calculi and Montague semantics), as well as to properly handle modalities (alethic, epistemic, deontic, etc.), generalized quantifiers, indexicals and anaphora.
3. **ARAs need to create and communicate programs.** Beyond LLM-driven code generation, they need to actually analyze and reason rigorously about of the meaning (semantics) of the programs they or other ARAs generate (cf. *program synthesis*).
4. **ARAs need to carry out planning and resource-sensitive reasoning.** ARAs should devise plans (sequences of actions) to achieve goals and reason about resources (time, objects, energy, etc.) consumed or produced.
5. **ARAs need to engage in strategic reasoning and coalition-building.** Their language should support reasosing about strategies or policies.
6. **ARAs need to reason reliably with partial and contradictory evidence.** In particular, paraconsistent reasoning, meaning that contradictory statements do not trivialize or collapse the inferential process (i.e. the traditional *ex contradictio quodlibet*).
7. **ARAs need to engage in counterfactual reasoning.** Considering hypothetical 'what ifs' to explore consequences of actions or assumptions (possibly contrary to known facts).
8. **ARAs need to represent and reason about argumentive discourse.** When ARAs contemplate multiple arguments (which may conflict), they need to evaluate which arguments prevail (cf. abstract argumentation frameworks).

Finally, we discuss current plans for implementing an agent-oriented, open-source framework built on the battle-tested Elixir/Erlang ecosystem for concurrent programming, with its rich ecosystem of libraries for machine learning and LLM interaction, as well for interaction with proof assistants.[2] Our framework shall allow ARAs to reliably invoke HOL-based theorem provers, AI-driven abductive oracles, and various SMT/SAT solvers, while storing, pruning, and sharing formally verified intermediate results. A first envisaged application of this framework consists in large-scale mathematical theory exploration using the previously discussed Isabelle/HOL combinatory logic bricks library.

[2] Cf. integration with Isabelle/HOL: https://github.com/davfuenmayor/isabelle_elixir.

References

[1] C. Benzmüller, X. Parent, and L. van der Torre. "Designing Normative Theories for Ethical and Legal Reasoning: LogiKEy Framework, Methodology, and Tool Support". In: *Artificial Intelligence* 287 (2020), p. 103348. DOI: 10.1016/j.artint.2020.103348.

[2] C. Benzmüller, A. Farjami, D. Fuenmayor, P. Meder, X. Parent, A. Steen, L. van der Torre, and V. Zahoransky. "LogiKEy Workbench: Deontic Logics, Logic Combinations and Expressive Ethical and Legal Reasoning (Isabelle/HOL Dataset)". In: *Data in Brief* 33.106409 (2020), pp. 1–10. DOI: 10.1016/j.dib.2020.106409.

[3] C. Benzmüller, D. Fuenmayor, and B. Lomfeld. "Modelling Value-oriented Legal Reasoning in LogiKEy". In: *Logics* 2.1 (2024), pp. 31–78. DOI: 10.3390/logics2010003.

[4] D. Fuenmayor and C. Benzmüller. "Higher-order Logic as a Lingua Franca for Logico-Pluralist Argumentation". In: *Logics for New-Generation AI. Second International Workshop. 10-12 June 2022, Zhuhai*. Ed. by B. Liao, R. Markovich, and Y. N. Wáng. College Publications, 2022, pp. 83–94. URL: http://www.collegepublications.co.uk/downloads/LNGAI00002.pdf.