

# The Fatio Protocol for Formal Dialogue in Isabelle/HOL

Luca Pasetto<sup>1</sup> and Christoph Benzmüller<sup>2,3</sup>

<sup>1</sup> University of Luxembourg

`luca.pasetto@uni.lu`

<sup>2</sup> Otto-Friedrich-Universität Bamberg

<sup>3</sup> Freie Universität Berlin

`christoph.benzmueller@uni-bamberg.de`

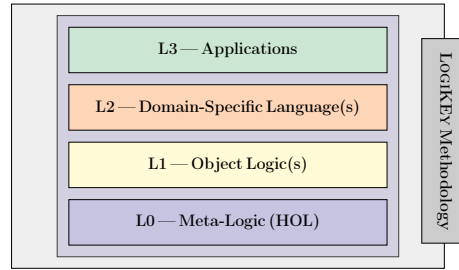
## Extended Abstract

Recent advancements in dialogue systems based on language models are pushing the boundaries of AI-human interaction. The rapid evolution of these technologies can set the stage for more intuitive and helpful AI systems in everyday life, but it also presents various risks, demonstrating the need for research into formal dialogue systems based on logical languages, where relevant properties can be formally and then mechanically verified.

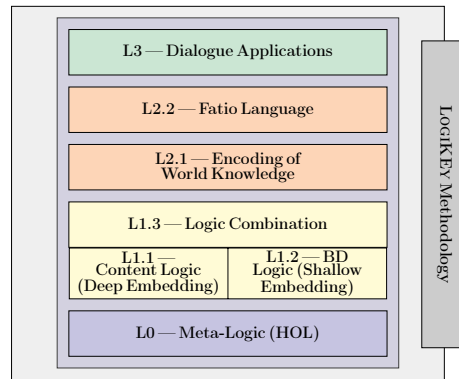
In this presentation, we discuss a modeling and implementation of *Fatio*[3], a formal dialogue system based on argumentation between multiple agents. The modeling follows the *LogiKEy*[2] framework and methodology, and is implemented using the Isabelle/HOL proof assistant system[4].

*LogiKEy*[2] is a framework and methodology for the design, engineering, and experimentation of reasoning systems, with a focus on ethico-legal applications and normative reasoning. The formal framework of *LogiKEy* is based on *shallow semantical embeddings* (SSE)[1], allowing combinations of various logics to be embedded in classical higher-order logic (HOL). This meta-logical approach enables the use of off-the-shelf theorem provers and model finders for HOL, assisting in the development of ethical intelligent agents without requiring the creation of entirely new tools. Furthermore, continuous improvements to theorem provers enhance LogiKEy’s reasoning capabilities without additional modifications. The methodology is illustrated in Figure 1: relevant object logics are encoded in HOL and can be used to express domain theories, facilitating experimentation with concrete applications and examples.

Fatio is a dialogue protocol that has been proposed by McBurney and Parsons[3] to extend the language for agent communications FIPA ACL with locutions for dialectical argumentation. The protocol is defined with a syntax, an axiomatic semantics, and an operational semantics. Its syntax is composed of five agent locutions, and its axiomatic semantics is defined in terms of the mental states (beliefs and desires) of the participating agents, that are influenced by what is uttered. Figure 2 shows the layers of the LOGIKEY methodology applied to Fatio.



**Fig. 1.** LOGiKEY development methodology



**Fig. 2.** LOGiKEY development methodology for Fatio

After presenting an implementation[5] of the axiomatic semantics of Fatio in the Isabelle/HOL proof assistant, we discuss whether and how such a formal dialogue system could be used to evaluate an LLM-based dialogue system.

**Acknowledgements** This work was supported by the Luxembourg National Research Fund (FNR) through the project Logical methods for Deontic Explanations (INTER/DFG/23/17415164/LODEX) and the project Deontic Logic for Epistemic Rights (OPEN O20/14776480).

## References

1. Benzmüller, C.: Universal (meta-)logical reasoning: Recent successes. *Science of Computer Programming* **172**, 48–62 (2019). <https://doi.org/10.1016/j.scico.2018.10.008>
2. Benzmüller, C., Parent, X., van der Torre, L.: Designing normative theories for ethical and legal reasoning: LogiKEy framework, methodology, and tool support. *Artificial Intelligence* **287**, 103348 (2020). <https://doi.org/10.1016/j.artint.2020.103348>
3. McBurney, P., Parsons, S.: Locutions for argumentation in agent interaction protocols. In: van Eijk, R.M., Huget, M.P., Dignum, F. (eds.) *Agent Communication*. pp. 209–225. Springer Berlin Heidelberg, Berlin, Heidelberg (2005)

4. Nipkow, T., Paulson, L.C., Wenzel, M.: Isabelle/HOL - A Proof Assistant for Higher-Order Logic, Lecture Notes in Computer Science, vol. 2283. Springer (2002)
5. Pasetto, L., Benzmüller, C.: Implementing the fatio protocol for multi-agent argumentation in logikey. In: Benzmüller, C., Otten, J., Ramanayake, R. (eds.) Proceedings of the 5th International Workshop on Automated Reasoning in Quantified Non-Classical Logics (ARQNL 2024). vol. 3875, pp. 38–47. CEUR Workshop Proceedings, CEUR-WS.org (2024), <https://ceur-ws.org/Vol-3875/>