

# Diálogos de Intenções em Sistemas Multi-Agentes Baseados em Ontologias e Argumentação

Débora Cristina Engelmann<sup>1,2</sup>, Alison R. Panisson<sup>3</sup>, Angelo Ferrando<sup>2</sup>,  
Viviana Mascardi<sup>2</sup>, Rafael H. Bordini<sup>1</sup>

<sup>1</sup>School of Technology – Pontifícia Universidade Católica do Rio Grande do Sul (PUCRS)  
Porto Alegre – RS, Brasil

<sup>2</sup>DIBRIS – University of Genoa (UniGe)  
Gênova, Itália

<sup>3</sup>Department of Computing – UFSC  
Araranguá – SC, Brasil

debora.engelmann@acad.pucrs.br, alison.panisson@ufsc.br,

{angelo.ferrando,viviana.mascardi}@unige.it, rafael.bordini@pucrs.br

**Abstract.** *Some areas of application, for example, healthcare, are known to resist the replacement of human operators by fully autonomous systems. It is typically not transparent to users how artificial intelligence systems make decisions or obtain information, making it difficult for users to trust them. To address this issue, we investigate how argumentation theory and ontology techniques can be used together with reasoning about intentions to build complex natural language dialogues to support human decision-making. Based on such an investigation, we propose MAIDS, a framework for developing multi-agent intentional dialogue systems, which can be used in different domains. Our framework is modular so that it can be used in its entirety or just the modules that fulfil the requirements of each system to be developed. Our work also includes the formalisation of a novel dialogue-subdialogue structure with which we can address ontological or theory-of-mind issues and later return to the main subject. As a case study, we have developed a multi-agent system using the MAIDS framework to support healthcare professionals in making decisions on hospital bed allocations. Furthermore, we evaluated this multi-agent system with domain experts using real data from a hospital. The specialists who evaluated our system strongly agree or agree that the dialogues in which they participated fulfil Cohen's desiderata for task-oriented dialogue systems. Our agents have the ability to explain to the user how they arrived at certain conclusions. Moreover, they have semantic representations as well as representations of the mental state of the dialogue participants, allowing the formulation of coherent justifications expressed in natural language, therefore, easy for human participants to understand. This indicates the potential of the framework introduced in this thesis for the practical development of explainable intelligent systems as well as systems supporting hybrid intelligence.*

**Resumo.** *Algumas áreas, como por exemplo a área da saúde, são conhecidas por resistirem à substituição de operadores humanos por sistemas totalmente*

*autônomos. Normalmente, não é transparente para os usuários como os sistemas de inteligência artificial tomam decisões ou obtêm informações, dificultando a confiança dos usuários. Para abordar essa questão, foi investigado como técnicas de teoria da argumentação e de ontologia podem ser usadas em conjunto com raciocínio sobre intenções para construir diálogos complexos em linguagem natural a fim de apoiar a tomada de decisão humana. Com base nessa investigação, foi proposto o MAIDS, um framework para desenvolvimento de sistemas de diálogo intencional multiagente, que pode ser usado em diferentes domínios. O framework é modular para que possa ser utilizado integralmente ou apenas os módulos que se encaixem nas necessidades de cada sistema a ser desenvolvido. O trabalho também inclui a formalização de uma nova estrutura de diálogo-sub-diálogo com a qual se pode resolver questões ontológicas ou de teoria da mente e posteriormente retornar ao assunto principal. Como estudo de caso, foi desenvolvido um sistema multiagente usando o framework MAIDS para auxiliar na tomada de decisão de profissionais da saúde sobre alocação de leitos hospitalares. Além disso, esse sistema multiagente foi avaliado com especialistas do domínio usando dados reais de um hospital. Os especialistas que avaliaram o sistema concordam fortemente ou concordam que o sistema contempla os desideratos de Cohen para sistemas de diálogo orientados a tarefas. No sistema desenvolvido os agentes têm a capacidade de explicar ao usuário como chegaram a determinadas conclusões. Além disso possuem representação semântica e de estados mentais dos participantes de diálogo permitindo a formulação de justificativas coerentes em linguagem natural, portanto de fácil compreensão para os participantes humanos. Isso demonstra o potencial do framework apresentado nessa tese para o desenvolvimento prático de sistemas explicáveis e sistemas de inteligência híbrida.*

## **1. Introdução**

A pesquisa se concentra em combinar técnicas de Teoria da Argumentação e Ontologia para dar suporte a diálogos complexos em linguagem natural. Em particular, foi criada uma abordagem para apoiar o desenvolvimento de sistemas de diálogo que tiram proveito dessa combinação de técnicas juntamente com agentes BDI (Belief-Desire-Intention) para auxiliar os humanos na tomada de decisão. Essa abordagem é adaptável e suficiente para ser aplicada em vários domínios.

Muitas áreas podem se beneficiar de um sistema de apoio à decisão. Na área da saúde, por exemplo, a alocação de leitos representa um desafio para os hospitais (principalmente em países em desenvolvimento como o Brasil), pois os leitos hospitalares são um recurso escasso. Além disso, os ambientes hospitalares são altamente dinâmicos e incertos, portanto, a alocação otimizada de leitos hospitalares desempenha um papel essencial no planejamento geral dos recursos hospitalares. Um sistema que sugere melhores alocações de leitos para o profissional responsável por essa tarefa é interessante nesse contexto.

O framework para Sistema de Diálogo Intencional Multiagente (MAIDS) que é apresentado nesta tese concentra-se na combinação de técnicas da Teoria da Argumentação [de Oliveira Gabriel et al. 2020], Ontologia [Gruber 1995] e Teoria da Mente (ToM) [Goldman 2012] para dar suporte a diálogos complexos em linguagem

natural. O MAIDS inclui componentes para apoiar o desenvolvimento de aplicações complexas de Sistemas Multi-Agentes (MAS), tais como: (i) suporte para a criação de diálogos em linguagem natural para facilitar a interação com operadores humanos; (ii) raciocínio baseado em argumentação e diálogos que permitem aos agentes raciocinar e comunicar informações bem fundamentadas; (iii) ontologias para ajudar os agentes a organizar o conhecimento do domínio e realizar o raciocínio semântico; e (iv) teoria da mente para permitir aos agentes inferir atitudes mentais de outros agentes levando a uma comunicação intencional. Todos esses componentes são desenvolvidos com base em uma Linguagem de Programação Orientada a Agentes (AOPL) baseada na arquitetura BDI, que fornece uma base adequada para todas essas técnicas.

## 2. Motivação

A utilização de sistemas de inteligência artificial no cotidiano está se tornando madura e ubíqua, resultando na crescente disponibilidade de sistemas onde agentes e humanos trabalham em conjunto, de onde surgem novos conceitos como a Inteligência Híbrida [Akata et al. 2020]. No entanto, existem muitos desafios relacionados a como humanos e agentes irão interagir nesses sistemas e como os agentes irão explicar suas decisões e estados mentais internos para que se tornem mais transparentes e confiáveis.

Em algumas áreas, há resistência à substituição de operadores humanos por sistemas totalmente autônomos. Sabe-se que uma delas, é a área da saúde. Isso é compreensível porque se trata de vidas humanas, então uma decisão errada de um sistema totalmente autônomo – ou seja, um sistema que também toma a decisão final por conta própria – levanta questões legais com as quais a maioria dos países ainda não consegue lidar (quem é responsável para uma decisão errada? O engenheiro que implementou a ferramenta? O cientista de dados que a treinou?...). Um sistema automatizado pode impactar seriamente a vida das pessoas. Por isso, em áreas como essa, é importante ter um sistema que auxilie na tomada de decisão, mas onde o operador humano tome a decisão final. Além disso, devemos considerar que se o sistema não conseguir explicar as decisões sugeridas, é possível que o usuário não entenda e, por isso, ignore a sugestão. Outro importante desafio é manter e aumentar a disposição do usuário em interagir com o sistema técnico. Nesses casos, um sistema de iniciativa mista, que suporta interação humano-computador, torna-se útil. No contexto da Inteligência Híbrida (HI), ela requer que humanos e sistemas inteligentes trabalhem juntos, e um dos principais desafios para alcançar essa parceria é a capacidade dos agentes de entender os atores humanos.

A abordagem criada utiliza técnicas da Teoria da Argumentação e de Ontologia como base para o raciocínio em diálogos de linguagem natural. Usando essas técnicas, as decisões podem se tornar mais transparentes e explicáveis aos usuários, com estruturas de diálogo mais naturais e ricas. Além disso, essa abordagem é adaptável o suficiente para ser aplicada em vários domínios, mas as avaliações foram concentradas na área da saúde.

Nessa área, a gestão efetiva de leitos hospitalares tem sido foco de diversas pesquisas, como o modelo IMBEDS que utiliza redes neurais artificiais e teoria de valor multi-atributo para tomada de decisão [Grübler et al. 2018]; abordagem estatística e de mineração de dados [Teow et al. 2012]; assim como, um modelo de otimização com algoritmo evolutivo para alocação de leitos [e Oliveira et al. 2020]. Embora todos os trabalhos citados anteriormente busquem melhorar o gerenciamento dos leitos, eles não fornecem

interação de linguagem natural e não oferecem suporte à decisão para que o profissional tenha controle total sobre as alocações, nem as decisões são explicáveis. Além disso, ainda existem poucos estudos que aplicam de forma prática os modelos formais de raciocínio e diálogo baseados em argumentação para apoiar interações entre software e usuários humanos [Essers et al. 2018]. Ademais, faltam trabalhos que avaliem empiricamente tais modelos com usuários humanos.

### **3. Principais Contribuições**

Dentre as contribuições desta tese destaca-se: (i) identificação das formas como a comunidade científica tem utilizado técnicas de argumentação para alcançar inteligência artificial explicável em sistemas de diálogo incluídas em levantamento bibliográfico publicado em [Engelmann et al. 2022a]; (ii) desenvolvimento da estrutura MAIDS [Engelmann et al. 2023] para apoiar o desenvolvimento de sistemas de diálogo explicáveis baseados em agentes BDI para auxiliar humanos na tomada de decisões; (iii) introdução e formalização de uma base de crenças multipartes para uma linguagem de programação de agentes BDI e uma abordagem estruturada para diálogos onde os agentes discutem sobre as informações do componente principal da base de crenças, mas podem passar para sub-diálogos para discutir questões específicas relacionadas ao componente ontológico ou o componente ToM da base de crença multipartes; (iv) criação do framework Dial4JaCa [Engelmann et al. 2021b, Engelmann et al. 2021a] para permitir que agentes inteligentes se comuniquem com humanos através da interação em linguagem natural; (v) criação do framework Onto4JaCa para dar aos agentes inteligentes a capacidade de usar e gerenciar as informações contidas nas ontologias durante seus processos de raciocínio [Ferreira et al. 2022]; (vi) criação do framework RV4JaCa [Engelmann et al. 2022b] que suporta o uso de verificação em tempo de execução em sistemas multi-agentes desenvolvidos na plataforma JaCaMo [Boissier et al. 2020]; (vii) implementação de um sistema explicável baseado na estrutura do MAIDS para auxiliar na tomada de decisões sobre alocação de leitos hospitalares; (viii) avaliação do sistema criado a partir de dados hospitalares reais e com o auxílio de profissionais responsáveis pela alocação de leitos em um hospital.

### **4. Resumo dos Resultados e Discussões**

Nesta tese, foram fornecidas ferramentas para o desenvolvimento prático de sistemas inteligentes explicáveis, bem como sistemas de suporte à inteligência híbrida. Visando a comunicação entre humanos e agentes de software, construiu-se o Dial4JaCa, que funciona como uma ponte entre sistemas multi-agentes desenvolvidos com JaCaMo e plataformas de processamento de linguagem natural como Dialogflow e Rasa. O Onto4JaCa, por sua vez, foi desenvolvido para permitir que os agentes utilizem, durante seu processo de raciocínio, as informações contidas nas ontologias, bem como as inferidas pelos raciocinadores semânticos. Cabe destacar que ambos, Dial4JaCa e Onto4JaCa tem sido usados por outros colegas para o desenvolvimento de sistemas multi-agentes em outros domínios. Além disso, foi construído o RV4JaCa, que possibilita a verificação formal do comportamento do software em tempo de execução. Esses três módulos formam a base do MAIDS, o framework para desenvolver sistemas de diálogo intencional multi-agente, que pode ser usado, juntamente com agentes específicos de domínio, em diferentes domínios. Como esse framework é modular, ele pode ser utilizado em sua totalidade ou

apenas os módulos que atendam os requisitos de cada sistema a ser desenvolvido. Esse trabalho também inclui a formalização de uma nova estrutura de diálogo-sub-diálogo com a qual pode-se abordar questões ontológicas ou de teoria da mente e depois retornar ao assunto principal do diálogo.

Para avaliar a aplicabilidade prática do framework, foi desenvolvido um sistema multiagente que oferece suporte à alocação de leitos hospitalares. Além dos módulos do framework MAIDS, também foram desenvolvidas algumas ferramentas específicas para este domínio, como uma ontologia de domínio, um validador de planos PDDL (Planning Domain Definition Language), um otimizador e duas propriedades RML (Runtime Monitoring Language) a serem verificadas em tempo de execução. O sistema desenvolvido foi avaliado por profissionais especializados em alocação de leitos do Hospital São Lucas – PUCRS, que concordaram em apoiar na avaliação da abordagem. Foram realizadas duas avaliações distintas, a primeira, com dados sintéticos, para verificar se as funcionalidades oferecidas pelo sistema atendiam às necessidades do hospital. A segunda, utilizando dados reais do hospital, teve como objetivo avaliar a expressividade de um sistema desenvolvido com o framework MAIDS baseado nos desideratos de Cohen [Cohen 2019] para sistemas de diálogo orientados a tarefas. Como resultado das avaliações, pode-se concluir que, segundo os avaliadores, o sistema atende a todos os elementos dos desideratos.

Dentre as limitações encontradas pode-se citar a dificuldade em utilizar grandes quantidades de dados na base de crenças do agente. Motivo pelo qual, durante os testes com usuários, foi utilizada uma quantidade limitada de dados. Além disso, durante a pandemia de COVID-19 houve um longo período de isolamento social e alta demanda nos hospitais, o que acarretou em dificuldades de contato com o hospital e a necessidade de repensar parte da estrutura desenvolvida especificamente para a avaliação do caso de uso após o contato ter sido reestabelecido. Outras limitações mais técnicas que podem ser citadas são as especificidades do Dialogflow, principalmente em relação ao tempo limite de resposta, a necessidade de usar um padrão específico para desenvolver a ontologia que é utilizada pelos agentes inteligentes através do Onto4JaCa e o uso de um monitor para cada propriedade diferente a ser verificada utilizando o RV4JaCa. Embora ainda haja muito trabalho a ser feito, em seu estado atual, o framework já contribui para o aumento da sofisticação em IA explicável, inteligência híbrida e sistemas de diálogo humano-agente.

## References

- Akata, Z., Balliet, D., de Rijke, M., Dignum, F., Dignum, V., Eiben, G., Fokkens, A., Grossi, D., Hindriks, K., Hoos, H., et al. (2020). A research agenda for hybrid intelligence: Augmenting human intellect with collaborative, adaptive, responsible, and explainable artificial intelligence. *Computer*, 53(8):18–28.
- Boissier, O., Bordini, R. H., Hubner, J., and Ricci, A. (2020). *Multi-agent oriented programming: programming multi-agent systems using JaCaMo*. MIT Press.
- Cohen, P. (2019). Foundations of collaborative task-oriented dialogue: What’s in a slot? In *Proceedings of the Annual SIGdial Meeting on Discourse and Dialogue*, pages 198–209.
- de Oliveira Gabriel, V., Panisson, A. R., Bordini, R. H., Adamatti, D. F., and Billa, C. Z. (2020). Reasoning in BDI agents using toulmin’s argumentation model. *Theoretical Computer Science*, 805:76–91.

- e Oliveira, B., de Vasconcelos, J., Almeida, J., and Pinto, L. (2020). A simulation-optimisation approach for hospital beds allocation. *International Journal of Medical Informatics*, 141:1–11.
- Engelmann, D., Damasio, J., Krausburg, T., Borges, O., Cezar, L. D., Panisson, A. R., and Bordini, R. H. (2021a). Dial4jaca – a demonstration. In *Proceedings of the International Conference on Practical Applications of Agents and Multiagent Systems*, pages 346–350.
- Engelmann, D., Damasio, J., Krausburg, T., Borges, O., Colissi, M., Panisson, A. R., and Bordini, R. H. (2021b). Dial4jaca – a communication interface between multi-agent systems and chatbots. In *Proceedings of the International Conference on Practical Applications of Agents and Multiagent Systems*, pages 77–88.
- Engelmann, D., Damasio, J., Panisson, A. R., Mascardi, V., and Bordini, R. H. (2022a). Argumentation as a method for explainable ai: A systematic literature review. In *Proceedings of the Iberian Conference on Information Systems and Technologies*, pages 1–6.
- Engelmann, D. C., Ferrando, A., Panisson, A. R., Ancona, D., Bordini, R. H., and Mascardi, V. (2022b). Rv4jaca – runtime verification for multiagent systems. In *Proceedings of the Workshop on Agents and Robots for reliable Engineered Autonomy*, pages 23–36.
- Engelmann, D. C., Panisson, A. R., Vieira, R., Hübner, J. F., Mascardi, V., and Bordini, R. H. (2023). Maids — a framework for the development of multiagent intentional dialogue systems. In *Proceedings of the International Conference on Autonomous Agents & Multiagent Systems*, pages 1–9.
- Essers, K., Rogers, R., Sturt, J., Sklar, E. I., and Black, E. (2018). Assessing the posture prototype: A late-breaking report on patient views. In *Proceedings of the International Conference on Human-Agent Interaction*, pages 344–346.
- Ferreira, C. E. A., Panisson, A. R., Engelmann, D. C., Vieira, R., Mascardi, V., and Bordini, R. H. (2022). Explaining semantic reasoning using argumentation. In *Proceedings of the International Conference on Practical Applications of Agents and Multiagent Systems*, pages 153–165.
- Goldman, A. I. (2012). Theory of mind. *The Oxford handbook of philosophy of cognitive science*, 1:1–25.
- Gruber, T. R. (1995). Toward principles for the design of ontologies used for knowledge sharing? *International journal of human-computer studies*, 43(5-6):907–928.
- Grübler, M. d. S., da Costa, C. A., Righi, R., Rigo, S., and Chiwiacowsky, L. (2018). A hospital bed allocation hybrid model based on situation awareness. *Computers, Informatics, Nursing*, 36(5):249–255.
- Teow, K. L., El-Darzi, E., Foo, C., Jin, X., and Sim, J. (2012). Intelligent analysis of acute bed overflow in a tertiary hospital in singapore. *Journal of Medical Systems*, 36:1873–1882.