Conference track «40.17 Philosophy of Cognitive Sciences and Artificial Intelligence»

Cognitive-Pragmatic Architecture of Industrial AI: A Four-Dimensional Framework for Ethics and Compliance Based on Gricean Maxims

Academic supervisor – Нагорная Александра

Чэн Чан
Student (master)
National Research University Higher School of Economics, Школа иностранных языков, Moscow, Россия
E-mail: cchazy2021@gmail.com

As large language models (LLMs) become deeply integrated into industrial applications, the pragmatic competence of AI systems—their ability to communicate effectively, appropriately, and cooperatively in context—emerges as a critical determinant of successful human-AI collaboration [2, 5]. However, current AI ethics research predominantly focuses on universal principles such as fairness and transparency, often overlooking the differentiated demands that various industries place on AI pragmatic capabilities [2]. This study addresses the core question: How should AI systems adapt their communicative behavior according to industry-specific characteristics to achieve both effective interaction and regulatory compliance?

Integrating insights from cognitive science and linguistic pragmatics, this research employs Grice's Cooperative Principles—the maxims of Quantity, Quality, Relation, and Manner—as foundational analytical tools [2, 4, 5]. We propose a "cognitive-pragmatic architecture" for industrial AI, comprising four core dimensions operationalized as industry-specific demands: (1) Quantity Demand: the requirement for appropriate informativeness; (2) Quality Demand: the requirement for truthfulness and reliability [1]; (3) Relation Demand: the requirement for contextual relevance [5]; and (4) Manner Demand: the requirement for clarity and unambiguity [4]. By transforming these maxims into operationalizable demand profiles, we construct a four-dimensional analytical framework to guide the ethical design and compliance assessment of AI systems.

To validate the framework's explanatory power, we analyze three representative industries, drawing on case study methodologies from human-AI interaction research [4, 5]. Medical diagnosis exhibits high Quality and high Manner demands: diagnostic suggestions must be accurate and uncertainty clearly communicated to avoid clinical errors [1]. Financial services require high Quality and high Quantity: comprehensive yet accurate disclosures are essential for regulatory compliance [1]. Content generation prioritizes high Quantity and high Relation: users expect rich, contextually relevant output, while Quality demands may be appropriately relaxed [5]. This analysis reveals systematic cross-industry variation in how Gricean maxims are weighted, variation that should directly inform AI system design constraints and ethical norms.

The theoretical contributions are threefold. First, the framework extends Gricean maxims from descriptive pragmatic tools to a normative design framework, providing a foundation for industrial adaptation of AI pragmatic competence [2]. Second, the concept of "cognitive-pragmatic architecture" bridges cognitive science and AI system design, aligning with research on integrating cognitive architectures with LLMs for enhanced decision-making [3, 6]. Third, it grounds ethical discussions in industrial contexts, offering actionable pathways for "responsible AI" implementation [1]. Practically, this framework provides design guidelines and evaluation criteria for AI product managers, engineers, and policymakers, facilitating systems that align with both human cognitive expectations and industrial compliance requirements.

Future research should explore cross-cultural applicability of Gricean maxims [2] and examine

the dynamic evolution of AI pragmatic competence alongside model iterations [3, 6].

Keywords: Cognitive-Pragmatic Architecture; Gricean Maxims; Industrial AI; AI Ethics

Источники и литература

1) Kaas, M. H. L., & Habli, I. (2025). Assuring AI safety: fallible knowledge and the Gricean maxims. AI and Ethics, 5, 1467–1480. https://doi.org/10.1007/s43681-024-00490-x
2) Krause, L., & Vossen, P. T. J. M. (2024). The Gricean Maxims in NLP - A Survey. In Proceedings of the 17th International Natural Language Generation Conference (pp. 470–485). Association for Computational Linguistics.
3) Setlur, V., & Tory, M. (2022). How do you converse with an analytical chatbot? Revisiting Gricean maxims for designing analytical conversational behavior. In Proceedings of the CHI Conference on Human Factors in Computing Systems (pp. 1–17). ACM. https://doi.org/10.1145/3491102.3501972
4) Wu, S., Oltramari, A., & Ritter, F. E. (2025). VSM-ACTR 2: a human-like decision making model with metacognition for manufacturing solutions. Computational and Mathematical Organization Theory, 31(4), 259–276. https://doi.org/10.1007/s10588-025-09405-5
5) Kim, S., et al. (2025). Applying the Gricean Maxims to a Human-LLM Interaction Cycle: Design Insights from a Participatory Approach. arXiv preprint arXiv:2503.00858.
6) Wu, S., Oltramari, A., Francis, J., Giles, C. L., & Ritter, F. E. (2024). Cognitive LLMs: Towards Integrating Cognitive Architectures and Large Language Models for Manufacturing Decision-making. arXiv preprint arXiv:2408.09176.