

Structure, Meaning, and Answerability of Automated Systems

Mattias Brännström

Department of Computing Science
Umeå University, Sweden
mattias.brannstrom@umu.se

Abstract

This paper presents an ongoing PhD research project that re-frames transparency in AI systems from a property of models to a participatory relation between a system and the people interacting with it. Rather than asking whether a model is interpretable, the project asks what a system and its deployment context must be like for the demands of those interacting with it — their questions, their need to act, their accountability and governance requirements — to be genuinely satisfiable. This alignment between system structure and observer demand is formalized as semantic transparency, using refineable graph structures grounded in logic and inquisitive semantics. The framework has been validated in sociotechnical decision processes through close collaboration with public sector organizations on EU AI Act compliance and recruitment AI, producing formal results and practical tools for accountability and fairness analysis. Ongoing work extends the framework to the causal structure of intentional agency and knowledge formation as mediated through AI systems.

Motivation, Problem, and Approach

People and organizations interacting with, or acting through, AI systems need to understand the tools they are using — not as isolated models, but as components of processes that shape decisions, knowledge, and agency. Existing approaches to explainability and fairness analyze the model in isolation, separated from the sociotechnical context in which it operates (Selbst et al. 2019), producing post-hoc approximations that may be pedagogically useful without being structurally true (Arrieta et al. 2020; Ananny and Crawford 2018). The result is that transparency remains a surface property rather than a substantive one: systems can be declared explainable while the questions that matter to those affected remain structurally unanswerable.

This project reframes transparency from a property of models to a participatory relation between a system and the people interacting with it. Rather than asking whether a model is interpretable, I ask what a system and its deployment context must be like for the demands of those interacting with it — their questions, their need to act, their accountability requirements — to be genuinely satisfiable. I call

this alignment between system structure and observer demand *semantic transparency* (Brännström, Xanthopoulou, and Jiang 2025b; Brännström 2023), and it is the central concept unifying this research.

To formalize this, I take a unified structural approach to representation, knowledge, and agency applicable to humans, automated systems, and their interaction. The core machinery consists of refineable graph structures grounded in logic and inquisitive semantics (Ciardelli, Groenendijk, and Roelofsen 2018; Pearl 2009): information flow graphs representing how decisions form across human and automated components, and structural causal models representing how goals, knowledge, and actions relate over time. These structures support analysis of transparency as answerability at the level of sociotechnical systems, knowledge representation, and human-AI agency.

Work Accomplished

The Semantic Transparency framework (Brännström 2023; Brännström, Xanthopoulou, and Jiang 2025b) formalizes the conditions under which a system composed of opaque components is decomposable relative to a set of questions — establishing what structural properties are necessary for transparency demands to be satisfiable at all. Building on this, and on earlier work structuring how high-level normative requirements attach to socio-technical context (Brännström, Theodorou, and Dignum 2022), the Information Flow Methodology (IFM) (Brännström et al. 2023; Brännström, Xanthopoulou, and Jiang 2025a,c; Brännström and Xanthopoulou 2025) operationalizes transparency as answerability in hybrid human-AI settings through a participatory modeling method that captures both automated components and the organizational processes surrounding them, separating sources of bias from stakeholder impact. Validated in close collaboration with public-sector organizations, AI developers, and stakeholder groups across close to 40 interviews, IFM has been applied to EU AI Act compliance assessment and recruitment AI: in the latter, structurally identified bias pathways were quantitatively confirmed, revealing an intersectional gender-by-nationality effect invisible to aggregate analysis (Brännström, Xanthopoulou, and Jiang 2025a). This demonstrates that structural modeling can surface accountability failures that statistical analysis cannot reach.

The user is not a passive recipient of answers but an agent whose knowledge and intentions are themselves shaped by and effected through the tools they use. I have developed the Intentional Frames framework to address this directly, constructing a causal account of intentional reach and agency as mediated through AI systems — modeling the causal connections between epistemic states, internal models, and intentions across both the user and the AI system. This work started with a HumaneAI-Net micro-project I initiated and led on explainable agent behavior (Giménez-Abalos et al. 2026), has continued through ongoing collaborations and the CLEAR workshop series on relational transparency, and has motivated the Intermediate Knowledge Problem as a formal account of the epistemic conditions that coherent intentional agency requires. This provides a formal basis for reasoning about what kinds of understanding and control are structurally possible in a given human-AI interaction, and what governance or design interventions can expand them.

The same question of epistemic coherence arises within automated and hybrid reasoning systems themselves. The Intermediate Knowledge Problem (IKP) (Brännström, Xanthopoulou, and Brännström 2026b) identifies and formalizes this: in multi-step reasoning processes, intermediate representations must preserve coherence, grounding, and goal-alignment across the steps linking evidence to conclusion — a condition that existing systems, from symbolic pipelines to LLMs, systematically fail to meet. The Hybrid-Human Grounded Theory workflow (Brännström, Xanthopoulou, and Brännström 2026a) takes a first step toward addressing this, introducing epistemic graph structures that maintain traceable paths from source material to the formal representations built on them.

Current and Future Work

The Intentional Frames framework is being formalized and extended through several ongoing projects: a study of frictional AI design in Human-in-the-Loop scenarios (accepted, HHAI workshop), the Keyhole Maze — a structured reinforcement learning environment applying the framework to agent-observing-agent scenarios (in preparation) — and a formal treatment of how observer mental models determine the functional role of an AI system, from opaque oracle to internalized tool (in preparation). A parallel effort is unifying the formal language of IF graphs and SCMs to enable coherent analysis across both decision processes and agent behavior. The expected outcome is a unified account of transparency as a participatory relation — structural, formally grounded, and applicable across the range of ways people and organizations interact with and through AI systems.

References

- Ananny, M.; and Crawford, K. 2018. Seeing without knowing: Limitations of the transparency ideal and its application to algorithmic accountability. *New Media & Society*, 20(3): 973–989.
- Arrieta, A. B.; Díaz-Rodríguez, N.; Del Ser, J.; Benetot, A.; Tabik, S.; Barbado, A.; Garcia, S.; Gil-López, S.; Molina, D.; Benjamins, R.; et al. 2020. Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. *Information Fusion*, 58: 82–115.
- Brännström, M. 2023. *Transparency of Complex Systems: The Semantic Transparency Framework*. Bachelor's thesis, Umeå University.
- Brännström, M.; Jiang, L.; Tubella, A. A.; and Dignum, V. 2023. Impact Based Fairness Framework for Socio-Technical Decision Making. In *Proceedings of the AEQUITAS Workshop on Fairness and Bias in AI at ECAI 2023*.
- Brännström, M.; Theodorou, A.; and Dignum, V. 2022. Let it RAIN for Social Good. In *Proceedings of the IJCAI-ECAI 2022 Workshop on Artificial Intelligence Safety (AISafety 2022)*.
- Brännström, M.; and Xanthopoulou, T. D. 2025. Information Flow Model (IFM): Methodological Guide. Zenodo, DOI: 10.5281/zenodo.16252143.
- Brännström, M.; Xanthopoulou, T. D.; and Brännström, A. 2026a. Hybrid-Human Grounded Theory with Multi-Step Epistemic Continuity. In *Proceedings of the 5th International Conference on Hybrid Human-Artificial Intelligence (HHAI 2026)*. To appear.
- Brännström, M.; Xanthopoulou, T. D.; and Brännström, A. 2026b. The Intermediate Knowledge Problem. In *Foundations of Information and Knowledge Systems (FoIKS 2026)*, volume 16475 of *Lecture Notes in Computer Science*, 334–340. Springer.
- Brännström, M.; Xanthopoulou, T. D.; and Jiang, L. 2025a. Modeling Fairness in Recruitment AI via Information Flow. In *Proceedings of the AEQUITAS Workshop on Fairness and Bias in AI at ECAI 2025*.
- Brännström, M.; Xanthopoulou, T. D.; and Jiang, L. 2025b. Transparency as Answerability: Role-Dependent Information Flow Modeling. Under review, *ACM Journal on Responsible Computing*.
- Brännström, M.; Xanthopoulou, T. D.; and Jiang, L. 2025c. Transparent Information Flow Modeling for Role-Aware AI Act Assessment. In *Proceedings of the CLEAR @ ECAI 2025*.
- Ciardelli, I.; Groenendijk, J.; and Roelofsen, F. 2018. *Inquisitive Semantics*. Oxford University Press.
- Giménez-Abalos, V.; Tormos, A.; Vázquez-Salceda, J.; Edström, F.; Brännström, M.; Lindqvist, J.; Cortés, U.; and Álvarez-Napagao, S. 2026. Unifying Explainable Agency via a Ladder of Intentions. *Information Systems Frontiers*, 1–26.
- Pearl, J. 2009. *Causality: Models, Reasoning, and Inference*. Cambridge University Press, 2nd edition.
- Selbst, A. D.; Boyd, D.; Friedler, S. A.; Venkatasubramanian, S.; and Vertesi, J. 2019. Fairness and Abstraction in Sociotechnical Systems. In *Proceedings of the Conference on Fairness, Accountability, and Transparency (FAccT)*, 59–68.