Precedent-Based Professional Role Ethics for AI Decision Analysis

Christopher Rauch

Drexel University cr625@drexel.edu

Abstract

Large language models (LLMs) are increasingly being used in professional fields such as healthcare, law, and engineering. In these domains, errors can lead to serious consequences. Many current AI ethics approaches do not reflect the structured codes and precedent-informed reasoning that guide professional conduct. This work introduces ProEthica, a system under development that combines LLMs with role-based ontologies to support structured ethical reasoning in professional settings. The system draws on principles from professional role ethics and incorporates ethical guidelines, practice standards, and prior case decisions. As a demonstration case, it applies the National Society of Professional Engineers (NSPE) Code of Ethics and Board of Ethical Review precedents. ProEthica includes an ontology based on engineering ethics, a precedent retrieval method using both vector similarity and ontological mappings, a framework for guiding and checking LLM outputs with structured constraints, and a validation process modeled on FIRAC (Facts, Issues, Rules, Analysis, Conclusion) reasoning. The system is intended to help professionals make ethical decisions that are consistent with established standards, not to replace human judgment. Preliminary evaluations using NSPE cases indicate that it can retrieve relevant precedents and produce structured analyses that align with engineering ethics.

Introduction

Professional ethics often draws on case-based reasoning that applies precedent to current circumstances using the principle of consistent treatment (Ashley and McLaren 1995). Although AI ethics research has advanced, it remains largely disconnected from the codified standards and reasoning practices used by professionals in the regulated fields mentioned above. This gap arises partially from the difficulty of translating high-level principles into enforceable constraints and functional tools. Efforts to bridge this gap are hindered by the lack of alignment between AI systems and the institutional norms of professional practice.

Recent studies identify key limitations in current AI tools. They cannot effectively enforce domain-specific standards without structured knowledge (Puri 2020), they often lack support for analogical reasoning with precedent, although some CBR systems provide partial solutions (Ashley and McLaren 1995; McLaren 2003), and few tools provide mechanisms to validate consistency between ethical constraints and generated outputs. In professional settings, decision making must furthermore conform to institutional codes and legal frameworks that typical LLMs do not understand natively.

Related Work

Computational ethics approaches have typically followed three broad traditions: rule-based systems grounded in deontological ethics, optimization methods aligned with consequentialist frameworks, and character-based modeling inspired by virtue ethics (Boddington 2017; Allen, Smit, and Wallach 2005; Wallach and Allen 2009). Although these paradigms provide useful starting points, they do not reflect the structured, precedent-informed reasoning common in professional domains, where ethical principles must be considered contextually and are not readily applied through deductive approaches alone.

Earlier systems such as SIROCCO and MedEthEx modeled ethical reasoning through principles and cases, but lacked integration with large language models or generalizable ontologies (McLaren 2003). SIROCCO employed an engineering ethics ontology tailored to its domain, and MedEthEx focused on constrained biomedical cases. Both systems predate current LLMs and were designed to assist human decision-making rather than reach autonomous ethical judgments.

Analogical reasoning has also been central to systems such as SIROCCO and Truth-Teller, which use structured comparisons to highlight ethically relevant similarities and differences between cases (Ashley and McLaren 1995; McLaren 2003). These implementations demonstrate the value of precedent-based reasoning, although they do not address integration with modern LLM pipelines or structured constraint validation.

Recent reviews emphasize that fine-tuned LLMs continue to struggle with domain-specific standards and require human oversight due to inconsistent outputs and the absence of robust validation mechanisms (Haltaufderheide and Ranisch 2024; Prem 2023). LLMs may refer to professional codes, but often lack procedural safeguards to ensure output consistency or alignment with institutional frameworks. Efforts such as SPECTRA have proposed testbeds to evaluate ethi-

cal models across paradigms, suggesting additional methods for implementation that can support transparency and cross-model consistency (Chhabra et al. 2024).

Methodology

This work applies concepts from professional role ethics and structured decision-making to develop an LLM-supported system for ethical reasoning in professional settings. ProEthica combines a modular ontology based on engineering ethics with a precedent retrieval process that uses both vector similarity and ontological mapping to find relations between concepts. The approach is informed by prior arguments that ethical reasoning requires context-sensitive, structured representations. It also addresses calls for better alignment between the outputs of ethical AI systems and professional codes.

The precedent retrieval process is designed more specifically to address some of the shortcomings observed in current LLMs, including hallucinated content and inconsistent output (Haltaufderheide and Ranisch 2024). To improve output reliability, the system uses FIRAC, a legal reasoning framework, for structured validation (Casellas 2011). Further, professional norms are encoded as constraints that are used to verify that the generated outputs meet domain-specific expectations. This structure promotes the desirable application of patterns seen in work on constitutional AI and reinforcement-based ethical systems (Bai et al. 2022; Abel, MacGlashan, and Littman 2016).

ProEthica incorporates causal reasoning and connects conclusions to identifiable precedent cases. The system avoids relying exclusively on top-down principles or data-driven models by combining explicit codes with analogical retrieval. This hybrid design is intended to maintain interpretability and ensure the relevance of retrieved results (Allen, Smit, and Wallach 2005).

Preliminary Results

Initial testing suggests that ProEthica can identify relevant precedents and apply structured decision-making principles to generate ethical evaluations. Compared to baseline LLM outputs, evaluations indicate improved alignment with domain-specific standards, which are similarly observed in systems that integrate ethical knowledge with constraint enforcement or reinforcement-based feedback (Bai et al. 2022; Abel, MacGlashan, and Littman 2016).

The incorporation of professional codes appears to support consistent ethical responses, aligning with findings that emphasize the value of converting principles into rules and validation procedures (Prem 2023). Precedent-based case matching contributes to improved reasoning quality, a strategy validated by previous systems using analogical comparison (Ashley and McLaren 1995).

Structured validation mechanisms enable traceable output review. Constraint-based evaluation frameworks clarify model behavior and support consistent decision-making (Chhabra et al. 2024). These early findings suggest that ProEthica's approach supports alignment with professional ethics, as well as greater transparency and evaluability.

Ongoing Work and Next Steps

Development is ongoing to extend the ProEthica ontology to include additional professional domains of healthcare and law practice. Refinements to the precedent retrieval algorithms will ideally improve sensitivity to context-specific elements in ethical cases. A comparative evaluation protocol is being constructed to assess system performance using expert-annotated baselines and metrics that include consistency, explainability, and alignment with domain codes and cases. Integration with interactive LLM prompting is being evaluated to support user-guided refinement of ethical justifications. Planned pilot studies will investigate usability and practitioner trust, while enhancements to causal reasoning modules are intended to improve the system's ability to model normative dependencies. These efforts are intended to support generalization across domains and ensure practical applicability in professional decision-making.

Conclusion

ProEthica demonstrates a method for operationalizing structured, precedent-based ethics within AI systems. It integrates professional codes, analogical reasoning, and constraint validation to support decision-making aligned with institutional standards. The system is designed to assist rather than replace human judgment, offering a practical framework for incorporating ethical considerations into professional workflows.

References

Abel, D.; MacGlashan, J.; and Littman, M. L. 2016. Reinforcement Learning as a Framework for Ethical Decision Making. In *AAAI Workshop: AI, Ethics, and Society*.

Allen, C.; Smit, I.; and Wallach, W. 2005. Artificial Morality: Top-down, Bottom-up, and Hybrid Approaches. *Ethics and Information Technology*, 7(3): 149–160.

Ashley, K. D.; and McLaren, B. M. 1995. A CBR Knowledge Representation for Practical Ethics. In Haton, J.-P.; Keane, M.; and Manago, M., eds., *Advances in Case-Based Reasoning*, 180–197. Springer.

Bai, Y.; Kadavath, S.; Kundu, S.; Askell, A.; Kernion, J.; Jones, A.; Chen, A.; Goldie, A.; Mirhoseini, A.; McKinnon, C.; Chen, C.; Olsson, C.; Olah, C.; Hernandez, D.; Drain, D.; Ganguli, D.; Li, D.; Tran-Johnson, E.; Perez, E.; Kerr, J.; Mueller, J.; Ladish, J.; Landau, J.; Ndousse, K.; Lukosuite, K.; Lovitt, L.; Sellitto, M.; Elhage, N.; Schiefer, N.; Mercado, N.; DasSarma, N.; Lasenby, R.; Larson, R.; Ringer, S.; Johnston, S.; Kravec, S.; Showk, S. E.; Fort, S.; Lanham, T.; Telleen-Lawton, T.; Conerly, T.; Henighan, T.; Hume, T.; Bowman, S. R.; Hatfield-Dodds, Z.; Mann, B.; Amodei, D.; Joseph, N.; McCandlish, S.; Brown, T.; and Kaplan, J. 2022. Constitutional AI: Harmlessness from AI Feedback.

Boddington, P. 2017. Codes of Professional Ethics. In Boddington, P., ed., *Towards a Code of Ethics for Artificial Intelligence*, 39–57. Cham: Springer International Publishing.

Casellas, N. 2011. *Legal Ontology Engineering: Methodologies, Modelling Trends, and the Ontology of Professional Judicial Knowledge*. Springer Science & Business Media.

Chhabra, J.; Sama, K.; Deshmukh, J.; and Srinivasa, S. 2024. Evaluating Computational Models of Ethics for Autonomous Decision Making. *AI and Ethics*.

Haltaufderheide, J.; and Ranisch, R. 2024. The Ethics of ChatGPT in Medicine and Healthcare: A Systematic Review on Large Language Models (LLMs). *npj Digital Medicine*, 7(1): 1–11.

McLaren, B. M. 2003. Extensionally Defining Principles and Cases in Ethics: An AI Model. *Artificial Intelligence*, 150(1-2): 145–181.

Prem, E. 2023. From ethical AI frameworks to tools: a review of approaches. *AI and Ethics*, 3.

Puri, A. 2020. Moral Imitation: Can an Algorithm Really Be Ethical? *Rutgers Law Record*, 48: 47–58.

Wallach, W.; and Allen, C. 2009. *Moral Machines: Teaching Robots Right from Wrong*. Oxford University Press.