“What if AI could reason with values the way we do—with nuance, with context, under pressure?”

AIM

MORAL CODE

Advancing Ethical AI

ICAD 2025 – IEEE Presentation

Presenter: Ran Hinrichs

A Methodology and Empirical Approach to the AI Moral Code

Ethical AI remains fragmented despite proliferation of guidelines

291+ global documents analyzed (2006–2025)

Extracted 12 canonical values through semantic and sectoral analysis

Values stratified and weighted across domains

AI Moral Code offers unified, testable framework

Grounded in global philosophy, built for governance

Background and Motivation

- AI is embedded in high-stakes decision-making
- Ethical risks persist across contexts and sectors
- Global guidelines show value convergence (2018–2020)
- Key frameworks: IEEE, OECD, Jobin, Fjeld, Floridi
- Ethical principles lack consistent implementation
- This project builds a systematized moral canon

Stratifying Ethical Priorities for Impact

- 12 values derived through empirical clustering
- Stratified into Core, Instrumental, and Conditional tiers
- Core: Ethical foundations (e.g., trust, dignity)
- Instrumental: Enable innovation and sustainability
- Conditional: Contextual values—privacy, autonomy, inclusivity
- Structured via NRBC framework: Normative → Conceptual

	Low Complexity (Static Inputs)	**High Complexity (Dynamic, Multistakeholder Inputs)**
Low Ethical Tension (Few conflicting values)	**Education AI** Learner autonomy vs algorithmic classification	**Cybersecurity IR** Transparency vs national security, human override vs AI
High Ethical Tension (Many conflicting values)	**Health Care Diagnostics** Trust vs uncertainty vs. privacy	**Climate Modeling** Uncertainity disclosure vs. panic vs power dynamics **Autonomous Vehicles** Passenger safety vs pedestrian harm tradeoff

Results – Weighted & Normalized

- 8% threshold = ethical convergence point
  a. **Represents statistically significant cross-sector agreement**
- Conditional values dominate ethical discourse
- Privacy, autonomy show sector-specific weighting
- Fairness & transparency diluted by overuse
- Government, NGO documents shape value gravity
- Sector Weight Index reveals hidden influence

Scaling Toward Agent-Level Moral Reasoning

- Semantic NLP applied to ethics detection
- Moving beyond exact match to fuzzy logic
- Scalable extraction across AI ethics documents
- Adaptive weighting for domains like cybersecurity
- Live ethical auditing & simulation validation
- Toward agent-level moral reasoning frameworks

Summary

Corpus → Value Extraction → Stratification → Weighting → Simulation Testing → Future Scaling

Closing and Contact

- AI ethics requires structure, not slogans
- Moral codes must be testable, adaptive, alive
- This framework bridges theory and application
- Grounded in global values, built for systems
- Join the movement toward ethical alignment
- Visit: aimoralcode.org Book forthcoming

aimoralcode.org

Ran Hinrichs

rhinrich@norwich.edu

Questions

Advancing Ethical AI

AIM

MORAL CODE

“Built with thanks to global ethics researchers and shaped by feedback from human-AI dialogues.”