GEORG-AUGUST-UNIVERSITÄT GÖTTINGEN

HPS

Ethical AI: Towards Defining a Collective Evaluation Framework

Aasish Kumar Sharma[1], *Dimitar Kyosev*[2], *Julian Kunkel*[1]

Georg-August-Universität, Göttingen, Germany[1], & *Alis Grave Nil Private Limited, Bulgaria*[2]

Outline

IEEE Paper

Ethical AI: Towards Defining a Collective Evaluation Framework

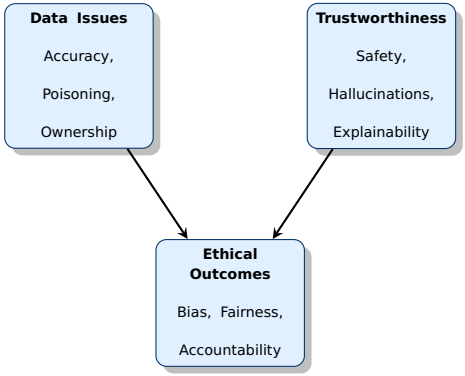
Published in: 49th Annual IEEE Computers, Software, and Applications Conference
DOI: 10.1109/COMPSAC65507.2025.00344

The AI Ethics Challenge

AI's Transformative Potential:

- Healthcare innovation
- Financial services
- Autonomous systems
- Enhanced productivity

Emerging Ethical Concerns:



Core Problem

- To cope with changes,
  - Guidelines and Legislation keep on changing.
  - AI systems often operate as "black boxes":
    - Generating decisions for critical digital infrastructure,
    - May not consider ethical considerations.

Research Motivation

- Need for comprehensive frameworks that:
  - Embed ethical principles directly into AI systems from design to deployment

Research Objectives

This work addresses key challenges by proposing:

1. **Unified Observation**
   → For AI outcomes aligned with EU AI Act and global regulations

2. **Ontological Blocks Evaluation**
   → Feasibility of using ontological blocks of meaning for ethical assessment

3. **FAIR Principles Integration**
   → Incorporating Findable, Accessible, Interoperable,
   → Reusable principles into ontological frameworks

Goal: Transparent, accountable, and fair AI systems

Research Methodology

Methodological Framework

→ **Type:** Descriptive research with qualitative observation
→ **Focus:** Exploring ethical AI pathways through synthesis

Key Variables Examined:

- **Ethical Outcomes** – Benchmarks set by experts in ethics, AI, and policy
- **Verification Methods** – Approaches to automate assessments against benchmarks
- **Scalability** – Applicability across varied AI systems and domains

Analyzed

- Synthesize literature, expert insights, and theoretical models:
  - To propose an actionable frameworks rather than empirical simulations

Ethical AI Landscape

Frameworks & Principles:

- WHO's 6 core principles
- EU AI Act regulations
- IEEE ethical guidelines
- Responsible Research & Innovation (RRI)

Technical Challenges:

- Deep RL as closed-box systems
- Limited explainability (XAI)
- Human-in-the-loop integration
- Real-time assessment needs

Current Limitations:

- Lack technical grounding
- Static, not dynamic
- Domain-specific gaps

State-of-the-Art:

- SHAP[a] & LIME[b] for interpretability
- Bias mitigation techniques
- Risk-based regulations

[a] SHAP → SHapley Additive exPlanations
[b] LIME → Local Interpretable Model-agnostic Explanations

Use Cases: Domain-Specific Ethical Concerns

Domain	**Key Concerns**	**Identified Gaps**
Finance	Credit scoring bias → Discriminatory outcomes	Independent bias detection tools needed
Healthcare	Patient privacy → Treatment fairness	Real-time explainable AI scalability challenges
Autonomous Vehicles	Safety decisions → Accountability	Real-time ethical decision-making

Common Thread

→ All domains require scalable, transparent, and auditable ethical assessment mechanisms

Framework: Ontological & FAIR Approaches

Ontological Models:

- Semantic Web standards (W3C)
- Structured ethical representations
- Scalable principle encoding
- Examples: Medical ontologies
  - National Cancer Institute Thesaurus,
  - Cancer Care

FAIR Principles:

- **F**indable: Persistent identifiers
- **A**ccessible: Open protocols
- **I**nteroperable: Standard vocabularies
- **R**eusable: Clear licensing

Application:

→ Support openness &
→ Standardization of ethical modules.

Challenge:

→ Manual effort for building &
→ Maintaining interoperable frameworks.

What are Ontological Blocks of Meaning?

Definition

→ *Ontological blocks* are discrete, machine-readable constructs,
→ That encode specific ethical principles while preserving their philosophical meaning.

Key Characteristics:

- Modular design
- Machine-interpretable
- Philosophically grounded
- Domain-agnostic
- Composable

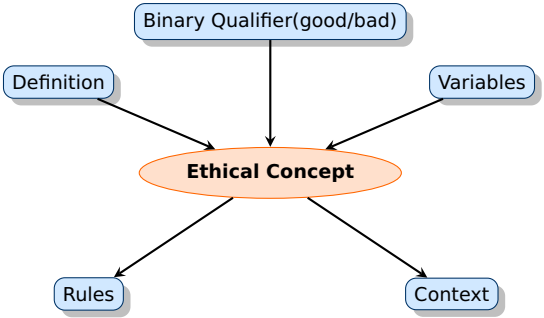
Creation Process:

1. Translate ethical concepts
2. Structure representations
3. Preserve ethical meaning
4. Ensure scalability
5. Enable precision

Analogy

→ Like evolving definitions in physics (Newton to Quantum),
→ Ethical terms must be redefined for AI while preserving core meaning

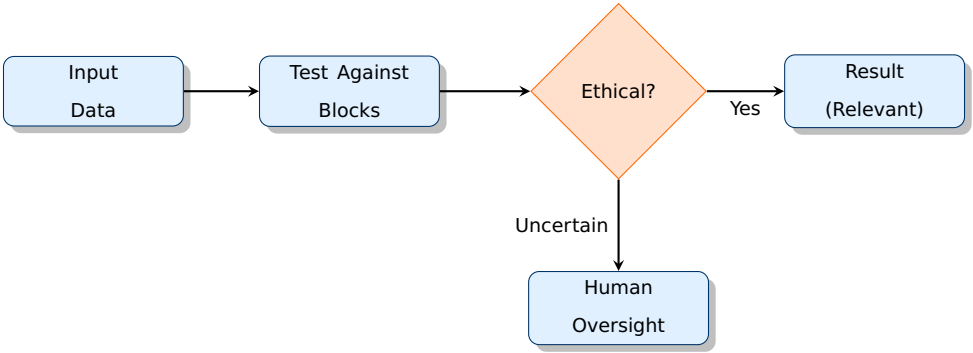
Ontological Block Structure



Example: "Stealing is bad"

- *Concept:* Unauthorized ownership transfer
- *Qualifier:* Bad (negative)
- *Variables:* Property, ownership, legitimate pathways

How It Wroks - Block Processing Workflow



Key Features:

- Generalization over rule-based examples
- Inference from incomplete data
- Human intervention for weak cross-references

Use Case: AI-Powered Investor Profiling

Context: Brokerage firms using AI voice assistants for investor classification

The Challenge

- Regulations require to distinguish:
  - Professional vs. retail investors to protect the latter from high-risk products,
  - But current binary systems overlook real-world behavioral risk tolerance.

Innovation: Ontological Product Classification

→ Financial products trigger "Riskier" blocks based on:

- *Financial risk* – Potential economic loss
- *Psychological risk* – Emotional reaction to stress

Enables real-time, context-aware, ethically grounded assessment

Comparison with Existing Approaches

Feature	Our Framework	SHAP/LIME	Rule-Based
Data Independence	Yes	No	Yes
Domain Flexibility	Yes	Limited	No
Explainability	Yes	Yes	Yes
Auditability	Yes	Limited	Yes
Real-time Capable	Yes	Yes	Yes
Automated Creation	Yes	Yes	No
Human-In-Loop	Yes	No	No

Unique Contribution

→ Open sourced framework with data independence, domain flexibility, & structured auditability.

Final Remarks

Core Message

- Ontological blocks offer a promising path toward explainable and auditable AI ethics,
- Balancing philosophical rigor with practical applicability.

Challenges Ahead:

- Scaling block creation
- Automating processes
- Community adoption
- Standardization efforts

Opportunities:

- Cross-sector collaboration
- Integration with existing frameworks
- Open-source community development
- Global standardization potential

The future of AI must be ethically grounded, transparent, and accountable

Acknowledgments

We thank:

Gesellschaft für wissenschaftliche Datenverarbeitung mbH Göttingen (GWDG)

For valuable contributions and support

NHR (National High Performance Computing)

Essential infrastructure for this work

Funding:

National KISSKI Project

Grant Number: 01-S22093A (Förderkennzeichen)

Special thanks to our reviewers and colleagues for constructive feedback

"Building Ethical AI Is Not A One-time Achievement,
Instead an ongoing commitment to transparency,
accountability, and collective progress."!

Questions & Discussion

Aasish Kumar Sharma
aasish-kumar.sharma@gwdg.de
ORCID: 0000-0002-7514-2340

Julian Kunkel
julian.kunkel@gwdg.de
ORCID: 0000-0002-6915-1179

Georg-August-Universität Göttingen

IEEE COMPSAC 2025