

Ethical AI: Towards Defining a Collective Evaluation Framework

1st Aasish Kumar Sharma
Mathematics and Computer Science
Georg-August-Universität Göttingen
 Göttingen, Germany
 aasish-kumar.sharma@gwdg.de,
 orcid: 0000-0002-7514-2340

2nd Dimitar Kyosev
Legal Affairs
Alis Grave Nil Private Limited
 Burgas, Bulgaria
 kyosev.dimitar@gmail.com

3th Julian Kunkel
Mathematics and Computer Science
Georg-August-Universität Göttingen
 Göttingen, Germany
 julian.kunkel@gwdg.de,
 orcid: 0000-0002-6915-1179

Abstract—Artificial Intelligence (AI) is transforming sectors such as healthcare, finance, and autonomous systems, offering powerful tools for innovation. Yet its rapid integration raises urgent ethical concerns related to data ownership, privacy, and systemic bias. Issues like opaque decision-making, misleading outputs, and unfair treatment in high-stakes domains underscore the need for transparent and accountable AI systems.

This article addresses these challenges by proposing a modular ethical assessment framework built on *ontological blocks of meaning*—discrete, interpretable units that encode ethical principles such as fairness, accountability, and ownership. By integrating these blocks with FAIR (Findable, Accessible, Interoperable, Reusable) principles, the framework supports scalable, transparent, and legally aligned ethical evaluations, including compliance with the EU AI Act.

Using a real-world use case in AI-powered investor profiling, the paper demonstrates how the framework enables dynamic, behavior-informed risk classification. The findings suggest that ontological blocks offer a promising path toward explainable and auditable AI ethics, though challenges remain in automation and probabilistic reasoning.

Index Terms—Responsible AI, Ethical AI, Machine Ethics, AI Acts, Explainability, Ontology, Workflows, Transparency.

I. INTRODUCTION

Artificial Intelligence (AI) has emerged as a transformative technology, revolutionizing fields such as healthcare, finance, and autonomous systems. While AI offers significant benefits, including enhanced productivity and innovative solutions to global challenges, its rapid evolution also introduces profound ethical challenges that require immediate and thorough attention.

One of the foundational challenges lies in the data used to train AI systems. AI relies heavily on vast datasets sourced from the internet, private organizations, and other repositories. However, issues related to data ownership, consent, and privacy remain unresolved. Questions such as “Who owns the data?” and “Was it ethically obtained?” highlight potential risks of misuse and exploitation, raising concerns about intellectual property rights and individual privacy.

Another pressing issue is trustworthiness. AI systems occasionally produce misleading or incorrect outputs, a phenomenon known as “hallucinations.” Moreover, ensuring the safety of AI systems and maintaining detailed records of their

decision-making processes are critical for fostering trust. Without robust mechanisms for safety and accountability, reliance on AI systems could lead to unforeseen risks, particularly in high-stakes domains such as healthcare, finance, and law enforcement.

However, the most profound challenges arise from the ethical implications of AI’s outcomes. AI systems, often operating as “black boxes”, generate decisions and recommendations with far-reaching consequences. The absence of embedded ethical considerations can result in biased, unfair, or even harmful outcomes. For example:

- Biased algorithms in hiring or lending processes can perpetuate societal inequalities.
- Unethical use of AI in surveillance can infringe on civil liberties.
- Lack of explainability can erode trust, particularly in life-critical applications.

Addressing these challenges requires a collaborative effort among technologists, policymakers, ethicists, and the public. Establishing robust frameworks for ethical AI development and deployment is imperative to ensure AI aligns with human values and serves humanity equitably and responsibly.

A. Key Ethical Challenges in AI

- *Data-Related Issues*: These include questions around data accuracy, inclusion of purposefully misleading data or even data poisoning
- *Trustworthiness*: AI systems must be reliable, safe, and explainable to avoid risks like hallucinations and over reliance on black-box outcomes.
- *Ethical Implications of Outcomes*: The potential for biased or harmful decisions underscores the need to embed ethical principles directly into AI systems.

As AI continues to integrate into society, the task of addressing these ethical dilemmas grows increasingly complex. While the challenges are significant, they also present an opportunity to build a future where AI aligns with the highest ethical standards. By fostering collaboration and establishing comprehensive frameworks, we can ensure that AI development progresses in a way that prioritizes transparency, fairness, and accountability.

The article is organized in five sections: Introduction (Section I), background (Section II), literature review (Section III), methodology and discussion (Section IV), then includes conclusion and recommendations (Section V).

II. BACKGROUND

Ethical AI refers to the development and deployment of artificial intelligence systems that emphasize fairness, transparency, accountability, and respect for fundamental rights [1]. AI ethics emphasises the impact of AI on individuals, groups and wider society. The goal is to promote safe and responsible AI use, to mitigate AI's novel risks and prevent harm. Much of the work in this area centres around four main verticals: Frameworks such as the six core principles of the WHO [2], the EU AI Act [3] highlight the importance of protecting autonomy, promoting equity, and fostering responsibility. Despite these efforts, ethical frameworks often lack technical grounding and do not address the dynamic nature of AI systems [1], [4]–[8].

For example, Reinforcement-Learning (RL), particularly deep reinforcement learning, operates as a closed-box system, posing challenges in explainability [9]–[11]. Human-in-the-loop (HITL) AI approaches [12] provide a potential solution by integrating human oversight at critical stages, but their implementation in Self Reinforcement Learning (SLR) remains limited.

Recent studies highlight several state-of-the-art advances in ethical AI and SRL. Mehrabi et al. [13] explore bias mitigation techniques such as data cleaning, model adjustments, and output tuning. [14], [15] discusses interpretability approaches such as SHAP and LIME, which enhance explainability in complex models. Human-in-the-loop systems [12] integrate ethical oversight effectively, while the EU AI Act [3] offers risk-based regulations tailored for high-stakes applications.

A. Objectives

This article addresses key challenges in ethical AI and aims to:

- 1) Propose a unified system for monitoring AI outcomes in line with the EU AI Act and emerging global regulations.
- 2) Evaluate the feasibility of using ontological blocks of meaning for ethical assessment.
- 3) Explore the integration of FAIR (Findable, Accessible, Interoperable, Reusable) principles into these ontological blocks.

Together, these objectives support the development of ethical AI frameworks that promote transparency, accountability, and fairness [16].

III. LITERATURE REVIEW

The rising importance of ethical Artificial Intelligence (AI) has driven researchers and institutions to develop frameworks, methods, and applications that promote transparency, accountability, and fairness. This review is organized into subsections covering finance, healthcare, autonomous vehicles, ontological and FAIR approaches, and global initiatives.

A. Ethical AI in Finance

AI-driven credit scoring in finance raises ethical concerns about bias and unequal treatment.

- Hassani [17] shows how societal biases in training data reinforce discrimination in credit assessments, disproportionately affecting marginalized groups.
- Packin [18] warns that lack of comprehensive data on disabled individuals can lead to exclusionary outcomes in AI-based scoring.

Identified Gap: Independent tools like ontological blocks are proposed to detect and mitigate embedded bias in financial AI systems.

B. Ethical AI in Healthcare

Healthcare—particularly oncology—illustrates the need for ethical AI frameworks to safeguard privacy, transparency, and fairness [19].

- Ontological blocks have been applied to sensitive patient data under FAIR principles, aiding in personalized treatment planning [20], [21].
- Monteith et al. [22] caution against misinformation in AI mental health tools, stressing the need for explainability and oversight.
- Antoniou et al. [23] highlight the need for explainable AI to manage complex diagnoses involving comorbidities.

Identified Gap: Real-time, scalable use of explainable AI methods (e.g., SHAP, LIME) remains a key challenge in high-stakes clinical settings [9].

C. Ethical AI in Autonomous Vehicles

Autonomous vehicles demand AI systems that prioritize safety, accountability, and ethical reasoning.

- Lin [24] and Jenkins et. al. [25] examines scenarios like the trolley problem, calling for transparent ethical decision-making in critical moments.
- Goodall [26] advocates for programming rules that minimize harm during unavoidable accidents.
- Nyholm and Smids [27] argue for clear accountability frameworks among manufacturers, users, and regulators.

Identified Gap: Ensuring real-time ethical decisions and system adaptability remains a core technical and ethical challenge.

D. Ontological and FAIR Approaches for Ethical AI

Ontological models and FAIR principles are emerging as foundational tools for ethical AI design.

- Ontological blocks, drawing from Semantic Web standards [28], [29], provide scalable representations of principles like non-maleficence and equity [2].
- FAIR principles [7]—Findable, Accessible, Interoperable, and Reusable—support the openness and standardization of ethical modules.
- Guizzardi et al. [30] emphasize the need for ontologies that model user-centric concerns like privacy, fairness, and risk.

Challenges: Implementing FAIR-aligned ontological blocks is constrained by the manual effort needed to build and maintain interoperable frameworks.

E. Global and Collaborative Efforts

International collaborations aim to mitigate societal-scale risks and build governance frameworks for ethical AI [8], [31], [32].

- The EU champions “ethics-by-design,” embedding ethics at the development stage [7].
- An international consortium proposes benchmarking societal risks, promoting transparency and cooperative oversight [33].
- Responsible Research and Innovation (RRI) offers a governance model that aligns AI with societal values [34]–[38].

Future Direction: These efforts underscore the importance of global standards and shared accountability mechanisms for ethical AI deployment.

IV. METHODOLOGY AND DISCUSSION

A. Methodology

This article uses descriptive research and qualitative observation to explore ethical AI pathways. Key variables include: *Ethical outcomes* – Benchmarks set by experts in ethics, AI, and policy; *Verification* – Methods to automate assessments against those benchmarks; *Scalability* – Applicability across varied AI systems and domains.

Rather than conducting empirical simulations, the study synthesizes literature, expert insights, and theoretical models to propose actionable frameworks.

B. Flexibility Requirements for an Ethical AI

Integrating ethical AI into legally binding decisions requires acknowledging that legal standards vary significantly by context. Ethical and legal thresholds are not fixed; they shift based on specific circumstances and competing interests. Therefore, ethical AI systems must be adaptable, aligning with legal norms while upholding fairness and accountability.

For example, in serious criminal investigations like murder, authorities may justifiably relax privacy protections to prioritize truth-seeking over individual rights. In contrast, financial transactions—such as a custodian bank managing client funds—demand stricter ethical and legal adherence, emphasizing fiduciary duty over operational flexibility. These scenarios highlight the need for dynamic ethical standards that adjust to context.

This variability challenges the one-size-fits-all ethics frameworks found in much of the AI literature, which often overlook the contextual nature of legal and ethical reasoning. An AI system cannot apply identical principles to a police investigation and a banking transaction, given their differing priorities and risks.

We propose a flexible, modular framework based on *ontological blocks of meaning*—discrete ethical principles that can be layered or combined. This allows AI systems to

dynamically tailor their ethical reasoning to specific legal domains. For instance, an AI analyzing financial data might prioritize fiduciary responsibility, while one used in criminal justice may focus on due process. Such a system ensures ethical adaptability, legal alignment, and transparent oversight.

C. Ontological Blocks of Meaning

Ontology—the philosophical study of definitions—offers a foundation for aligning human ethical concepts with AI systems. Just as the concept of *light* evolved from Newtonian particles to quantum duality, terms like *responsibility* or *ownership* must be redefined for AI interpretation while preserving their ethical core.

In AI ethics, this means translating abstract ethical terms—*right*, *wrong*, *ownership*, *responsibility*—into constructs that AI can process. These must be both philosophically robust and operationally viable. Ontological blocks are modular, structured representations of ethical principles that can be integrated into AI operations. Creating them involves: Translating ethical concepts into machine-readable constructs; Embedding them into modular, systematic representations; Preserving ethical meaning while ensuring scalability and precision.

This ensures AI systems align with human values while providing a functional, auditable ethical framework.

Practically, creating ontological blocks involves developing a structured system to categorize ethical principles, values, and guidelines. Drawing from ontological practices in fields like oncology—e.g., the Cancer Care Treatment Outcome Ontology [21] and the NCI Thesaurus [20], [39]—this process starts by defining core ethical dimensions such as fairness, accountability, and transparency, and mapping them to relevant AI domains.

Collaboration with domain experts ensures term accuracy and contextual relevance. Ethical AI frameworks such as those by Prem [14] support alignment with current standards. Recent models like the Ontology for Ethical AI Principles (AIPO) use vocabularies such as Dublin Core, SKOS, FOAF, and DCAT2 to generate dynamic knowledge graphs, enabling semantic querying and systematic analysis [40]. These approaches support transparency, consistency, and accountability in ethical AI development.

D. Designing Ontological Blocks for Algorithmic Use

Ontological blocks translate ethical concepts into structured, quantifiable forms. For instance, defining “*stealing is bad*” involves breaking it down into core concepts like *property* and *ownership*.

Example: A dataset element (e.g., a car) represents property, while ownership (e.g., “Tom owns the car”) provides context. *Generalization:* Instead of rule-based examples, AI evaluates principles of ownership, avoiding reliance on predefined scenarios. *Incomplete data:* AI may infer missing elements through cross-referencing. If cross-references are weak, human oversight is required. *Short format:* The point of the ontological block is that it would refer a single concept (e.g.

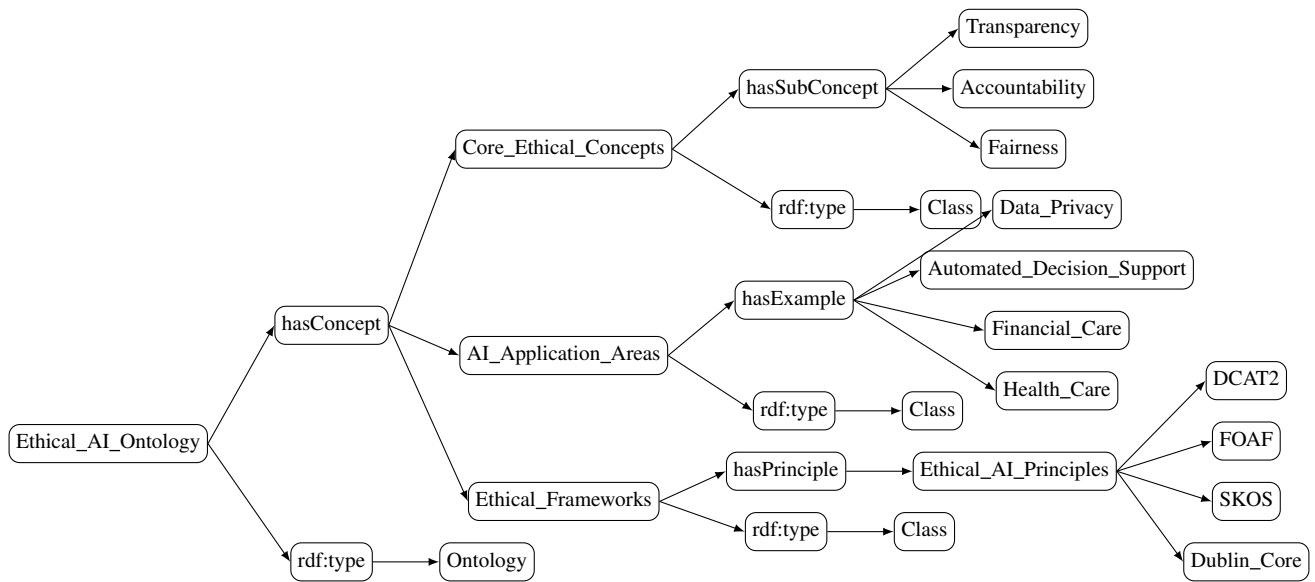


Fig. 1. An example of ontological block for ethical AI in RDF (Resource Description Framework) form

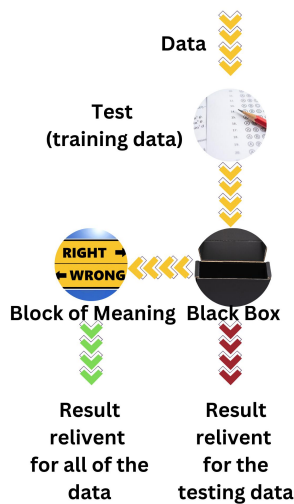


Fig. 2. Sketch: (a) Abstract view of ontological blocks processing for ethical frameworks.

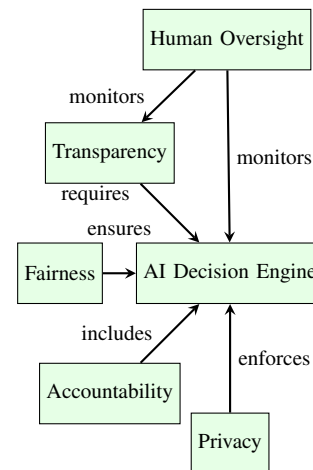


Fig. 3. Sketch: (b) Representation of ontological blocks for responsible AI. Each node represents a principle, and the relationships define their role in the AI system.

ownership), therefore more complex ethical questions would be described as a sum of multiple ontological blocks.

E. Common Structure of Ontological Blocks of Meaning

A major challenge in ethical AI is ensuring ontological blocks follow a common structure to support uniform testing and consistent term relationships. The proposed structure is simple: a primary concept paired with a binary ethical qualifier (e.g., *good* or *bad*).

For example, in the block “*stealing is bad*”, *stealing* is defined as the unauthorized transfer of ownership outside quantifiable pathways like trade, gift, bequest, or taxation. This format supports clear, consistent, and transparent ethical evaluations.

Block variables are selected by designers—researchers, practitioners, or stakeholders—ensuring both expert insight and public auditability. For complex concepts (e.g., *medical emergency*), multiple simpler blocks can be combined. This modular structure supports consistency and scalability across domains.

These are the foundational principles that could guide visualizing ethical AI ontological block. They include:

Fairness: Prevents bias and ensures equitable treatment. **Accountability:** Assigns responsibility for AI outcomes. **Transparency:** Promotes understandable, open AI processes **AI Application Areas:** Domains where ethical principles are applied. **Ethical Frameworks:** Structured sets of principles guiding ethical AI development.

Together, these components provide a robust structure for

designing ethical AI systems that are principled, transparent, and aligned with current standards.

F. Pros and Cons of the Proposed Framework

1) *Advantages:* Independence from training data, enabling robust ethical evaluations. Flexibility to combine blocks across fields, such as medicine or law. Auditability of standardized blocks, improving transparency and accountability.

2) *Drawbacks:* Labor-intensive creation of non-contradictory, quantifiable blocks. Reliance on large language models (LLMs) or external tools to infer missing relationships, introducing potential biases. Additional processing power required for integration and maintenance.

Despite challenges, the framework's flexibility, independence, and auditability make it a promising foundation for building ethically aligned AI systems.

G. A Use-Case: AI-Powered CRM and Ontological Profiling for Investor Protection

Overview: Brokerage firms are increasingly using AI-powered voice assistants to enhance investor classification and risk profiling. Moving beyond static binary systems, these tools enable more precise, ethical, and adaptive assessments of investor risk tolerance, protecting retail clients from unsuitable high-risk products [41].

1) *The Challenge:* Regulations require firms to distinguish between professional and retail investors to prevent the latter from accessing high-risk instruments like synthetic options. However, current assessments often overlook real-world risk tolerance, especially under stress, leading to potential harm and regulatory noncompliance.

2) *The Innovation:* This use case introduces *ontological blocks of meaning* for product classification. A financial product might trigger a "Riskier" block based on:

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

This enables real-time, context-aware, and ethically grounded risk assessment.

3) *How It Works:* AI voice assistants engage clients with adaptive questions (e.g., "How did you react to your last investment loss?"). A response like "Not well" may indicate emotional sensitivity. Using NLP and probabilistic modeling, the system estimates a behavioral risk profile.

If there's a high probability (e.g., 60%) of emotional vulnerability, a corresponding ontological block is triggered, ethically restricting access to high-risk products. This creates a personalized, behavior-informed investor classification framework.

H. Discussion

The evaluation shows that combining ontological blocks with FAIR principles offers a strong foundation for ethical AI monitoring. Key insights include:

- The unified monitoring system effectively identifies ethical risks and supports compliance (e.g., EU AI Act).

- Ontological blocks provide structured, modular, and interpretable ethical encoding.
- FAIR principles enhance discoverability, usability, and scalability.

Remaining challenges include the manual effort required to build blocks and the reliance on probabilistic reasoning with incomplete data. Future work should focus on automating block creation and improving data integration.

V. CONCLUSION AND RECOMMENDATION

This framework introduces ontological blocks as standardized, expert-designed tools for ethical AI assessment. These modular units enable transparent evaluations across diverse ethical contexts.

Key strengths include independence from training data and flexibility across domains. However, challenges remain in automating block creation and managing dependencies on probabilistic models.

By integrating with FAIR principles, the framework supports ethical, transparent, and accountable AI systems—laying the groundwork for responsible AI development across sectors.

ACKNOWLEDGMENT

The authors sincerely thank the Gesellschaft für wissenschaftliche Datenverarbeitung mbH Göttingen (GWDG, Germany) for their valuable contributions. We are especially grateful for support from NHR¹, essential for presenting this work, and appreciate the constructive feedback from our peers. This research was funded by the EU KISSKI Project under grant number 01—S22093A (Förderkennzeichen).

REFERENCES

- [1] A. Jobin, M. Ienca, and E. Vayena, "The global landscape of ai ethics guidelines," *Nature Machine Intelligence*, vol. 1, no. 9, pp. 389–399, 2019. [Online]. Available: <https://www.nature.com/articles/s42256-019-0088-2>
- [2] A. A. Abujaber and A. J. Nashwan, "Ethical framework for artificial intelligence in healthcare research: A path to integrity," *World Journal of Methodology*, vol. 14, no. 3, p. 94071, 9 2024, pMID: 39310239; PMCID: PMC11230076. [Online]. Available: <https://pmc.ncbi.nlm.nih.gov/articles/PMC11230076/>
- [3] "Regulation (EU) 2024/1689 of the european parliament and of the council of 13 june 2024 laying down harmonised rules on artificial intelligence and amending regulations (EC) no 300/2008, (EU) no 167/2013, (EU) no 168/2013, (EU) 2018/858, (EU) 2018/1139 and (EU) 2019/2144 and directives 2014/90/EU, (EU) 2016/797 and (EU) 2020/1828 (artificial intelligence act) (text with EEA relevance)," legislative Body: CONSIL, EP. [Online]. Available: <http://data.europa.eu/eli/reg/2024/1689/oj/eng>
- [4] N. Mika, G. Nadezhda, L. Jaana, and K. Raija, "Ethical ai for the governance of the society: Challenges and opportunities," in *CEUR Workshop Proceedings*, vol. 2505, 2019, pp. 20–26. [Online]. Available: <https://ceur-ws.org/Vol-2505/paper03.pdf>
- [5] D. Leslie, "Understanding artificial intelligence ethics and safety: A guide for the responsible design and implementation of ai systems in the public sector," *arXiv preprint arXiv:1906.05684*, 2019. [Online]. Available: <https://zenodo.org/record/3240529>
- [6] N. Bostrom and E. Yudkowsky, "The ethics of artificial intelligence," pp. 57–69, 2018, accessed: 2025-02-07. [Online]. Available: <https://www.taylorfrancis.com/chapters/edit/10.1201/9781351251389-4/ethics-artificial-intelligence-nick-bostrom-eliezer-yudkowsky>

¹<http://www.nhr-verein.de/en/ourpartners>

- [7] P. Brey and B. Dainow, "Ethics by design for artificial intelligence," *AI and Ethics*, vol. 4, no. 4, pp. 1265–1277, 2024. [Online]. Available: <https://link.springer.com/article/10.1007/s43681-023-00330-4>
- [8] A. B. Hanssen and S. Nichele, "Ethics of artificial intelligence demarcations," *Ethics in Information Technology*, pp. 133–142, 2019. [Online]. Available: https://link.springer.com/chapter/10.1007/978-3-030-35664-4_13
- [9] P. Linardatos, V. Papastefanopoulos, and S. Kotsiantis, "Explainable AI: A review of machine learning interpretability methods," vol. 23, no. 1, p. 18. [Online]. Available: <https://www.mdpi.com/1099-4300/23/1/18>
- [10] G. Vasan, Y. Wang, F. Shahriar, J. Bergstra, M. Jagersand, and A. R. Mahmood, "Revisiting sparse rewards for goal-reaching reinforcement learning," *arXiv preprint arXiv:2407.00324*, 2024. [Online]. Available: <https://arxiv.org/abs/2407.00324>
- [11] S. Ibrahim, M. Mostafa, A. Jnadi, H. Salloum, and P. Osinenko, "Comprehensive overview of reward engineering and shaping in advancing reinforcement learning applications," *IEEE Access*, vol. 12, pp. 175 473–175 500, 2024. [Online]. Available: <https://ieeexplore.ieee.org/document/10763475>
- [12] E. Mosqueira-Rey, E. Hernandez-Pereira, D. Alonso-Rios, J. Bobes-Bascaran, and A. Fernandez-Leal, "Human-in-the-loop machine learning: a state of the art," *Artificial Intelligence Review*, vol. 56, no. 4, pp. 3005–3054, 4 2023. [Online]. Available: <https://doi.org/10.1007/s10462-022-10246-w>
- [13] N. Mehrabi, F. Morstatter, N. Saxena, K. Lerman, and A. Galstyan, "A survey on bias and fairness in machine learning," *ACM Computing Surveys (CSUR)*, vol. 54, no. 6, pp. 1–35, 2021. [Online]. Available: <https://doi.org/10.1145/3457607>
- [14] E. Prem, "From ethical ai frameworks to tools: a review of approaches," *AI and Ethics*, vol. 3, no. 3, pp. 699–716, 2023, accessed: 2025-02-07. [Online]. Available: <https://doi.org/10.1007/s43681-023-00258-9>
- [15] A. M. Salih, Z. Raisi-Estabragh, I. Boscolo Galazzo, P. Radeva, S. E. Petersen, K. Lekadir, and G. Menegaz, "A perspective on explainable artificial intelligence methods: Shap and lime. [Online]. Available: <https://arxiv.org/abs/2305.02012v3>
- [16] F. Rossi and N. Mattei, "Building ethically bounded ai," *Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society*, pp. 68–74, 2018. [Online]. Available: <https://doi.org/10.48550/arXiv.1812.03980>
- [17] B. K. Hassani, "Societal bias reinforcement through machine learning: a credit scoring perspective," *AI and Ethics*, vol. 1, no. 3, pp. 239–247, 2021. [Online]. Available: <http://dx.doi.org/10.2139/ssrn.3625691>
- [18] N. G. Packin, "Disability discrimination using ai systems, social media and digital platforms: Can we disable digital bias?" *Journal of International and Comparative Law*, vol. 8, no. 2, pp. 487–512, 2021. [Online]. Available: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3724556
- [19] D. Talati, "Artificial intelligence (ai) in mental health diagnosis and treatment," *Journal of Knowledge Learning and Science Technology*, vol. 2, no. 3, pp. 251–253, 2023. [Online]. Available: <https://doi.org/10.60087/jklst.vol2.n3.p253>
- [20] A. Kumar and B. Smith, "Oncology ontology in the nci thesaurus," in *Artificial Intelligence in Medicine: 10th Conference on Artificial Intelligence in Medicine, AIME 2005, Aberdeen, UK, July 23-27, 2005. Proceedings 10*, Springer. Springer, 2005, pp. 213–220. [Online]. Available: https://link.springer.com/chapter/10.1007/11527770_30
- [21] F. P. Lin, T. Groza, S. Kocbek, E. Antezana, and R. J. Epstein, "Cancer care treatment outcome ontology: a novel computable ontology for profiling treatment outcomes in patients with solid tumors," *JCO clinical cancer informatics*, vol. 2, pp. 1–14, 2018. [Online]. Available: <https://ascopubs.org/doi/pdf/10.1200/CCI.18.00026>
- [22] S. Monteith, T. Glenn, J. R. Geddes, P. C. Whybrow, E. Achtyes, and M. Bauer, "Artificial intelligence and increasing misinformation," *The British Journal of Psychiatry*, vol. 224, no. 2, pp. 33–35, 2024. [Online]. Available: <https://doi.org/10.1192/bjp.2023.136>
- [23] G. Antoniou, E. Papadakis, and G. Baryannis, "Mental health diagnosis: a case for explainable artificial intelligence," *International Journal on Artificial Intelligence Tools*, vol. 31, no. 03, p. 2241003, 2022. [Online]. Available: <https://doi.org/10.1142/S0218213022410032>
- [24] P. Lin, "Why ethics matters for autonomous cars," *Autonomous driving: Technical, legal and social aspects*, pp. 69–85, 2016. [Online]. Available: https://link.springer.com/chapter/10.1007/978-3-662-45854-9_4
- [25] R. Jenkins, D. Cerný, and T. Hrfbek, *Autonomous vehicle ethics: the trolley problem and beyond*. Oxford University Press, 2022. [Online]. Available: <https://academic.oup.com/book/44058>
- [26] N. J. Goodall, "Machine ethics and automated vehicles," *Road vehicle automation*, pp. 93–102, 2014. [Online]. Available: https://link.springer.com/chapter/10.1007/978-3-319-05990-7_9
- [27] S. Nyholm and J. Smids, "The ethics of accident-algorithms for self-driving cars: An applied trolley problem?" *Ethical theory and moral practice*, vol. 19, no. 5, pp. 1275–1289, 2016. [Online]. Available: <https://link.springer.com/article/10.1007/s10677-016-9745-2>
- [28] D. L. McGuinness, F. Van Harmelen *et al.*, "Owl web ontology language overview," *W3C recommendation*, vol. 10, no. 10, p. 2004, 2004. [Online]. Available: <https://static.twoday.net/71desa1bif/files/W3C-OWL-Overview.pdf>
- [29] G. Antoniou and F. v. Harmelen, "Web ontology language: Owl," *Handbook on ontologies*, pp. 91–110, 2009. [Online]. Available: https://link.springer.com/chapter/10.1007/978-3-540-92673-3_4
- [30] R. Guizzardi, G. Amaral, G. Guizzardi, and J. Mylopoulos, "An ontology-based approach to engineering ethicality requirements," *Software and Systems Modeling*, vol. 22, no. 6, pp. 1897–1923, Jul. 2023. [Online]. Available: <https://link.springer.com/article/10.1007/s10270-023-01115-3>
- [31] J. S. Roberts and L. N. Montoya, "Contextualizing artificially intelligent morality: A meta-ethnography of theoretical, political and applied ethics," *Journal of Artificial Intelligence Research*, vol. 75, pp. 482–501, 2022. [Online]. Available: https://link.springer.com/chapter/10.1007/978-3-031-28073-3_35
- [32] R. Benjamins and I. Salazar, "Towards a framework for understanding societal and ethical implications of artificial intelligence," *CoRR*, vol. abs/2001.09750, 2020. [Online]. Available: <https://arxiv.org/abs/2001.09750>
- [33] R. Gruetzmacher, A. Chan, K. Frazier, C. Manning, S. Los, J. Fox, J. Hernández-Orallo, J. Burden, M. Franklin, C. N. Ghuidhir, M. Bailey, D. Eth, T. Pilditch, and K. Kilian, "An international consortium for evaluations of societal-scale risks from advanced ai," Nov. 2023, arXiv:2310.14455 [cs]. [Online]. Available: <http://arxiv.org/abs/2310.14455>
- [34] R. Leaders, "The role of rri in the ethical governance of AI | RRI-leaders," 2023. [Online]. Available: <https://www.rri-leaders.eu/the-role-of-rri-in-the-ethical-governance-of-ai/>
- [35] S. M. Field, J. Thompson, S. De Rijcke, B. Penders, and M. R. Munafò, "Exploring the dimensions of responsible research systems and cultures: a scoping review," *Royal Society Open Science*, vol. 11, no. 1, p. 230624, 2024. [Online]. Available: <https://doi.org/10.1098/rsos.230624>
- [36] S. Ivanova, C. Reichetzer, A. Martinuzzi, F. Findler, and K. Miko-Schefzig, "Frames, interests, and incentives—a typology of institutionalizing rri in the business sector derived from ten pioneering projects," *Journal of Responsible Innovation*, vol. 10, no. 1, p. 2267736, 2023. [Online]. Available: <https://doi.org/10.1080/23299460.2023.2267736>
- [37] T. Völker, R. Slaattelid, and R. Strand, *Translations of Responsibility: Innovation Governance in Three European Regions*. Taylor & Francis, 2024. [Online]. Available: <https://library.oapen.org/handle/20.500.12657/86044>
- [38] H. Pacifico Silva, P. Lehoux, F. A. Miller, and J.-L. Denis, "Introducing responsible innovation in health: a policy-oriented framework," *Health research policy and systems*, vol. 16, pp. 1–13, 2018. [Online]. Available: <https://link.springer.com/article/10.1186/s12961-018-0362-5>
- [39] W. Ceusters, B. Smith, and L. Goldberg, "A terminological and ontological analysis of the nci thesaurus," *Methods of information in medicine*, vol. 44, no. 04, pp. 498–507, 2005, accessed: 2025-02-07. [Online]. Available: <https://www.thieme-connect.com/products/ejournals/abstract/10.1055/s-0038-1634000>
- [40] J. Harrison, R. Smith, and L. Thompson, "An ontology for ethical ai principles," *AI and Society*, vol. 36, pp. 123–136, 2021, accessed: 2025-02-07. [Online]. Available: <https://doi.org/10.1007/s00146-020-01046-7>
- [41] N. S. Egbuhuzor, A. J. Ajayi, E. E. Akhigbe, O. O. Agbade, C. P.-M. Ewim, and D. I. Ajiga, "Ai and data-driven insights: Transforming customer relationship management (crm) in financial services," *Gulf Journal of Advance Business Research*, vol. 3, no. 2, pp. 483–511, 2025. [Online]. Available: https://link.springer.com/chapter/10.1007/978-3-031-28073-3_35