Check for updates

Commentary

FAIR Connect
Volume 3: 1–6
© The Author(s) 2025
Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/2949799X251403108
journals.sagepub.com/home/fco
Sage | IOS Press

FAIR data principles and AI ethics: Exploring convergence and gaps

Anastasios G Papadiamantis[1,2], Iseult Lynch[1,3], Georgia Melagraki[4] and Antreas Afantitis[1,2,5]

Abstract
Mapping AI-ethics frameworks to the FAIR principles and their extensions shows good conceptual alignment. While not directly referenced in AI-ethics frameworks, FAIR can enhance ethical AI use through traceability, accountability, auditability and transparent data governance.

Keywords
FAIR principles (http://purl.obolibrary.org/obo/T4FS_0000447), ethics (http://purl.obolibrary.org/obo/NCIT_C39796), artificial intelligence (http://purl.obolibrary.org/obo/NCIT_C16309), machine learning (http://purl.obolibrary.org/obo/NCIT_C176231)

Received: 16 June 2025; accepted: 16 November 2025

1 Introduction

Technological advancement and the production of large volumes of data (so-called big data) have led to the increasing use of machine learning (ML) and artificial intelligence (AI) in all aspects of everyday life, including science, technology and education. These advances in computing and data processing have sparked discussions on digitising Research and Innovation (R&I) by industry[1] and the European Commission.[2]

The rapid growth of AI has raised ethical concerns[3–5] about its responsible use and potential unintended impacts on people and the environment. AI system behaviour is unpredictable, and an unreliable predictor of future actions, as learning and new data can change future responses.[3] AI can also have unintended consequences, such as re-identifying anonymised sensitive data in previously unforeseen ways.[3]

Data governance[6] is key to ethical AI, given the challenges of managing large, dynamic and diverse datasets. Data governance ensures availability, usability, integrity and security, becoming a core focus across all data-driven fields. The 2016 introduction of FAIR (Findable, Accessible, Interoperable, Reusable) principles[7] has supported ongoing improvements in scientific, and wider, data governance and management. Similarly, the extension of FAIR to Computational Workflows[8] and research software (FAIR4RS),[9] supports FAIR practices for ML/AI models and related webtools.

Thus, the following questions emerge:

1. Should the FAIR data principles be part of ethical AI development?
2. Do existing AI-ethics frameworks align with the FAIR principles?

[1]Entelos Institute, Larnaca, Cyprus
[2]NovaMechanics Ltd, Nicosia, Cyprus
[3]School of Geography, Earth and Environmental Sciences, University of Birmingham, Birmingham, UK
[4]Division of Physical Sciences & Applications, Hellenic Military Academy, Vari, Greece
[5]NovaMechanics MIKE, Piraeus, Greece

Corresponding author:
Anastasios G. Papadiamantis, Entelos Institute, Larnaca, Cyprus; NovaMechanics Ltd, Nicosia, Cyprus.
Email: papadiamantis@entelos.eu

Creative Commons CC BY: This article is distributed under the terms of the Creative Commons Attribution 4.0 License (https://creativecommons.org/licenses/by/4.0/) which permits any use, reproduction and distribution of the work without further permission provided the original work is attributed as specified on the SAGE and Open Access page (https://uk.sagepub.com/en-gb/eur/open-access-at-sage).

Table 1. Alignment level between the FAIR, FAIR for Computational Workflows and FAIR4RS principles and the major ethical AI frameworks and their respective requirements.

Alignment level	Requirements
Strong	The framework includes well-developed practices and mechanisms that directly support the outcomes intended by the FAIR principles
Moderate	The framework supports multiple FAIR-related outcomes, but without complete coverage or technical specificity.
Partial	The framework touches on a few aspects of FAIR but lacks broader scope or operational clarity.
Implicit	The framework promotes adjacent values like auditability, equity, or accountability, which could benefit from FAIR practices but makes no mention of data governance mechanisms.
Weak	The framework provides no substantive consideration of data access, documentation, or lifecycle governance.

This alignment can potentially transform key pillars of AI-ethics frameworks like transparency, auditability, accountability and remedy of harmful outcomes into testable and machine-actionable processes using FAIR's infrastructure, for example, Globally Unique, Persistent, and Resolvable Identifiers (GUPRIs), standardised metadata, and structured vocabularies. Thus, data stewards can utilise existing tools and develop roadmaps for FAIR-aligned ethical AI development and assessment.

2 Overview of AI ethical frameworks

A wide range of AI-ethics frameworks have been published by public, private and research organisations. Although these frameworks address similar ethical issues like transparency, fairness, non-maleficence (or do-no-harm), responsibility and privacy, their approaches vary significantly.[10] The key high-level ethical guidelines are the EU's Ethics Guidelines for Trustworthy AI,[11] the OECD AI Principles,[12] UNESCO's Recommendation on the Ethics of AI,[13] the IEEE Ethically Aligned Design,[14] Singapore's Model AI Governance Framework,[15] The Montreal Declaration for Responsible AI,[16] the Toronto Declaration,[17] the Beijing AI Principles[18] and the G20 AI Principles.[19] A qualitative analysis was performed to evaluate how these frameworks align with the FAIR, FAIR for Computational Workflows and FAIR4RS principles, through critical text review and analysis. Each identified, to the authors' opinion, FAIR alignment was evaluated as Strong, Moderate, Partial, Implicit or Weak based on predefined criteria (Table 1), for example, whether mechanisms like GUPRIs, structured vocabularies, licensing, are mentioned or a compatible term was stated, for example, transparency, auditability. The results of this analysis are presented as a heatmap in Figure 1, while the tables corresponding to the heatmap can be found in reference.[20] Terms similar to the FAIR principles' requirements, like 'open and shared', 'traceability', 'access' and 'documentation' are common across ethical AI guidelines. The EU Guidelines for Trustworthy AI emphasise data quality, integrity and mitigating implicit bias through careful dataset governance.[11] The OECD AI Principles[12] require robust risk management and traceability to address (meta)data reliability and bias. UNESCO[13] promotes non-discrimination, traceability and interoperability, while Beijing Principles[18] require fairness, bias handling, and traceable and auditable systems, and open and shared platforms. The Montreal[16] and Toronto[17] declarations warn against data-related discrimination, with the latter outlining specific prevention measures. The G20 AI Principles[19] promote transparent, explainable AI, aligning with FAIR principles F2 and R1 on rich metadata. Furthermore, G20[19] calls for digital AI ecosystems based on free data flow and interoperable resources, echoing FAIR aspects of findability and accessibility (FAIR principles F4, A1, A1.1, A1.2, I1, I2, I3).

Given their high-level, policy-focused and domain-agnostic nature, these guidelines do not mention technical details or requirements. Terms like open data or traceable systems lack concrete solutions, like GUPRIs or (meta)data schemas, that align with FAIR,[7] FAIR for Computational Workflows[8] and FAIR4RS recommendations.[9] The EU Guidelines[11] and OECD AI Principles[12] mention accountability and documentation, but omit practical solutions like version control or licensing. While preventing discrimination through balanced data is a shared goal, the AI-ethics frameworks lack practical measures like using (meta)data schemas or controlled vocabularies. Calls for clarity, transparency and accountability in AI data align with the FAIR principles and the 'as open as possible, as closed as necessary' FAIRness approach. References to auditability, explainability and risk management align with FAIR's emphasis on organised, discoverable and reproducible data, while calls for robust metadata, oversight and data quality echo FAIR's core goals.

3 FAIR data principles and their relevance to AI ethics

The FAIR,[7] Computational Workflow[8] and FAIR4RS[9] principles guide data, workflow and software management to support digitisation and improve machine actionability of all (meta)data. Rich metadata aim to support community consensus on enhancing (meta)data accountability, reproducibility and transparency. FAIR data are often mistaken for open data, but they primarily mean well-documented, securely managed and machine-actionable datasets. This distinction is significant, as digital R&I and AI-driven process optimisation, including health and safety aspects, rely, in part, on managing sensitive human and commercial data. As a result, the line between FAIR and AI ethics becomes blurred.

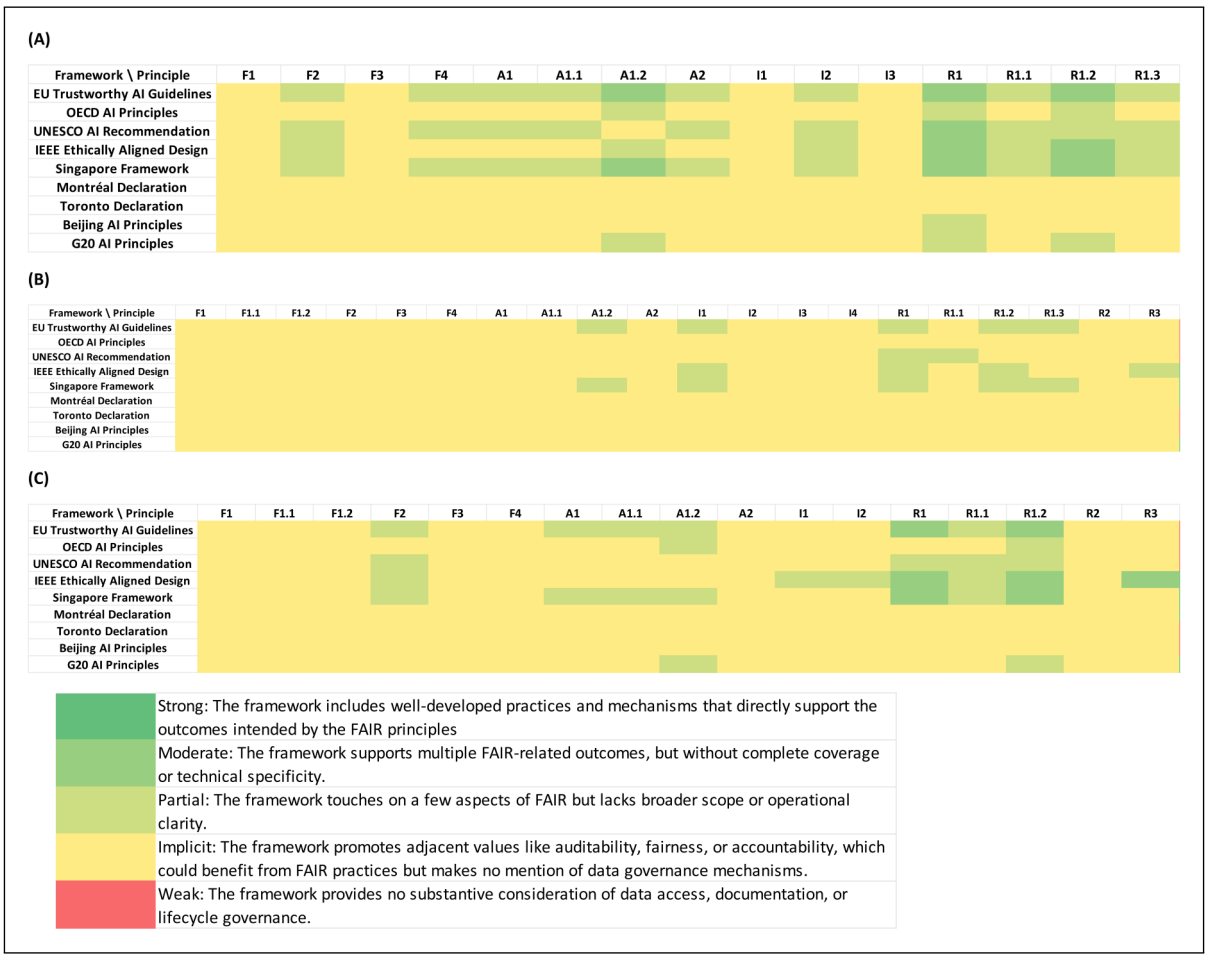

Figure 1. Alignment heatmap of the major AI ethical frameworks with the FAIR (A), FAIR for Computational Workflows (B), and FAIR for Research Software (FAIR4RS) (C) principles.

Operationalising FAIR through tools like GUPRIs, rich metadata, standard formats, licensing and provenance supports machine actionability and aligns closely with ethical AI concerns, which together form the basis of data governance. Furthermore, FAIR's metadata richness, sharing, long-term preservation and transparency, are reflected in frameworks like the EU Guidelines for Trustworthy AI,[11] the OECD AI Principles[12] and Singapore's AI Governance Framework.[15]

The EU Guidelines[11] stress the need for data governance to ensure 'data quality and integrity', highlighting traceability and verifiability as prerequisites for transparency and accountability. Like FAIR, the OECD AI Principles[12] recommend robust data governance to enable risk assessment, reproducibility and user confidence. Singapore's framework[15] addresses data provenance, lifecycle management and internal and external audit mechanisms, directly aligning with FAIR's Accessible, Interoperable and Reusable pillars.

Another way to understand the connection between the FAIR principles and AI-ethics frameworks is the nomenclature used. Traceability and openness are explicitly referenced by the Beijing[18] and the G20[19] AI Principles, while rich metadata are implicit via transparency, disclosure and auditability in the Toronto Declaration.[17] UNESCO's recommendations[13] advocate, in practice, FAIR-aligned data governance, stressing safeguarded accessibility, traceability and auditability, increased interoperability of tools and data, and the use of robust datasets for scientific, R&I and societal benefit. These values align with the FAIR principles and promote AI transparency and accountability, allowing validation, explainability and ethical AI monitoring.

Considering this correspondence, FAIR, FAIR for Computational Workflows and FAIR4RS principles can be mapped to major AI-ethics frameworks (presented as a heatmap in Figure 1), based the authors' critical qualitative evaluation deeming the alignment to be strong, moderate, partial, implicit or weak. For example, the FAIR principles emphasis on findable well-documented data align with the transparency and auditability requirements of the EU Guidelines[11] and Toronto Declaration.[17] The emphasis on controlled and standardised access, using free and open protocols and, where required, authentication and authorisation, supports the AI-ethics frameworks' calls for independent review.[12,13,19] FAIR's focus on data reusability aligns with the IEEE design[14] and EU Guidelines,[11] which mandate robust documentation and provenance. Similarly, the FAIR for Computational Workflows and FAIR4RS follow the same approach regarding GUPRIS, versioning, licensing and execution provenance, which are prerequisites for AI reproducibility and auditability.

Table 2. Technical solutions to operationalise the verification of ethical-AI pillars using the FAIR, FAIR for Computational Workflows, and FAIR4RS principles, targeted to data stewards.

Solution	FAIR relevance	Ethical pillar enabler
GURPIs	FAIR: F1, Computational Workflows: F1, F1.2, FAIR4RS: F1, F1.2	Traceability, audit trails
Machine actionable records, e.g., JSON-LD profile, domain-relevant metadata schema	FAIR: F2, R1, Computational Workflows: F2, R1, FAIR4RS: F2, R1	Transparency, explainability, reproducibility
HTTPS/REST APIs and KeyCloak, OAuth2, AAI	A1, A1.1, A1.2	Independent review, lawful access
Metadata persistence	A2	Remedy, ex-post investigation
Vocabularies, qualified links	FAIR: I1–I3, Computational Workflows: I1–I4, FAIR4RS: I1–I2	Cross-system audits, bias studies
SPDX, CC, Apache licences	R1.1	Lawful reuse, accountability
PROV, RO-Crate, CWLProv	FAIR: R1.2, Computational Workflows: R1.2, R1.3, FAIR4RS: R1.2	Accountability, reproducibility
Versioning of code, data, workflows	F1.2, R1.2 (Computational Workflows / FAI4RS)	Change audit, comparability

4 Challenges and opportunities

Although the FAIR principles and their extensions can substantially support ethical AI, explicit mention and utilisation remain absent from current AI-ethical frameworks. Like the FAIR principles, most AI-ethics frameworks promote traceability, transparency, auditability and accountability, without dictating ways to achieve these. Likewise, the FAIR principles are not intended, by design, to be part of ethical AI development.[21]

A major gap across many current AI-ethics frameworks is the lack of specificity in data governance, providing limited technical solutions, for example, GUPRIS, metadata schemas. In line with the FAIR,[7] FAIRification of computational workflows,[8] and FAIR4RS[9] principles, the EU Guidelines for Trustworthy AI[11] stress (meta)data auditability and integrity but lack specifics on metadata standards, GUPRIs, or curation processes required to assure these. The G20 AI Principles[19] promote transparency, traceability and accountability but offer no technical guidance for managing or structuring data to ensure interoperability or reusability. The Beijing AI Principles[18] mention data openness and sharing but lack guidance on controlled access, licensing or use of GUPRIs. Singapore's Model AI Governance Framework[15] offers risk-tiered guidance on data traceability (provenance, quality and lineage) and human oversight (human-in-the-loop, human-over-the-loop, human-out-of-the-loop), but omits explicit FAIR practices like domain-specific metadata standards, ontologies and version control.

Despite existing gaps, ethical AI frameworks conceptually align with the FAIR principles on data traceability, auditability and accountability. While not explicitly endorsing FAIR, the UNESCO Recommendation on AI Ethics[13] strongly supports robust data governance, transparency and scientific reproducibility. The OECD AI Principles[12] and Toronto Declaration[17] emphasise transparency, auditability and remedy of harmful outcomes, including from algorithms, which are facilitated by Findable and Reusable data. The IEEE Ethically Aligned Design guidelines[14] support long-term data stewardship and human-understandable records, aligning with human actionability and FAIR principle A2 on persistent metadata even when the data no longer exist.

In summary, although current AI-ethics frameworks don't explicitly reference FAIR, FAIR for Computational Workflows, or FAIR4RS, they are closely aligned (Figure 1). Although alignment varies by framework, there are many opportunities for more descriptive and practical alignment to accelerate ethical AI adoption in data, model and software production. Metadata standards, traceability, interoperability protocols and responsible access models are just some of the opportunities. Adopting FAIR and its extensions to bridge AI ethics gaps offers a promising path forward, even for AI system architecture.

Implementing FAIR into ethical AI frameworks and practice faces several barriers. Legal constraints, for example, GDPR, can limit data accessibility. Interoperability challenges include missing metadata standards, limited ontologies and the need for community consensus. AI ethics challenges arise when reusability or accessibility conflict with confidentiality, IP rights or community data ownership. Practical solutions include secure data enclaves, differential privacy, data stewards and structured access and data governance, while interdisciplinary ethics boards can help resolve conflicts, ensuring FAIR practices balance innovation with ethical safeguards.

A starting point for data stewards, responsible for ethical AI management, could be the solutions identified and presented in Table 2, which links technical solutions to the respective FAIR-related principles and the ethical AI pillar they would enable. A practical data steward roadmap for implementation should (i) include GUPRIs, for traceability and unambiguous referencing, (ii) distinct machine-actionable and domain-specific metadata records, for persistence, explainability and reproducibility, licencing, human- and machine-actionability, to enable lawful and independent review, (iii) community-related vocabularies, for cross-domain bias checks and linkage and (iv) versioning, for change-auditing and comparability. The Research Data Alliance FAIR4ML group, which implements domain-specific and agnostic FAIR in ML by extending the FAIR4RS principles,[22] converged on a practical top 10 for FAIRifying ML models using a Delphi process.[23] Their findings

align well and inform our data steward roadmap (Table 2), which also considers the FAIR for Computational Workflows and FAIR4RS principles, as reported in Figure 1.

5 Conclusion

This paper examines the convergence between the FAIR data principles, and their extensions for computational workflows and research software, and widely adopted ethical AI frameworks. While not explicitly referenced, there is strong alignment with FAIR in principles like transparency, traceability and responsible data governance. The practical and ethical benefits of integrating FAIR, for example, enhanced accountability, reproducibility and inclusivity will benefit the development of responsible AI systems. To strengthen this alignment, we recommend actions including the development and adoption of standardised metadata schemas for ML and AI models, FAIR-aligned technical systems, and the explicit reference of FAIR, FAIRification of computational workflows and FAIR4RS in ethical AI guidelines. Future work should explore measures to integrate these principles and how their adoption impacts AI system trustworthiness and sustainability.

ORCID iDs

Anastasios G Papadiamantis https://orcid.org/0000-0002-1297-3104
Iseult Lynch https://orcid.org/0000-0003-4250-4584
Georgia Melagraki https://orcid.org/0000-0001-7547-2342
Antreas Afantitis https://orcid.org/0000-0002-0977-8180

Funding

The authors disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work was supported by the European Union Horizon Europe project SSbD4Chem (Grant Agreement No. 101138475), and Innovate UK via the Horizon Europe Guarantee Fund (grant no. 10045979) for participation in Horizon Europe partnership PARC (Grant Agreement no. 101057014).

Declaration of conflicting interests

The authors declared the following potential conflicts of interest with respect to the research, authorship, and/or publication of this article: AA and AGP are employed by NovaMechanics, a Cheminformatics and Materials Informatics company.

References

1. Materials 2030 Manifesto. Systemic approach of advanced materials for prosperity: a 2030 perspective, 2022. https://www.ami2030.eu/wp-content/uploads/2022/06/advanced-materials-2030-manifesto-Published-on-7-Feb-2022.pdf.
2. EC. Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions: Advanced Materials for Industrial Leadership, 2024. https://research-and-innovation.ec.europa.eu/document/download/0fcf06ea-c242-44a6-b2cb-daed39584996_en?filename=com_2024_98_1_en_act_part1.pdf.
3. Stahl BC. *Ethical issues of AI: artificial intelligence for a better future: an ecosystem perspective on the ethics of AI and emerging digital technologies*. Cham: Springer International Publishing, 2021, pp.35–53.
4. Yin RK. *Applications of case study research*. Thousand Oaks, CA, USA: Sage publications, 2011.
5. Yin RK. *Case study research: design and methods*. Thousand Oaks, CA, USA: Sage, 2009.
6. Janssen M, Brous P, Estevez E, et al. Data governance: organizing data for trustworthy artificial intelligence. *Gov Inf Q* 2020; 37: 101493.
7. Wilkinson MD, Dumontier M, Aalbersberg IJ, et al. The FAIR guiding principles for scientific data management and stewardship. *Sci Data* 2016; 3: 160018.
8. Wilkinson SR, Aloqalaa M, Belhajjame K, et al. Applying the FAIR principles to computational workflows. *Sci Data* 2025; 12: 328.
9. Barker M, Chue Hong NP, Katz DS, et al. Introducing the FAIR principles for research software. *Sci Data* 2022; 9: 622.
10. Jobin A, Ienca M and Vayena E. The global landscape of AI ethics guidelines. *Nat Mach Intell* 2019; 1: 389–399.
11. Commission E. Directorate-General for Communications Networks Content & Technology (DG CNECT). Ethics guidelines for trustworthy AI: Publications Office, 2019.
12. OECD. AI principles, 2024. https://www.oecd.org/en/topics/sub-issues/ai-principles.html.
13. UNESCO. Recommendation on the Ethics of Artificial Intelligence: United Nations Educational, Scientific and Cultural Organisation, 2022. https://unesdoc.unesco.org/notice?id=p::usmarcdef_0000381137.
14. IEEE. IEEE Ethically Aligned Design: Institute of Electrical and Electronics Engineers, 2024. http://standards.ieee.org/wp-content/uploads/import/documents/other/ead_v2.pdf.
15. Singapore Personal Data Protection Commission. Model Artificial Intelligence Governance Framework, 2020. https://www.pdpc.gov.sg/help-and-resources/2020/01/model-ai-governance-framework.
16. Montréal Declaration. Montreal declaration for responsible AI, 2018. https://montrealdeclaration-responsibleai.com/the-declaration/#:~:text=The%20Montreal%20Declaration%20is%20addressed,to%20it%2C%20to%20be%20able.
17. Amnesty International, Access Now. The Toronto Declaration, 2018. https://www.torontodeclaration.org/declaration-text/english/.
18. International Research Center for AI Ethics and Governance. The Beijing AI Principles, 2019. https://ai-ethics-and-governance.institute/beijing-artificial-intelligence-principles/.

19. G20. G20 AI Principles, 2019. https://oecd.ai/en/wonk/documents/g20-ai-principles.
20. Papadiamantis AG, Lynch I, Melagraki G, et al. FAIR data principles and AI ethics: exploring convergence and gaps - qualitative alignment, 2025. https://osf.io/zphjq/.
21. Mons B, Schultes E, Liu F, et al. The FAIR principles: first generation implementation choices and challenges. *Data Intell* 2020; 2: 1–9.
22. Limani F, Tofik L, Latif A, et al. FAIR for machine learning model: principles and assessment metrics. Zenodo, 2024.
23. Osmenaj E, Sharma CJ, Moschini U, et al. FAIR Principles implementation in ML/AI-findings from Skills4EOSC Delphi study. *Int J Digital Curation* 2025; 19: 6.