Ethics is a quality for software products: Ethical label for transparency and awareness to users

Marco Autili, University of L'Aquila Italy, **Riccardo Corsi,** University of Pisa/GSSI, Italy,
Martina De Sanctis, Gran Sasso Science Institute, Italy **Paola Inverardi**, Gran Sasso Science Institute, Italy
Patrizio Pelliccione, Gran Sasso Science Institute

NB: This interview will be audio-recorded for later analysis.

The study goal and background

Motivation of the work

Do modern systems require new quality dimensions?

Table 1. Stakeholders needs emerged from the analysed literature.

Name	Sources
Societal and environmental well-being	[5,16,13]
Accountability	[5,16,13]
Privacy and data governance	[5,10,1,16,17]
Human agency and oversight, human autonomy	[5,9,16,17]
Transparency	[5,16,13]
Explicability, Explainability	[9,10,1,16,13,17]
Diversity, non-discrimination and fairness	[5,9,10,1,16,13,17]

Why an ethics label?

Assessing ethics as a quality for modern digital systems in a user-centered perspective.

The goal of our work is to move beyond mere compliance with standards, proposing an ethics label that helps users understand how software systems impact human, societal, and environmental values—both throughout their development and during real-world use.

How do we build the ethics label?

(i) reviewing existing guidelines, framework, recommendations, laws and regulations on digital systems to identify human, societal, and environmental drivers of innovation;

(ii) identifying and analyzing existing standards, e.g., the SQuaRE family

(iii) realizing an ethical label user-centered.

Identified limitations

(i) Need of new qualities

(ii) Flourishing of standards

(iii) Focus mostly on product-oriented aspects

(iv) Limitation of CE marking / legal compliance

(v) Need of transparency and digital-ethical literacy

Considered sources - Guidelines, Frameworks, Laws, Regulations, Standards

Guidelines, frameworks, regulations

- High-Level Expert Group on AI: Ethics guidelines for trustworthy AI (2019/2024)
- Beijing AI Principles (2019)
- OECD.AI Policy Observatory: OECD AI Principles overview (2019/2024)
- UNESCO: Recommendation on the ethics of artificial intelligence (2022)
- G7 Hiroshima Process: International Guiding Principles for Advanced AI Systems (2023)
- United States Government: Blueprint for an AI Bill of Rights Making Automated Systems Work for the American People (2022)
- USA Algorithmic Accountability Act (2022)
- Council of Europe Framework Convention on AI and Human Rights (2024)
- China’s Deep Synthesis Provisions (2023)
- China’s Interim Measures on Generative AI (2023)
- China’s Governance principles for the New Generation Artificial Intelligence (2019)
- China’s Ethical Norms for the NGAI (2021)
- Government Data Quality Frameworks: UE; Canada

SQuaRE Standards

- ISO/IEC 25010:2023 Software Product Quality
- ISO/IEC 25059:2023 Quality model for AI systems
- ISO/IEC 25012:2008 Data quality model
- ISO/IEC 5259:2024 - Quality model for data
- analytics and AI based on ML
- ISO/IEC 25019:2023 Quality-in-use model

Human, Societal, Environmental (HSE) Drivers and ISO Standards SQuaRe

ISO Standards SQuaRe

ISO/IEC 25010: 2023 Software Product Quality + ISO/IEC 25059: 2023								
Functional Suitability	**Performance efficiency**	**Compatibility**	**Interaction capability**	**Reliability**	**Security**	**Maintainability**	**Flexibility**	**Safety**
- Functional completeness - Functional correctness - Functional appropriateness - Functional adaptability	- Time behaviour - Resource utilization - Capacity	- Co-existence - Interoperability	- Appropriateness recognizability - Learnability - Operability - User error protection - User engagement - Inclusivity - User assistance - Self-descriptiveness - User controllability - Transparency	- Faultlessness - Availability - Fault tolerance - Recoverability - Robustness	- Confidentiality - Integrity - Non repudiation - Accountability - Authenticity - Resistance - Intervenability	- Modularity - Reusability - Analizability - Modifiability - Testability	- Adaptability - Scalability - Installability - Replaceability	- Operational constraint - Risk identification - Fail safe - Hazard warning - Safe integration

ISO Standards SQuaRe

ISO/IEC 25019:2023 - Quality in use model + ISO/IEC 25059:2023			
Beneficialness	**Freedom from risk**	**Acceptability**	**Satisfaction**
- Usability - Accessibility - Suitability	- Freedom from economic risk - Freedom from environmental and societal risk - Freedom from health risk - Freedom from human life risk - Societal and ethical risk mitigation	- Experience - Trustworthiness - Compliance	Transparency

ISO/IEC 25012:2008 + ISO/IEC 5259:2024 - Quality model for data analytics and AI based on ML			
Inherent data quality	**Inherent and system-dependent data quality**	**System-dependent data quality**	**Additional characteristics**
- Accuracy - Completeness - Consistency - Credibility - Currentness	- Accessibility - Compliance - Confidentiality - Efficiency - Precision - Traceability - Understandability	- Availability - Portability - Recoverability	- Auditability - Identifiability - Effectiveness - Balance - Diversity - Relevance - Representativeness - Similarity - Timeliness

HSE Drivers

Table 1: Categories and subcategories of HSE drivers. The blue text is the extension with respect to [9].

Societal and Environmental Well-being	**Accountability and Responsibility**	**Privacy and Data Governance**	**Human Agency and Oversight**	**Transparency and Explainability**	**Diversity, Fairness, and Non-discrimination**
• Sustainable and environmental friendliness • Societal and social impact • Society and democracy • Respect of the rule of law: normativeness • AI and digital literacy • Openness and plurality	• Auditability • Minimization and reporting of negative impacts • Tradeoffs and redress	• Accuracy, completeness, consistency, timeliness, uniqueness in data quality, interpretability, coherence • Security and privacy considerations, controllability, adaptability, supervisability, intellectual property • Access to data, accessibility, interoperability, reusability, findability, reliability of outputs	• Ensuring human-in-the-loop approaches • Preventing excessive automation	• Explainability • Traceability, predictability, supervisionability, interpretability • Communication	• Avoidance of unfair bias • Accessibility and universal design • Stakeholder participation • Promoting equity of opportunity

Ethical label

Environmental wellbeing	Societal wellbeing	Human wellbeing	Technical/General
Energy consumption: end-user understandable estimation of energy consumption for the services offered or the training of AI algorithms. For instance, producers can make comparison with the energy needed by houses, saunas, etc.	*Society and democracy*: includes aspects concerning the influence of digital systems on democratic processes, institutions, political engagement, public deliberation as well as broader societal conditions, institutional transparency, and media pluralism. For instance, a news app may support democracy by showing articles from many sources, without hiding certain opinions or spreading false news, but rather checking them.	*Principle of autonomy*: if and how the system compromises user autonomy, understood as the ability of humans to act according to their informed beliefs, and limit human control of the system. For instance, human being able to take control over an autonomous system, or explaining when this is not possible, allowed or beneficial for humans.	*Minimization of risk/negative impact*: if and how the system has a plan to monitor, eliminate, or limit risks or negative impact as much as possible. For instance, organizations can have standardized risk management practices and processes for managing both existing and newly detected ethical risks.
Greenhouse Gas Emissions: end-user understandable estimation of gas emission for the services offered or the training of AI algorithms. To make it understandable to end-users, producers can make a comparison with the emission of vehicles.	*Openness and plurality/Cultural Sensitivities*: the system is applicable in different local contexts, respecting plurality and various cultures. For instance, an application that allows users to choose language, symbols, or customs based on their culture is inclusive.	*Privacy and data governance/ intellectual property*: Privacy must be respected throughout data collection, use, and sharing, with clear and accessible information provided to users. Intellectual property rights must also be safeguarded, particularly during AI training processes, preventing misuse or legal infringements. The governance structure of the system—including who controls and who accesses data—must be transparently disclosed.	*Mitigation of risk/negative impact*: Recognizing the inevitability of some risks, it is important to consider if and how the system has strategies to address the aftermath of a problem, as well as the steps that can be taken beforehand to reduce adverse and potentially long-term effects. For instance, when a collision is unavoidable, vehicles should minimize accelerations that humans experience to reduce the severity of injuries.
Sustainability in Use: end-user understandable explanation of the system's resource consumption during use that should be compared with its sustainability potential.	*Inclusive and Participatory Design*: The system should adhere inclusive design principles have been followed and stakeholders related to the system have been identified and consulted during development. The producer may clearly state whether the system was tested and shaped with feedback from different user groups, such as people with disabilities, elderly users.	*Transparency and explainability*: it is made clear to those who use or interact with a digital system, e.g., an AI-powered system, that AI is being used and that the resulting outcomes are transparent. Another perspective involves explaining the rationale behind the decisions made by the system.	*Reporting negative impact*: there is a strategy to make available to the public and users potential negative impacts of the system. For instance, there can be a plan and strategies to inform users about potential data leaks, together with the impact and the scope of it.
Indirect Resource Consumption: end-user understandable explanation of the indirect resource consumption of the system, e.g. in terms of water consumption. The explanation should clearly make examples and comparisons, e.g., in terms of average daily water consumption of a person.	*Diversity, fairness, and non-discrimination*: if and how the system aligns with the ideal of justice and promotes fairness, inclusion, respect for diversity, and equality of opportunity.	*Beneficialness*: to what extent the use of a product, system, or service is beneficial for humans and not harmful. The producer should clearly describe what are the benefits brought by the system and why its usage is not harmful. For instance, reminding users to take a break from screen exposure to reduce eye strain, or to take breaks while driving to avoid accidents.	
Sustainability Potential in Application: if and how the system exploits the sustainability potential, e.g., by promoting sustainable products or considering sustainability in the decision process or when computing recommendations. For instance a search engine could rank based also on sustainability.	*Respect of the rule of law*: the system must respect the rule of law of specific countries. Additionally, it should inform end-users on the correct way of using systems to avoid potential violations of laws. Examples include ethical filters in LLMs or autonomous systems.		

Ethic label: an application example

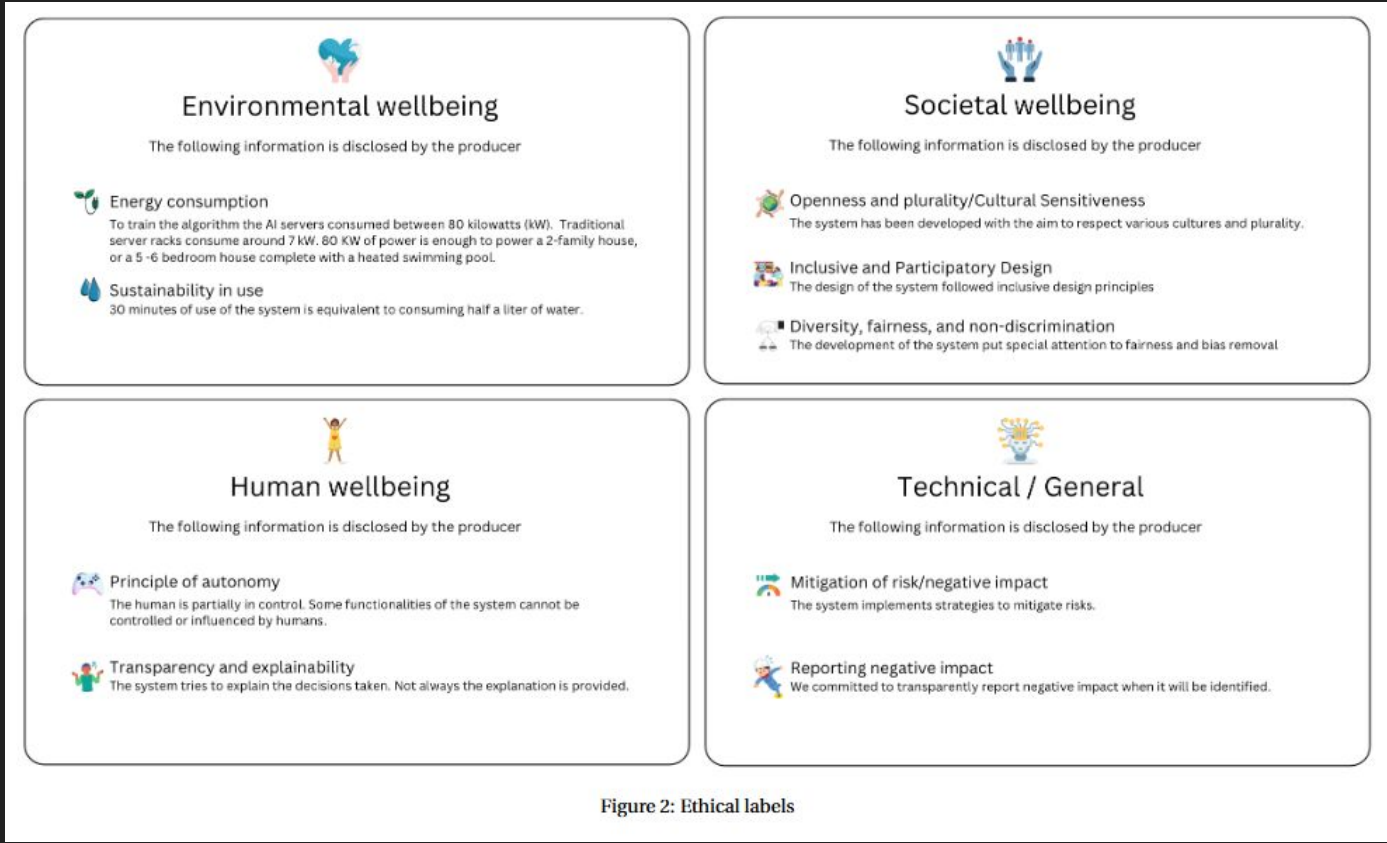
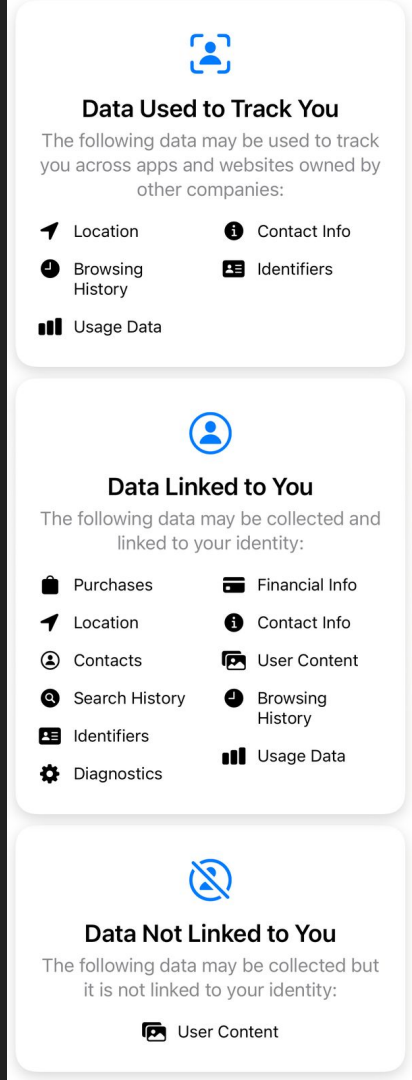


Figure 2: Ethical labels

Discussion

Q1: Have you ever heard of privacy or nutrition labels?

Privacy Label: example



Q2: Is the proposed ethics label easily understandable?

Evaluating Comprehensibility / Understability

Q3: Does the proposed label appropriately address Human, Societal, and Environmental drivers?

HSE Drivers

Table 1: Categories and subcategories of HSE drivers. The blue text is the extension with respect to [9].

Societal and Environmental Well-being	**Accountability and Responsibility**	**Privacy and Data Governance**	**Human Agency and Oversight**	**Transparency and Explainability**	**Diversity, Fairness, and Non-discrimination**
• Sustainable and environmental friendliness • Societal and social impact • Society and democracy • Respect of the rule of law: normativeness • AI and digital literacy • Openness and plurality	• Auditability • Minimization and reporting of negative impacts • Tradeoffs and redress	• Accuracy, completeness, consistency, timeliness, uniqueness in data quality, interpretability, coherence • Security and privacy considerations, controllability, adaptability, supervisability, intellectual property • Access to data, accessibility, interoperability, reusability, findability, reliability of outputs	• Ensuring human-in-the-loop approaches • Preventing excessive automation	• Explainability • Traceability, predictability, supervisionability, interpretability • Communication	• Avoidance of unfair bias • Accessibility and universal design • Stakeholder participation • Promoting equity of opportunity

Q4: The proposed label is complete in integrating ethics as a quality for end-users? Are there any important ethical dimensions which have been overlooked? If any, which dimensions do you feel are missing or require further clarifications?

Anything to add?

Thanks!