Rethinking Holistic AI Development Through Social Diversity, Interdisciplinary Collaboration and Integrative Knowledge Production

Cinzia Leone[1], Angela Celeste Taramasso[2] and Anna Siri[3]
[1]Istituto Italiano di Tecnologia, Italy
[2]Università degli Studi di Genova, Italy
[3]Università Telematica Pegaso, Italy

cinzia.leone@iit.it
Angela.Celeste.Taramasso@unige.it
anna.siri@unipegaso.it

Abstract: The rapid deployment of AI reveals persistent socio-technical and data-driven biases that reflect profound epistemic limitations in knowledge production. These biases are not accidental, but symptomatic of deeper epistemic limitations in the way AI knowledge is produced — often by homogeneous teams within technocentric paradigms that exclude alternative perspectives. This paper argues that the underrepresentation of diverse social actors in AI development not only perpetuates inequality, but also severely limits the epistemic and ethical robustness of AI systems. The focus of this paper arises in particular from the preliminary findings obtained in the Horizon Europe project STEP, which highlight the potential of the framework to improve the inclusivity and trustworthiness of AI. The central thesis is that social diversity must be considered as an epistemic condition and not just an ethical or demographic ideal. Drawing on sociology, psychology and educational science, the authors show how integrating plural forms of knowledge, lived experiences and cultural perspectives into the design and development process can lead to AI systems that are more context-sensitive, equitable and trustworthy. Rather than proposing inclusion as an external corrective, this paper discusses a paradigm shift in AI development - a paradigm shift that embeds diversity into the infrastructure of knowledge production itself. The contribution of this paper is twofold. First, it proposes a theoretical model of integrative knowledge production that identifies mechanisms through which interdisciplinary collaboration can challenge dominant epistemologies and promote systemic reflexivity. Second, a participatory design framework is outlined to operationalise this model through concrete methodological tools, including dialogic co-design workshops, ethnographic participation in data selection and cross-functional team structuring. These practises aim to break through technocratic compartmentalisation by creating space for social critique and situated intelligence within AI development cycles. Finally, the authors reflect on the transformative potential of this approach and suggest that rethinking who is involved in AI knowledge production will not only change the outcomes of AI systems, but also the normative foundations of the technological future. From this perspective, ethical AI is not just explainable or compliant — it is structurally inclusive, responsive to different lifeworlds and open to critical reinvention.

Keywords: Epistemic diversity, Algorithmic justice, Participatory AI, Interdisciplinary design, Sociology of technology, Inclusive innovation, Knowledge infrastructures, Human-centred AI, Bias

1. AI and Biases: Introduction and Background

AI systems that collect and process data on humans and human behaviour often become a mirror of social inequality (Ferrara, 2024). This bias is not only encoded in the data sets, but also results from the cultural, institutional and professional environment in which AI tools are designed and developed. This is often due to the lack of diversity of people involved in the development of AI systems and tools, as well as biases within science and culture/humanities. As Snow (1969) warned, the growing divide between the sciences and the humanities hinders the solution of complex problems and a gap that is still evident in AI knowledge production. He also argued that this growing divide is a significant obstacle to effectively tackling the complex challenges of the modern world. Indeed, even today we can observe that the process of knowledge production in the field of AI is dominated by homogeneous teams and technocentric worldviews and tends to exclude alternative knowledge systems, reinforcing existing hierarchies (Ferrara, 2024). These rely too heavily on data sets that favour, for example, the Western perspective and stereotypical assessments of the validity of information (Nersessian & Mancha, 2020; Zowghi & da Rimini, 2023). AI harbours the risk of perpetuating existing biases by reinforcing the perspectives and interests of dominant groups, while holistic, interdisciplinary knowledge and different cultural values fade into the background (Ahrweiler et al., 2025; Gichoya et al., 2023). This dynamic can lead to an overrepresentation of dominant cultural norms and an underrepresentation of minority identities and alternative perspectives or diversity (Roche et al., 2023). In this way, AI models become biassed tools that reflect and reproduce historical inequalities since their inception. This is particularly evident in biometric technologies, predictive policing, recruitment algorithms and health diagnostics - where AI often fails marginalised groups. The profound epistemic limitation reflects the societies and dominant group that create, implement and govern AI, in a vicious cycle that has proven flawed in recent years (De Sousa Santos, 2024).

Considering the aftermath of the rapid development of AI and the challenge of keeping up with its technological, scientific and substantive advances (Jungherr, 2023), this paper identifies specific scientific gaps. Indeed, this paper presents a fundamental response to this challenge: it proposes to invert the traditional model of AI development by integrating diverse voices - those that speak different (social) languages and have different worldviews and scientific backgrounds. This approach encourages developers and engineers to redesign tools through the lived experiences of their communities, moving from a model of technological imposition to a model of participatory co-creation and co-design, in a conscientious, reliable and human-centred design and development of the upcoming research and development of AI (Bencivenga et al., 2025).

The authors believe that this will help train a new generation of interdisciplinary researchers who are able to conceptualise and use AI to design and shape more inclusive and equitable societies by considering the opportunities and challenges related to gender, diversity and inclusion and shaping future planning and decision-making in a way that incorporates diverse perspectives and promotes equality (Ferrara, 2024). Capitalising on the synergy of diverse disciplines such as computer science, engineering, economics, sociology, law, pedagogy and education, this new AI development approach will tailor not only educational systems but entire social ecosystems to the dynamics of the future, while working towards integrity in scientific research, which would lead to more complete, consistent and holistic outcomes.

In response to the growing concern, the European Union has been at the forefront of regulation. The provisional agreement on the Artificial Intelligence Act (December 2023) (European Parliament, 2023) introduces a comprehensive framework of safeguards, restrictions and prohibitions, coupled with sanctions for non-compliance. The AI Act is the first of its kind in the world and sets a precedent for the regulatory framework worldwide. It was complemented by the *Ethics Guidelines for Trustworthy AI* (European Commission, 2019), issued by the High-Level Expert Group, which include the key principles of acting humanely, transparency, accountability and non-discrimination. At the same time, the European Commission has made significant progress in establishing ethical guidelines for AI researchers and the topic of AI ethics has been a focus of various projects and initiatives. The guidelines were a decisive step in this direction. Following an open consultation that received more than 500 comments, these guidelines have been thoroughly revised since their first draft was published in December 2018. However, they often remain abstract and offer limited practical tools for implementing diversity and inclusion. There is still a gap between the high-level ethical obligations and the daily reality of AI research and application (Floridi et al., 2018). Furthermore, legal restrictions, although necessary in some cases, can unintentionally stifle innovation (Aghion et al., 2023) - especially in socially sensitive areas of research where nuanced understanding is essential (Nersessian & Mancha, 2020). Moreover, it should be emphasised that there is little concrete practical advice to date on how to ensure that diversity and inclusion considerations are embedded in both specific AI systems and the broader global AI ecosystem (Zowghi, da Rimini, 2023). In addition, AI's reliance on historical data poses a challenge due to inherent biases, particularly in relation to gender equality and other equality, diversity and inclusion (EDI) approaches. Dealing with these biases is crucial for the integral and fair application of AI.

Today, the future appears not as a deterministic tool, but as part of a broader ecosystem of interventions that together bring about meaningful change. While the linear approaches of 20th century social sciences provide insights, today we recognise the complex interplay of numerous interacting variables and think it is time to explore the transformative role of new knowledge guided by both AI and components of EDI. Our framework operationalises these high-level ethical commitments by embedding diversity into AI development practises at the design stage.

2. From Epistemic Blind Spots to Participatory Knowledge

To address the persistent systemic problems that underlie the production of knowledge and AI-related tools today and reproduce the inherent biases of the society that produces them, we propose to rethink the foundations of AI knowledge production. This requires a move away from ethics as compliance and towards epistemic diversity as a fundamental principle. Particularly in recent decades, the ethical principle has become increasingly established as a principle that underpins the production of knowledge but also serves as a check in case of discrepancies, inconsistencies or misuse (Anderljung et al., 2025; Brundage et al., 2018). In this approach, biases or prejudices were often only identified after knowledge had been created, i.e. through a retrospective intervention after the knowledge had been produced (Brundage et al., 2018).

We propose a paradigm shift in which inclusion is not an afterthought or a subsequent parameter, but an embedded property of AI infrastructures - a property that influences every phase from design to implementation.

We draw on three interrelated areas:

- The sociology of technology, which explores how power relations shape scientific and technological practises.
- Educational science, which investigates how knowledge is co-constructed and passed on in different populations.
- Psychology, which explores how cognitive biases and group dynamics influence decision-making and creativity.

Together, these fields offer insights into how plural perspectives can be mobilised as epistemic enrichment rather than treated as marginal additions.

Drawing on the sociology of knowledge (Berger & Luckmann, 1966) and critical pedagogy (Freire, 1970), we define epistemic diversity as the integration of diverse perspectives, cultural narratives and experiential knowledge into the AI knowledge production system. Rather than assuming objectivity through abstraction, this perspective emphasises contextual, lived realities as sources of innovation and critical insight. As Pragya affirms,

> *Supposed neutrality and distance created by technology makes it harder to bring the power imbalance and inequalities to light. The manner in which the use of such tools has caught on in the research and academic sector makes it crucial to talk about who makes knowledge (Pragya, 2025: 2341).*

We are aware that this is a controversial topic that can be analysed from many different angles (Marti & Recupero, 2021; Friedman, & Nissenbaum, 1996). We are dealing here with an evolving science that requires the interpenetration and cross-fertilisation of different fields of knowledge and disciplines, and we certainly do not claim to be exhaustive in this paper. But we are definitely interested in asking the question of where knowledge originates and develops so that we can better analyse the biases and biases that very often underlie it (Jungherr, 2023; Jasanoff, 2004).

At the same time, we do not want to give too general an answer to a specific and precise problem. We are aware that it is not a question of proposing a black and white alternative to the ethical issues surrounding the development of artificial intelligence. Nor is the proposal of the EDI approach a simple or easy solution, precisely because of the complexity of the question we are trying to answer. We also recognise that the concepts associated with an inclusive approach that embraces equality and diversity are themselves under scrutiny (Stinson & Vlaad, 2024; Noble, 2018), and the debate remains open and evolving (Benjamin, 2019).

It is therefore essential to anchor any ethical framework in a reflexive and interdisciplinary dialogue that is sensitive to context, power dynamics and historical inequalities — rather than applying universal or standardised models that run the risk of reproducing the same biases they are supposed to mitigate. Such a shift requires a deep epistemological engagement with the question of what inclusion in AI development really means and who is involved in its definition (Costanza-Chock, 2020; Stinson & Vlaad, 2024).

3. Participatory Design and Implementation Framework

In order to operationalise epistemic diversity and further develop the above, we propose a participatory framework (Ahrweiler et al. 2025, Özer, 2024), which has already experimented in the ongoing Horizon Europe STEP project - *STEM and Equality, Diversity and Inclusion: an open dialogue for research enhancement in Portugal* (2023-2025), G.A. N. 101078933, with three methodological pillars:

1) Dialogic co-design workshops: These would bring together interdisciplinary teams with community representatives to collaboratively develop design goals, evaluate modelling assumptions and define success metrics. These workshops follow a dialogue-based pedagogy (Freire, 1970) and are based on theories of co-production (Jasanoff, 2004).

Each workshop consists of three iterative phases:

- Pre-workshop: community representatives and developers jointly select use cases through participatory mapping (e.g. Miro boards with 20+ stakeholders[1]).
- Implementation: 2-day sessions with role-playing bias scenarios (e.g. simulating how a hiring algorithm could exclude non-native speakers).
- Post-workshop: Voting on design priorities using the Delphi method (Özer, 2024), with anonymised feedback integrated into Jira/GitHub tickets[2].

Selection criteria: at least 40% participants from marginalised groups (as per UN SDG 10 indicators), with gender-balanced facilitation teams.

2) Ethnographic data practises: Rather than simply extracting data or using pre-existing databases, researchers from different scientific fields and disciplines collaborate where possible to understand how data is contextualised and interpreted within communities. This avoids epistemic extractivism (Medina, 2013; De Sousa Santos, 2024) and promotes relational accountability.

Data collection combines:

- Immersion: Researchers spend more than 15 days in the target communities and use dense description (Geertz, 1973) to document contextual factors (e.g., how older users interact with voice AI in a noisy environment).
- Tools: open-source platforms such as Open Ethnographer for semantically analysing field notes validated by member checking.
- Quality control: annotations of datasets must achieve 0.8+ Krippendorff's alpha for intercoder reliability of bias labels (Nili et al., 2017).

3) Cross-competence teaming: structuring of AI development teams must involve various experts from the social sciences and humanities (SSH) from the outset (Marti & Recupero, 2021).

The effective implementation of the framework requires consciously assembled teams that reflect epistemic diversity. One possible configuration envisages different but interdependent roles, including technical developers who integrate fairness specifications and translate participatory insights into concrete technical changes. Complementing this are social scientists who monitor societal impact through intersectional approaches, and community liaisons who reconcile lived experiences with technical design by collaboratively developing culturally appropriate testing protocols and shared deployment decisions. While hypothetical, this configuration is based on emerging principles in the literature and can serve as a blueprint for further empirical research. These teams are trained in EDI principles alongside core AI competences and are tasked with continuous critical evaluation of development trajectories.

Effective implementation of the proposed framework requires not only teams that reflect epistemic diversity, but also mechanisms to manage potential disagreements, particularly on contentious issues. While the framework emphasises full participation and engagement, it recognises that not all concerns can be addressed simultaneously. To overcome this challenge, the inclusion of a facilitator role is essential, acting as a transversal mediator throughout the development process. This role, akin to a facilitator or coordinator, would ensure that different perspectives are constructively integrated and that, where necessary, adjustments are made to maintain a co-operative and inclusive environment. Such a role is critical to managing conflict and fostering productive dialogue, ensuring that development remains focused and aligned with the overall goals of the project.

To ensure reproducibility, each pillar follows a standardised protocol based on the principles of participatory action research (PAR) (Cornwall & Jewkes, 1995; Chevalier & Buckles, 2019). The overarching goal is to elucidate PAR's efficacy as a collaborative, inclusive, and empowering research paradigm that actively involves

---

[1] Miro Boards are collaborative online whiteboards that allow multiple users to work together in real time, regardless of location. In the context of your project, using Miro Boards with more than 30 stakeholders means that more than thirty participants— - e.g. community representatives, developers, social scientists, and moderators — work together in the same virtual space to capture use cases, share ideas, and visually organise priorities.

[2] After the workshops, participants provide anonymous feedback, which is systematically recorded and integrated into the technical development process with the help of tools such as Jira or GitHub. "Jira" is a project management tool to track tasks, bugs, and features. "GitHub" is a platform for collaborative programming where issues and feature requests are logged along with code repositories. By converting the anonymous workshop feedback into Jira/GitHub tickets, the community members' insights are translated directly into realisable technical tasks. This ensures that participants' contributions are not lost, but become a visible part of the development pipeline, subject to accountability and follow-up.

stakeholders in the knowledge co-creation process (Kemp et al., 2019). Originating from critical social theory and adult education, PAR distinguishes itself through its commitment to social justice and democratic principles, fostering a research environment where affected communities are not merely subjects but co-investigators and agents of change, thereby shifting research control to those most impacted by its outcomes (Torre et al., 2015). The participatory framework (Figure 1) integrates iterative co-design, ethnographic engagement and structured evaluation to ensure that societal considerations are embedded throughout the AI development lifecycle. The model emphasises cyclical exchanges between technical teams and community stakeholders, supporting both the identification of potential biases and the refinement of system requirements. It can be deduced that the three concepts are intertwined, cross-fertilise each other and have common intersections that can contribute even more to the advancement of knowledge from its initial production, based on a methodology that is as innovative as it is simple.

Figure 1: Participatory framework concept

From a qualitative perspective, the preliminary results from the experiences of the above-mentioned EU project underline the value of dialogue-based and participatory approaches in the design of responsible AI development. The co-design workshops have shown that structured dialogue, participatory mapping, and iterative feedback mechanisms can effectively translate diverse community perspectives — especially from marginalised/rejected groups — into actionable design goals. Complementing this, ethnographic data practises offered rich insights into the contextual dynamics of data use, and emphasised the importance of relational accountability and validation of annotations through member checking and intercoder reliability measures. Furthermore, early experiments with cross-competency teams demonstrated the potential of epistemically diverse teams to embed fairness constraints and cultural sensitivity into technical processes. Taken together, these experiences show that inclusive methods not only broaden the range of design priorities, but also promote trust and inclusivity in AI practise. However, it is important to point out that these findings are still preliminary, as they come from a single implementation within the project and will be the subject of further research and publication. The following section therefore critically analyses limitations and outlines future research opportunities.

The ultimate goal of participatory design is to cultivate researchers who are both technically proficient and EDI-aware and ethically conscious. Taken together, these practises aim to break through the technocratic compartmentalisation that often characterises AI development and create space for social critique, situated intelligence and pluralistic values (Avellan et al., 2020). Rather than seeing ethics as a constraint or add-on, the proposed model understands ethical engagement as a form of epistemological intervention - one that is integral to the innovation process itself (Bencivenga et al., 2025). The result is a more holistic, adaptive and equitable approach to AI development that recognises the socio-technical nature of intelligent systems and the multiplicity of people they affect. In this way, the degree of innovation can be measured not only with the well-known Technological Readiness Level (TRL, Figure 1) (Yfanti & Sakkas, 2024), but it is also possible to extend the application to the Societal Readiness Level (SRL, Figure 2) (Innovation Fund Denmark, 2019; Leone et al., 2024) and embody its critical and active role in the design and adoption of inclusive technologies (Bernstein et al., 2022). We therefore propose a value scale, as shown in the figure below, that considers the value of innovation based on its acceptability at the societal level (Obrien et al., 2024).

Phase	TRL	Description
Research	1	Basic principles observed and described
Research	2	Technology concept and/or applications formulated
Research	3	Analytical or experimental proof of principal functions and/or concept characteristics
Development	4	Validation of components and/or mock-ups in laboratory
Development	5	Validation of components and/or mock-ups in representative environment
Development	6	Demonstration of a prototype or system/subsystem model in a representative environment
Deployment	7	Demonstration of a system prototype in an operational environment
Deployment	8	Actual system completed and qualified through tests and demonstrations
Deployment	9	Actual system completed and qualified through successful operational missions

Figure 1: TRL scale

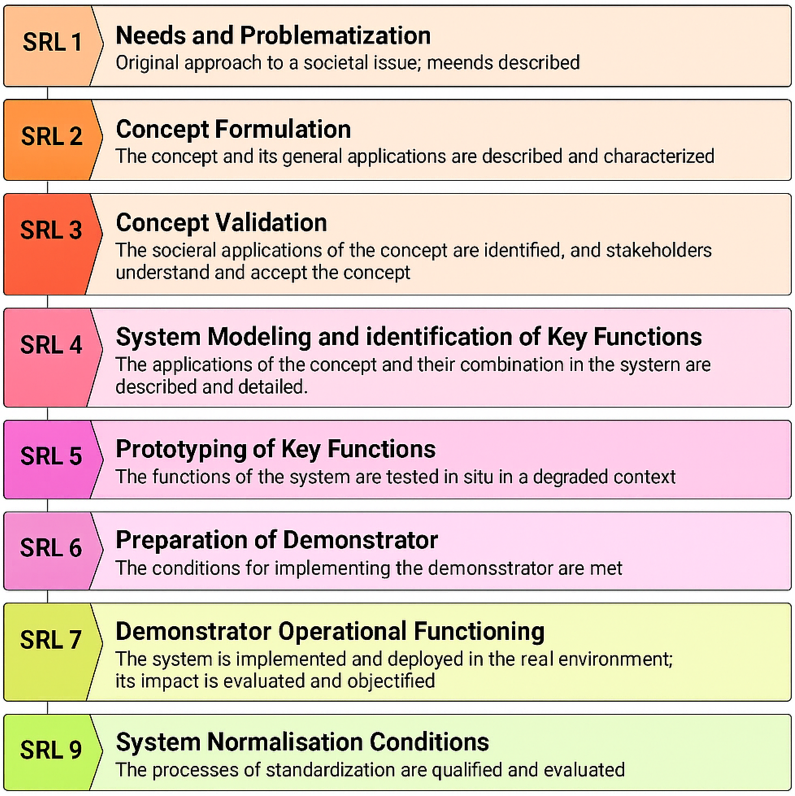

Figure 2: SRL scale

The SRL scale (Verma & Allen, 2024; Owen et al., 2012) assesses the social acceptability of innovation across 9 levels, ranging from problem recognition (level 1) to cultural integration (level 9) (Bruno et al., 2020). The scale is now used internationally and has several variants. Our intervention plan within the STEP project aimed to increase the SRL from 3 to 7, which represents a significant improvement in societal acceptance. By linking SRL and TRL, developers can assess both technical readiness and societal acceptance, mitigating the risks of socially misguided innovation.

4. Limitations and Future Implementation

The proposed framework opens up several critical avenues for future research and practical development at the intersection of AI and social inclusion and equality. While the STEP project demonstrates its potential to improve societal readiness and reduce bias, further work is needed to consolidate its methodological robustness, ensure its scalability and support its adaptation to diverse global contexts.

A first priority is the development of standardised metrics that can be used to assess the actual impact of participatory processes on model performance and bias reduction. Future research should explore the use of fairness metrics such as the Disparate Impact Ratio across large data sets and introduce sub-indices of SRL, including a Gender Bias Index and a Cultural Sensitivity Score, to capture intersectional inequalities that aggregate SRL levels cannot fully uncover. In addition to these tools, the concept of return on investment for inclusion (ROI) (Hoffman et al., 2011) could help quantify the long-term benefits of early participatory design, such as lower mitigation costs and higher social acceptability relative to the upfront investment required.

The here proposed framework offers a transformative approach to inclusive AI development, but several limitations and challenges are recognised. Firstly, the approach is very resource intensive. Dialogic co-design workshops and ethnographic dives require a significant investment of time and human input, which can limit scalability - especially for small projects or underfunded public initiatives. Secondly, persistent power asymmetries can undermine participatory processes. Even in an inclusive environment, community participants or junior contributors may self-censor when experienced developers or researchers are present.

Thirdly, while the SRL metric is innovative in terms of assessing social acceptability, it lacks the granularity to capture intersectional differences in highly heterogeneous social contexts. Several mitigation strategies are currently being developed to remedy this. In addition, the possible development of SRL sub-indices — such as a Gender Bias Index and a Cultural Sensitivity Score and others — should allow for a more nuanced assessment of inclusion outcomes.

Despite these efforts, there remains a fundamental tension between the depth of participatory inclusion and the efficiency of technical development. Applications where the stakes are high, such as AI in healthcare, may warrant extended and longer participatory phases, even if this results in delayed adoption, while commercial applications may require lighter but more frequent bias audits. Future implementations will likely rely on configurable levels of participation that calibrate this trade-off according to sectoral and contextual needs.

Taken together, these future implementations point to the evolution of the proposed framework into a globally adaptable model for inclusive AI. By combining rigorous metrics, decolonised practises, sustainable participation strategies and scalable design methods, the approach can ensure that AI systems are not only technically robust, but also socially legitimate, culturally responsive and sustainable in the long term, bridging the persistent gap between technological innovation and societal trust.

5. Conclusions

This paper proposes a participatory framework for the development of AI that incorporates epistemic diversity in the design phase and builds a bridge between technical and societal readiness. This is necessary to find breakthrough solutions to the current gaps in dealing with algorithmic biases and the design, production and implementation of AI. While much of the research focuses on identifying biases in AI systems and trying to avoid biases in post-production, we believe that we should take a broader action-oriented approach by developing tools and methods to avoid, prevent and mitigate these biases, especially those related to EDI. Including diverse voices and cultural perspectives in AI development from the design stage represents a significant step forward in creating equitable and inclusive technologies that can thus certify integrity (Almufareh, 2025). Incorporating EDI principles and gender-sensitive design principles into AI from the outset therefore requires developers to change the system of data collection and represents a fundamental shift in knowledge: finding and incorporating sources that are not biassed is a challenge, but it can and should still be done. While it is undeniable that, especially in recent years, powerful players — whether in companies or at government level — have often resisted EDI frameworks or dismissed the urgency of the debate with algorithmic bias, it is precisely this resistance that emphasises the need for systemic reflexivity in AI development. The reluctance of influential stakeholders to recognise structural gaps highlights the asymmetry between those who benefit from current technological trajectories and those who are disproportionately exposed to their risks. This divergence of interest illustrates why integrative knowledge production, based on interdisciplinary collaboration and participatory approaches, is so important: it creates mechanisms through which alternative epistemologies and lived experiences can exert influence, even when political or market incentives are misaligned. While supranational organisations such as the European Union may have stronger regulatory influence to counterbalance these power dynamics, other contexts demonstrate the vulnerability of smaller governments to industry capture. Therefore, the proposed framework not only addresses the technical challenges, but also situates AI development within the broader struggles of governance, legitimacy, and societal accountability.

In contemporary innovation processes, it is becoming increasingly essential to assess not only the technical feasibility of a particular technology, but also its potential for social acceptance, adoption and integration within the specific communities or social groups for which it is intended. The success of an innovation depends not only on its technological maturity, but also on its fit with the cultural, ethical and contextual realities of its users. Consequently, the integration of TRL with SRL becomes an important analytical and strategic framework. Such integration enables researchers, developers and policy makers to assess both the technical development and the societal viability of innovations simultaneously. This dual consideration helps to minimise the risk of technological failure due to social resistance, misalignment with user needs or a lack of contextual awareness. Ultimately, aligning TRL and SRL increases the likelihood that innovations are both technically robust and socially meaningful, contributing to more sustainable and responsible innovation outcomes.

In this regard, it is crucial that the development of innovation — especially in the field of artificial intelligence — is guided from the outset by considerations of social readiness and participatory engagement. Rather than treating SRL, EDI and participatory approaches as secondary or downstream concerns, they should be embedded from the conceptualisation and design phase of technological development. This means that various stakeholders — including end users, civil society actors, representatives of marginalised groups and others — are actively involved from the earliest stages of ideation and prototyping. Such participatory approaches not only increase the social legitimacy and relevance of AI systems, but also help to recognise potential ethical, cultural and structural impacts before deployment. The early inclusion of societal concerns in the innovation process ensures that AI solutions are not only technically advanced, but also context-sensitive and in line with collective values and long-term societal goals.

Incorporating SRL-driven participatory methods from the earliest stages of AI development strengthens societal trust and supports ethically robust innovation.

Acknowledgements

This work was partly funded by the Horizon Europe STEP Project, under Grant Agreement No. 101078933.

Ethical statement: The research presented in this article did not involve human participants, collection of personal data, or interventions requiring formal ethical approval according to the authors' institutional guidelines. Therefore, no ethical approval was required. The authors confirm that the study was conducted in accordance with the principles of research integrity and the applicable ethical standards of their respective institutions.

Statement on the use of artificial intelligence tools: The authors declare that they have not used any artificial intelligence tools to generate, modify, or revise the content of this article, with the exception of the title. All texts, analyses, and conclusions are the exclusive result of the authors' work.

References

Aghion, P., Bergeaud, A., & Van Reenen, J. (2023). The Impact of Regulation on Innovation. *American Economic Review, 113*(11), 2894–2936. https://doi.org/10.1257/aer.20210107

Ahrweiler, P., Späth, E., Siqueiros García, J. M., Capellas, B. L., & Wurster, D. (2025). Inclusive technology co-design for participatory AI. Participatory Artificial Intelligence in Public Social Services: From Bias to Fairness in Assessing Beneficiaries. In Ahrweiler, P. (Ed.). *Participatory Artificial Intelligence in Public Social Services: From Bias to Fairness in Assessing Beneficiaries* (p. 35-62). Springer Nature. https://doi.org/10.1007/978-3-031-71678-2_2

Almufareh, M. F., Kausar, S., Humayun, M., & Tehsin, S. (2024). A conceptual model for inclusive technology: advancing disability inclusion through artificial intelligence. *Journal of Disability Research, 3*(1), art. N. 20230060. https://doi.org/10.57197/JDR-2023-0060

Anderljung, M., Hazell, J., & von Knebel, M. (2025). Protecting society from AI misuse: when are restrictions on capabilities warranted?. *AI & Society, 40*, 3841-3857.

Avellan, T., Sharma, S., & Turunen, M. (2020). AI for all: defining the what, why, and how of inclusive AI. In *Proceedings of the 23rd International Conference on Academic Mindtrek,* 142-144. https://doi.org/10.1145/3377290.3377317

Bencivenga, R., Boutenbat, H., Taramasso, A.C., Leone, C. (2025, *in print*) Langage Inclusif et Societal Readiness Level (SRL): Garantir des Innovations Technologiques Accessibles et Adoptables. In: *Atti della Journée des Etudes Inclusion, communication institutionnelle et traduction*, University of Torino, 5 December 2024.

Benjamin, R. (2019). *Race after technology: Abolitionist tools for the new Jim code*. Polity Press.

Berger, P. L. & Luckmann, T. (1966). *The Social Construction of Reality*. Anchor Books.

Bernstein, M.J., Nielsen, M.W., Alnor, E. et al. (2022). The Societal Readiness Thinking Tool: A Practical Resource for Maturing the Societal Readiness of Research Projects. *Sci Eng Ethics*, *28*(6). https://doi.org/10.1007/s11948-021-00360-3

Brundage, M., Avin, S., Clark, J., Toner, H., Eckersley, P., Garfinkel, B., Dafoe, A., Scharre, P., Zeitzoff, T., Filar, B., Anderson, H., Roff, H., Allen, G.C., Steinhardt, J., Flynn, C., Éigeartaigh, S.Ó., Beard, S., Belfield, H., Farquhar, S., Lyle, C., Crootof, R., Evans, O., Page, M., Bryson, J., Yampolskiy, R., Amodei, D. (2018). *The malicious use of artificial intelligence: forecasting, prevention, and mitigation.* Report edited by Future of Humanity Institute, University of Oxford, Centre for the Study of Existential Risk, University of Cambridge, Center for a New American Security, Electronic Frontier Foundation, OpenAI. https://maliciousaireport.com (last accessed 1/8/2025).

Bruno, I., Lobo, G., Covino, B.V., Donarelli, A., Marchetti, V., Panni, A.S. and Molinari, F. (2020). Technology readiness revisited: a proposal for extending the scope of impact assessment of European public services. *Proceedings of the 13th international conference on theory and practice of electronic governance*. Association for Computing Machinery. Athens, Greece, 23-25 September, 369-380.

Chevalier, J.M. & Buckles, D.J. (2019). *Participatory Action Research. Theory and Methods for Engaged Inquiry*. Routledge. https://doi.org/10.4324/9781351033268

Cornwall, A. & Jewkes, R. (1995). What is participatory research?. *Social science & medicine*, *41*(12), 1667-1676. https://doi.org/10.1016/j.socscimed.2009.11.005

Costanza-Chock, S. (2020). *Design Justice: Community-Led Practices to Build the Worlds We Need*. MIT Press.

De Sousa Santos, B. (2024). AI and the epistemologies of the South. *Journal of World-Systems Research*, *30*(2), 635-645. https://doi.org/10.5195/JWSR.2024.1291

European Commission (2019). Ethics Guidelines for Trustworthy AI. High-Level Expert Group on Artificial Intelligence. https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai (last accessed 10/7/2025).

European Parliament (2023). *Artificial Intelligence Act: deal on comprehensive rules for trustworthy AI*. https://www.europarl.europa.eu/news/en/press-room/20231206IPR15699/artificial-intelligence-act-deal-on-comprehensive-rules-for-trustworthy-ai (last accessed 10/7/2025).

Ferrara, E. (2024). Fairness and Bias in Artificial Intelligence: A Brief Survey of Sources, Impacts, and Mitigation Strategies. *Sci*, *6*(1), 3. https://doi.org/10.3390/sci6010003

Floridi, L., Cowls, J., Beltrametti, M., Chatila, R., Chazerand, P., Dignum, V., ... & Schafer, B. (2018). AI4People - An ethical framework for a good AI society: Opportunities, risks, principles, and recommendations. *Minds and Machines*, 28(4), 689-707.

Freire, P. (1970). *Pedagogy of the Oppressed*. Continuum.

Friedman, B., & Nissenbaum, H. (1996). Bias in computer systems. *ACM Transactions on Information Systems, 14*(3), 330–347.

Gichoya, J. W., Thomas, K., Celi, L.A., Safdar, N., Banerjee, I., Banja, J. D., ... & Purkayastha, S. (2023). AI pitfalls and what not to do: mitigating bias in AI. *The British Journal of Radiology, 96*(1150), 20230023. https://doi.org/10.1259/bjr.20230023

Hoffman, S., Lane, R., & Posner, D. Measurement: Proving the ROI of global diversity and inclusion efforts. In Meadows, A.J. and Tapia, A.T. (Eds) (2011). *Global Diversity Primer* (pp 129-136). Diversity Best Practices.

Innovation Fund Denmark (2019). *The Technology Readiness Index primer*. https://innovationsfonden.dk/sites/default/files/2019-03/societal_readiness_levels_-_srl.pdf (last access 29/8/2025).

Jasanoff, S. (2004). *States of Knowledge: The Co-production of Science and the Social Order*. Routledge.

Jungherr, A. (2023). Artificial Intelligence and Democracy: A Conceptual Framework. *Social Media + Society*, *9*(3). https://doi.org/10.1177/20563051231186353

Kemp, L., Bailey, D., & Barnard, A. (2019). *Doing Participatory Action Research: Reflections on Criticality and Social Justice from the Researchers' Perspective*. University of Lodz Press. https://doi.org/10.18778/8142-348-9.15

Leone, C., Bencivenga, R., Siri, A (2024). Advancing Science and Society: Unveiling the Societal Readiness Level (SRL) for Holistic Integration of Innovations. In: *Proceedings of the COMPASS Conference: Transferable Skills for Research & Innovation,* 4-5/10/2023, Helsinki, Finland, Haaga Helia Green O/A https://julkaisut.haaga-helia.fi/en/compass_topic_7/ (last accessed on 3/7/2025).

Marti, P. & Recupero, A. (2021). Thinking inclusion through making. In: *Conference Proceedings of the AIUCD 2021. DH per la società: e-guaglianza, partecipazione, diritti e valori nell'era digitale. Raccolta degli abstract estesi della 10a conferenza nazionale (pp.62-69).* http://hdl.handle.net/11365/1170587 (last accessed 11/7/2025).

Medina, J. (2013). *The Epistemology of Resistance: Gender and Racial Oppression, Epistemic Injustice, and Resistant Imaginations*. Oxford University Press.

Muldoon, J., & Wu, B. A. (2023). Artificial intelligence in the colonial matrix of power. *Philosophy & Technology, 36*(4), 80. https://doi.org/10.1007/s13347-023-00687-8

Nersessian, D., & Mancha, R. (2020). From automation to autonomy: legal and ethical responsibility gaps in artificial intelligence innovation. *Mich. Tech. L. Rev.*, *27*(55).

Nili, A., Tate, M., & Barros, A. (2017). A critical analysis of inter-coder reliability methods in information systems research. *ACIS 2017 Proceedings, 99*. https://aisel.aisnet.org/acis2017/99 (last accessed 1/8/2025).

Noble, S. U. (2018). *Algorithms of oppression: How search engines reinforce racism*. NYU Press.

Obrien, K., Roberts, D.-L., Johnson, W., Meschewski, E., Guske, E., Figueroa, J., Musaad, S. (2024). Including Societal Benefits in Energy Project Development: The Evolution of the Societal Readiness Level (SRL), *Proceedings of the 17th Greenhouse Gas Control Technologies Conference (GHGT-17) 20-24 October 2024.* http://dx.doi.org/10.2139/ssrn.5023609

Owen, R. Macnaghten, Ph., & Stilgoe, J. (2012). Responsible research and innovation: From science in society to science for society, with society. *Science and Public Policy, 39*, 751-760. http://dx.doi.org/10.1093/scipol/scs093

Özer, M., Perc, M., & Suna, H. E. (2024). Participatory management can help AI ethics adhere to the social consensus. *İstanbul Üniversitesi Sosyoloji Dergisi, 44*(1), 221-238. https://doi.org/10.26650/SJ.2024.44.1.0001

Pragya, A. (2025). Generative AI and epistemic diversity of its inputs and outputs: call for further scrutiny. *AI & Soc* 40, 2341–2342. https://doi.org/10.1007/s00146-024-02097-6

Roche, C., Wall, P.J. & Lewis, D. (2023). Ethics and diversity in artificial intelligence policies, strategies and initiatives. *AI Ethics* 3, 1095–1115. https://doi.org/10.1007/s43681-022-00218-9

Snow, C.P. (1969). *The two cultures: and a second look.* Cambridge University Press.

Stinson, C., & Vlaad, S. (2024). A feeling for the algorithm: Diversity, expertise, and artificial intelligence. *Big Data & Society*, *11*(1), https://doi.org/10.1177/20539517231224247

Torre, M. E., Cahill, C., & Fox, M. (2015). *Participatory Action Research in Social Research*. Elsevier BV. https://doi.org/10.1016/b978-0-08-097086-8.10554-9

Verma, A., & Allen, T. (2024). A Sociotechnical Readiness Level Framework for the Development of Advanced Nuclear Technologies. *Nuclear Technology*, *210*(9), 1722–1739. https://doi.org/10.1080/00295450.2024.2336355

Yfanti, S. and Sakkas, N. (2024). Technology Readiness Levels (TRLs) in the Era of Co-Creation. *Applied System Innovation, 7*(2), 32. https://doi.org/10.3390/asi7020032

Zowghi, D., da Rimini, F. (2023). *Diversity and Inclusion in Artificial Intelligence*. https://doi.org/10.48550/arXiv.2305.12728