Journal of Responsible Innovation

Routledge
Taylor & Francis Group

ISSN: 2329-9460 (Print) 2329-9037 (Online) Journal homepage: www.tandfonline.com/journals/tjri20

ELSA Labs for responsible AI: a novel approach for addressing ethical, legal, social issues

Hao Wang , Vincent Blok & Mireille van Hilten

To cite this article: Hao Wang , Vincent Blok & Mireille van Hilten (2025) ELSA Labs for responsible AI: a novel approach for addressing ethical, legal, social issues, Journal of Responsible Innovation, 12:1, 2563944, DOI: 10.1080/23299460.2025.2563944

To link to this article: https://doi.org/10.1080/23299460.2025.2563944

Published online: 10 Oct 2025.

JOURNAL OF RESPONSIBLE INNOVATION
2025, VOL. 12, NO. 1, 2563944
https://doi.org/10.1080/23299460.2025.2563944

Routledge
Taylor & Francis Group

RESEARCH ARTICLE

OPEN ACCESS

Check for updates

ELSA Labs for responsible AI: a novel approach for addressing ethical, legal, social issues

Hao Wang[a], Vincent Blok[b] and Mireille van Hilten[c]

[a]Faculty of Social Sciences, Wageningen University, Wageningen, the Netherlands; [b]Faculty of Philosophy Department, Erasmus University, Rotterdam, the Netherlands; [c]Wageningen Social & Economic Research, Wageningen University & Research, Wageningen, the Netherlands

ABSTRACT

Artificial Intelligence (AI) is rapidly transforming our society, offering remarkable opportunities but also raising significant Ethical, Legal, and Social Aspects (ELSA) that should be addressed for responsible development. Some existing approaches to responsible AI successfully translate ELSA into concrete AI design practices but risk overlooking power dynamics and structural issues, while others excel at fostering dialogue yet struggle to turn insights into real design changes. This paper develops the ELSA Lab approach as a promising way to bridge this gap. Building on research in Responsible Research and Innovation (RRI), Social Labs, and Quadruple Helix (QH) collaboration, we show how this approach combines the strengths of practical, solutionist strategies with sufficient negotiation and reflexivity. We not only outline the key features of this ELSA Lab approach theoretically but also present a hands-on work process for putting it into practice. This approach aims to drive a systemic shift toward more responsible AI.

ARTICLE HISTORY
Received 23 December 2024
Accepted 16 September 2025

KEYWORDS
ELSA; responsible research and innovation; social lab; Quadruple Helix stakeholder; trustworthy AI

Introduction

Artificial Intelligence (AI) has the potential to address some of our most challenging social and environmental issues ranging from healthcare and food security to poverty and ecological sustainability (Floridi 2023; UNESCO 2021). However, the development and use of AI also bring up some significant ethical, legal, and social aspects (ELSA) that we need to address. On the ethical side, for example, there are worries about potential bias and discrimination (Leslie 2019), loss of autonomy and freedom (Yeung 2017), lack of transparency (Sax and Wang 2025; Wang 2022a), and harmful manipulation (Susser, Roessler, and Nissenbaum 2019). Some legal aspects include the misuse of personal data (Menard and Bott 2025), lack of accountability and liability (Scherer 2015), and intellectual property violation (Samuelson 2023). Socially, AI could impact marginalized communities (Birhane et al. 2022), lead to job displacement (Tiwari 2023), harm the environment (Van Wynsberghe 2013), concentrate economic power (Zuboff 2019), and erode social trust (Wang 2022b). Given all this, there is a widespread belief that we should align AI development with social values and interests to ensure more responsible AI (AI HLEG 2019; Dignum 2019).

In recent years, two main types of approaches have emerged to make AI developments more responsible. The first is by-design approaches, such as Value Sensitive Design of AI (Sadek, Calvo, and Mougenot 2023; Umbrello and Van de Poel 2021), Ethics by Design (Brey and Dainow 2024), and Design for Values (Buijsman, Klenk, and Van den Hoven 2025), among others. These approaches often take a practical and solutionist strategy, aiming to translate abstract ethical values into concrete design requirements and solutions for AI engineers (Brey and Dainow 2024). This translation process makes ethics actionable for AI engineers, helping them more easily address issues related to specific AI artefacts and thus making it easier to embed ethics into practical AI development. But some critics have shown that these approaches tend to focus on isolated AI artefacts and translating implementable values (Smolka and Böschen 2023),

CONTACT Hao Wang hao2.wang@wur.nl Faculty of Social Sciences, Wageningen University, P.O. Box 8130, 6700 EW Wageningen, the Netherlands

often overlooking broader power dynamics and structural issues that require social negotiation rather than by technical design alone (Ryan et al. 2024; Wang and Blok 2025; Bolte and Van Wynsbergh 2025). In this sense, by-design approaches excel in embedding ethics into AI development, but run the risk that this embeddedness comes at the expense of sufficient reflexivity on these more structural ELSA aspects (Blok 2024).

In contrast, there are also more systemic approaches for responsible AI, such as Responsible Research and Innovation (RRI) (Stilgoe, Owen, and Macnaghten 2013) and Social Lab (SL) (Marschalek et al. 2022; Timmermans et al. 2020), that stresses the role of public engagement and view AI as a complex socio-technical ecosystem intertwined with social and power structures. These approaches emphasize negotiation and social inclusion, recognizing that heterogeneous values often conflict and ensuring that diverse stakeholders, particularly disadvantaged groups, have a voice in shaping AI (Ludwig et al. 2022). These approaches are effective at identifying power dynamics and complex social-ethical issues, but a persistent challenge is how to translate these negotiated insights and complex nuances into concrete design changes (Fisher and Maricle 2015). Numerous frameworks have provided valuable insights into embedding social scientists within scientific projects to offer real-time input and more effectively influence the direction of research and innovation. These include Real-Time Technology Assessment (Guston and Sarewitz 2002), Upstream Engagement (Wilsdon and Willis 2004), Anticipatory Governance (Barben et al. 2008), Socio-Technical Integration Research (STIR) (Fisher and Schuurbiers 2013), Human Practices (Rabinow and Bennett 2012), and the post-ELSI approach (Balmer et al. 2015), among others. However, in practice, ethical perspectives are often added onto existing projects as secondary considerations, limiting their ability to drive meaningful change (Conley et al. 2020). Moreover, complex socio-ethical issues (like the power dynamics) that matter to the public may not be prioritized by technical teams – or they may not know how to translate such structural concerns into concrete design choices (Ryan et al. 2024). Hence, the challenge remains: how can social and ethical insights exert real, tangible influence on AI development instead of just having a seat at the table?

In this paper, we argue that the emerging ELSA Lab approach has the potential to combine the solutionist strategies of by-design approach and the negotiation and reflexivity of systemic approach, offering a more effective way to develop responsible AI. Over the past few years, the ELSA (Ethical, Legal, and Social Aspects) Lab approach has emerged as an important framework for developing responsible and trustworthy AI systems (Ryan and Blok 2023; Veenstra, Zoonen, and Helberger 2021). A prominent example is the large-scale funding initiative by the National Growth Fund programme AiNed, implemented by the Dutch AI Coalition (NL AIC), to create a wide range of AI ELSA Labs in the Netherlands.[1] The purpose of ELSA Labs is to create a collaborative environment in which Quadruple Helix stakeholders (academia, civil society, government, and industry) can work together to develop responsible, human-centred AI that serves humans and the public good (NL AIC 2023). Driven by this initiative, 27 ELSA Labs have already been established across diverse domains, including defense, media, healthcare, agrifood, public safety, transportation, and energy, among others.[2] Many more ELSA Labs are expected in the near future, fuelled by continued investments from the Dutch government.[3] As we will explain later, while the concept of ELSA is not new, the renewed interest in the ELSA Lab approach signals a shift from what might be called ELSA 1.0 – focused primarily on retrospective, consequence-oriented assessment – to ELSA 2.0, which emphasizes early intervention, participatory design, and the integration of ELSA considerations throughout the AI ecosystem. This shift reflects a broader ambition to more deeply embed responsibility and human values into the development and application of AI.

The ELSA Lab approach may seem similar to existing RRI and Social Lab approaches, which all aim to align technological advancements with public interests through dialogue, participation, and early intervention (Ryan and Blok 2023). However, the ELSA Lab may have a unique edge in its integrating negotiation with practical, solution-oriented strategies. On the one hand, ELSA Labs prioritize dialogue by involving Quadruple Helix (QH) collaboration at the core of their work (Veenstra, Zoonen, and Helberger 2021). This collaboration actively involves not only social and technical scientists, but companies, policymakers, and the public – especially those affected and disadvantaged groups – in the co-design process (Ryan and Blok 2023). On the other hand, the ELSA Lab approach also has a strong solution-focused and pragmatic orientation, making it easier to integrate ethical values into concrete technological designs. Zwart and Nelis (2009) have shown that ELSA has a natural 'proximity' to scientific projects, allowing ELSA researchers to assess the immediate effects of technologies and determine how to address them, making it well-integrated into scientific research. The renewed ELSA policy initiative also explicitly indicates that ELSA Labs will adopt a 'pragmatic approach' to 'implement human centric AI successfully in practice'.[4] Therefore, this combination of negotiation and solution-oriented strategies offers a unique opportunity to address power dynamics and structural challenges, while also making these insights actionable within AI development. Notably, the ELSA Labs' institutional setup can further support the development and experimentation of integrating solutionist and negotiationist strategies, as it was designed as an ELSA-focused study with embedded AI cases, where ELSA is the primary focus, not just a secondary component of existing scientific projects (NAIC 2025; Veenstra, Zoonen, and Helberger 2021).

However, it remains unclear how the policy ideal of developing ELSA Labs as a governance framework can be both theoretically grounded and practically applied in real-world AI contexts. Theoretically, the ELSA Lab approach often refers to existing frameworks like Responsible Research & Innovation (RRI) and Value Sensitive Design (VSD) (Veenstra, Zoonen, and Helberger 2021). But it does not fully clarify, for instance, why we need to adopt an experimental approach to responsible AI, what makes it truly experimental, or how it builds on and differs from existing ethical and governance frameworks. On the practical side, while the ELSA lab approach emphasizes that it is experimental in nature, there is little empirical guidance on how to set up or implement these labs effectively in real-world settings. For example, ELSA Labs stress early stakeholder involvement in Research & Innovation (R&I) to prevent ethical, legal, and social risks and boost adoption (Long et al. 2020). However, it is still not clear how technological, economic, and ELSA aspects should be balanced in the design process, or how Quadruple Helix stakeholders (Carayannis and Campbell 2009) can be engaged effectively in the co-creation of AI technologies. This lack of theoretical and empirical knowledge largely restrains the strength of ELSA Lab as a new approach for designing responsible and trustworthy AI.

This paper is a first attempt to develop a science-based ELSA Lab approach informed by the rich literature in the field of RRI, and offers a practical work process for putting it into practice. While it does not claim to solve all the challenges identified in RRI and Post-ELSI research, such as inclusion, interdisciplinarity, and aligning stakeholder expectations (Balmer et al. 2015; Smolka and Fisher 2024), it provides a structured way to explore ELSA and experiment with different designs and interventions. To do this, we first consider existing responsible AI frameworks to set the stage for why an ELSA Lab approach is necessary to address ethical, legal and social aspects (Section 2). The paper then introduces the main theories that the ELSA Lab approach builds on: Responsible Innovation & Research (RRI), Social Labs (SL), and Quadruple Helix (QH) stakeholders collaboration (Section 3). Next, we analyze how these theories can inform the ELSA Lab approach and outline five key process requirements for developing this approach. These range from identifying multi-level ELSA aspects to its experimental nature and dedication to systemic change (Section 4). After this, a concrete work process to implement the ELSA Lab approach is provided, giving a detailed description of what a specific ELSA Lab process looks like (Section 5). The paper concludes with a critical reflection on the ELSA Lab approach, analyzing some possible challenges in putting it into practice (Section 6).

Towards an ELSA Lab approach for responsible AI

Why do we need an ELSA Lab approach

Many different ethical frameworks have been proposed to help develop responsible AI. For this paper, we are not going to list all those possible frameworks, as they may vary widely by use case, region, and scale of implementation (Corrêa et al. 2023; Jobin, Ienca, and Vayena 2019). But we can generally categorize them into three main types to give a brief overview of the responsible AI literature, which are the guidelines approach, the by-design approach, and the systemic approach.

The first type is the guidelines approach, a top-down method where governing bodies, NGOs, or companies publish concrete guidelines to make AI more responsible. In recent years, there has been a global push to establish ethical guidelines and principles to regulate AI risks (Corrêa et al. 2023; Jobin, Ienca, and Vayena 2019). Notable guidelines include the OECD AI Principles (OECD 2019, 2022) and the European Commission's ethics guidelines on Trustworthy AI (AI HLEG 2019). Big tech companies like Google, Microsoft, and DeepMind have also published their own ethical guidelines for governing AI (de Laat 2021).

Guidelines often come with checklists and evaluation forms to self-check if their technologies have any ethical and legal issues (AI HLEG 2020; Hagendorff 2020). These guidelines are useful for identifying relevant values and principles required to develop responsible AI, such as transparency, justice and fairness, non-maleficience, responsibility, and privacy (Jobin, Ienca, and Vayena 2019). However, many critics argue that this approach can be too abstract and general to be put into practice, potentially reducing ethics to just a checkbox exercise (Balayn et al. 2023; Kijewski, Ronchi, and Vayena 2025). This could misguide the AI development and ethical concerns may still be left unaddressed as part of the AI design itself.. As a result, ethical codes can become ‘toothless’ (Resseguier and Rodrigues 2020), a marketing tool or ‘ethics washing’ (Wagner 2018), or ‘dangerous distraction’ from genuine social issues where organizations project an image of responsibility without substantive action (Munn 2022; Resseguier and Rodrigues 2020).

The by-design approach is the second type, which focuses on integrating ethical values into AI development by translating abstract values into practical design steps (Dignum 2019; Sadek, Calvo, and Mougenot 2023). These values are often drawn from well-known guidelines like the HLEG’s trustworthy AI guidelines or the OECD’s guidelines (Brey and Dainow 2024). But sometimes it also involves engaging stakeholders to identify the values that matter most to them (Sadek, Calvo, and Mougenot 2023). This approach translated values into actionable design requirements and specific tasks that AI developers can easily implement. Unlike ethical guidelines that outline broad principles, the by-design approach goes further by embedding these ethical and legal considerations directly into the entire lifecycle of AI technology, from design to use (Buijsman, Klenk, and Van den Hoven 2025). For example, the Ethics by Design (EbD) approach ‘starts with high-level values for AI, which are then translated into design requirements for AI systems, and further into specific measures at key points in the design process’ (Brey and Dainow 2024). Other examples include Value Sensitive Design (Umbrello and Van de Poel 2021), Values in Design (Dignum 2019), and Design for Values (Buijsman, Klenk, and Van den Hoven 2025).

The by-design approach takes a practical, solution-focused approach, offering clear tasks for AI developers to follow, which makes it particularly useful for addressing ELSA aspects, especially for those with limited ethical or legal expertise. Its main strength is to directly embed ethical considerations into AI development, making it easier to address ELSA issues throughout the project. However, critics argue that this approach might overlook more complex issues that require broader discussions, social engagement, and negotiation (Ludwig et al. 2022). By focusing on specific AI products, by-design methods may fail to consider larger social and organizational issues, such as social impacts, value conflicts, and power dynamics (Blok 2023; Davis and Nathan 2015). This is often because the approach prioritizes values that are simpler to implement in design. For example, when aiming for ‘trustworthy AI,’ the focus tends to be on values like privacy, transparency, security, and accountability, which are relatively easy to formalize and turn into technical solutions (Hagendorff 2020). While the by-design approach is important, they might not sufficiently address the more complex ELSA challenges that cannot be solved by technical design alone.

The third type is the systemic approach, which views AI as part of a socio-technical system, shaped by power dynamics and social forces, rather than just a standalone technical artifact. This approach focuses on addressing both specific issues and the larger, more complex challenges in AI design. For example, Responsible Research & Innovation (RRI) (Stilgoe, Owen, and Macnaghten 2013) uses technological and social innovations to address ‘wicked problems,’ which are complex social challenges that are difficult to solve (Rittel and Webber 1973). The Social Lab is another approach developed to tackle such large, interconnected problems by bringing together diverse stakeholders to work on these issues (Marschalek et al. 2022; Timmermans et al. 2020). The systemic approach takes a negotiation-based view, recognizing that conflicting values are often at play, and ensures that all stakeholders – not just the most powerful ones – are heard. For instance, RRI emphasizes the importance of power, norms, and representation, focusing on political and social aspects of research and innovation (Pellé and Reber 2015). While RRI aims to integrate these perspectives into scientific projects, its application in practice is often limited (Conley 2020; Glerup, Davies, and Horst 2017). RRI generally focuses more on the broader governance of research, which may overlook specific ethical, legal, and social challenges that technical scientists and engineers encounter in research and development (Ryan and Blok 2023). The Table 1 summarizes the three types of responsible AI approaches:

This comparison of different types of approaches to responsible AI provides some useful insights. The by-design approach, with its solutionist strategies, can effectively embed values into technical design

Table 1. A generic typology of responsible AI approaches.

Types	Typical examples	Analysis
Guidelines approach	• Ethics guidelines for Trustworthy AI (AI HLEG 2019) • OECD AI Principles (OECD 2019)	This approach helps identify key values and principles needed to develop responsible AI, but it is abstract and generic in nature, making it difficult to implement ethics into specific contexts.
By-design approach	• Ethics by Design for AI (Brey and Dainow 2024) • Value sensitive AI (Sadek, Calvo, and Mougenot 2023; Umbrello and Van de Poel 2021)	This approach focuses on solutionist strategies to effectively translate values into concrete design requirements and tasks to directly address ELSA aspects. But it runs the risk of missing the complex ELSA issues (e.g. power dynamics) that cannot be solved by design of an artefact alone but require broader discussions, negotiations, and social engagement.
Systemic approach	• Responsible Research & Innovation (Stilgoe, Owen, and Macnaghten 2013) • Social Labs (Marschalek et al. 2022; Timmermans et al. 2020)	The approach follows a more negotiation path, which offers a way for diverse stakeholders to address complex ELSA aspects (e.g. power issues, structural issues) in AI through negotiation. However, it may struggle to translate and operationalize these negotiated nuances into concrete changes in AI (re)design.

practices, but it may fall short when it comes to addressing power dynamics and other structural ELSA issues (Wang and Blok 2025). In contrast, the systemic approach excels in democratizing technical innovation and fostering social collaboration, but it may struggle to translate these nuanced insights into design practices. In this paper, we propose the ELSA Lab approach to combine the strengths of both solutionist and negotiationist strategies for more effective responsible AI development. Before moving on, we would like to give a brief overview of ELSA.

The evolution of ELSA: from assessment to co-creation lab

The ELSA concept itself is not new. It originated as a way to address issues in the life sciences (mainly human genomics) (Fisher 2005; Forsberg 2015). In 1989, the United States introduced ELSI (ethical, legal, and social implications), while Europe adopted ELSA (ethical, legal, and social aspects) in 1994. The two terms generally refer to the same thing, with no clear methodological differences – only a variation in name and emphasis (Zwart and Nelis 2009). The term 'implications' tends to have a more negative connotation, often referring to harmful consequences that need to be addressed (Veenstra, Zoonen, and Helberger 2021). In contrast, 'aspects' is sometimes seen as more neutral, encompassing both positive and negative ethical, legal, and social dimensions of technology. In this paper, we primarily focus on ELSA due to its renewed interest, although we sometimes use the two terms interchangeably.

The original idea behind ELSA was mainly retrospective, which followed genomics research to analyze its consequences, identify major concerns, and develop policy responses (Meslin et al. 1997). The assumption was that scientific advancements could have significant social impacts, so ELSA research focused on recognizing and addressing potential ethical, legal, and social issues. This harm-prevention approach made it easier to integrate into scientific projects, but it also meant ethics was often treated as a way to manage risks rather than actively shaping technology for the common good (Ryan and Blok 2023). In von Schomberg's words, it 'see(s) the ethical aspect of new technologies as constraints and restrictions' that should be prevented, instead of 'positive contributions' people 'wish to obtain from research and innovation' (Von Schomberg 2013, 16). Another critique is that focusing on consequences often means acting too late; by the time ethical concerns are identified, the technology may already be too late to change (Myskja, Nydal, and Myhr 2014).

A more serious critique of ELSA research is that its focus on the 'consequences' of scientific developments can sidestep more pressing political and structural issues in research and innovation. In Langdon Winner's sharpening words, this type of approach 'conduct research about the ethical dimensions of emerging technology to gravitate towards the more comfortable, even trivial questions involved, avoiding issues that might become a focus of conflict' (Winner 2004). Rabinow and Bennett (2012) go even further, arguing that the ELSI program within the Human Genome Project was deliberately designed not to interfere with scientific research, preventing socio-ethical issues from being raised in the first place. As a result, this retrospective approach risks reinforcing the interests of technology proponents rather than fostering critical reflection. Instead of truly interrogating the social impact of new technologies, it may serve as a legitimizing tool of serving technology advocates' interest, ensuring that research proceeds with minimal disruption (Rip 2009; Zwart and Nelis 2009).

That is why some scholars have called for a 'post-ELSI' or 'post-ELSA' shift (Balmer et al. 2015). The 'post-' label reflects an effort to move beyond traditional ELSA frameworks. This shift stems from the assumption that traditional ELSA has often been 'non-collaborative' or only operated in parallel to scientific research (Myskja, Nydal, and Myhr 2014, 10). In response, the post-ELSA approach emphasizes the need for deeper integration between social and technical aspects (Balmer et al. 2015). However, this assumption may oversimplify ELSA's development. It is important to distinguish between how ELSA was framed by funding agencies (top-down) and how it was shaped by research communities (bottom-up).[5] From a top-down perspective, funding policies in 1990s and 2000s positioned ELSA as a parallel component alongside large scientific projects, especially in biotechnology (Forsberg 2015). In contrast, from a bottom-up perspective, interdisciplinary collaboration has always been a fundamental aspect of ELSA. As Zwart and Nelis (2009, 541) highlight, 'interdisciplinarity' and 'proximity' to scientific projects are two of ELSA's four key features, along with 'interactivity' and 'early anticipation'. Similarly, Myskja et al. (2014) show that in Norwegian ELSA projects and many other EU funding projects, ELSA scholars have actively engaged with scientific research rather than simply running alongside them. All these facts challenge the idea of a strict break between ELSA and post-ELSA. Despite this, a key limitation of ELSA has been its close proximity to scientific projects, which has sometimes come at the expense of critical reflexivity.

The renewed interest in ELSA, like the new funding of ELSA Labs in the Netherlands, signal a shift towards a more embedded but also more reflexive ELSA. We call this shift 'ELSA 2.0.' On the one hand, ELSA 2.0 is a natural extension of the original ELSA approach, building on decades of interdisciplinary research, methods, and theories that integrate social and technical aspects. These insights can provide valuable strategies for how to embed ELSA more practically. On the other hand, it represents a more reflexive mode of engagement, moving beyond just reviewing scientific developments to actively shaping them. This shift also changes the role of ELSA scholars. Instead of being passive observers, they now actively participate in the decision-making process of developing responsible technologies. They are empowered to be able to define what responsible technology should look like and to have a real, tangible influence on scientific development. For example, they will have the opportunity to actively design, experiment with, and test interventions that combine solution- and negotiation-based strategies to promote a more effective responsible AI.

However, the ELSA Lab approach is still in its early stages, and it remains unclear how this policy-driven funding initiative will be structured and implemented to effectively integrate both embedded and reflexive ELSA. For the rest of the paper, we will: (1) build the ELSA lab on existing literature on RRI, Social Labs, and Quadruple Helix collaboration; (2) identify distinct features of the ELSA Lab methodology; and (3) develop a working process to implement this methodology.

Theoretical grounding of the ELSA Lab approach

RRI: a framework of developing responsible technology

To address the criticism that the original ELSA (ELSA 1.0) is too precautious, it is helpful to build on the Responsible Research & Innovation (RRI) framework. RRI works by integrating social considerations and values shared by a wide range of stakeholders into the early stages of science and technology innovations (von Schomberg 2013). Therefore, RRI is largely proactive, as it aims to promote responsibility by aligning science and technology with socio-ethical needs (Dignum 2019). RRI outlines four well-known process requirements – anticipation, reflexivity, inclusion, and responsiveness (Stilgoe, Owen, and Macnaghten 2013) – the integration of which can not only enrich the ELSA concept but make ELSA more reflexive and proactive (Ryan and Blok 2023). Anticipation refers to identifying socio-ethical outcomes of technological development. Reflexivity demands that 'holding a mirror up to one's own activities, commitments and assumptions, being aware of the limits of knowledge, and being mindful that a particular framing of an issue may not be universally held' (Stilgoe, Owen, and Macnaghten 2013, 1571). Inclusion requires involving various affected stakeholders throughout the entire lifecycle of innovation to ensure that human values are integrated into the innovation process (Stilgoe, Owen, and Macnaghten 2013). Responsiveness means that innovators should take suitable actions based on new insights from the other three dimensions, such as redesigning the technology (Rose et al. 2021) or changing institutional structures (Fisher and Maricle 2015). These four process requirements of RRI can form a solid foundation for structuring ELSA Lab concept.

The ELSA Lab methodology does not just apply these RRI elements; it can also move debates in RRI forward. At first sight, ELSA and RRI seem quite similar in terms of aims and means, since both aim to align science and technology developments with human values, using ways of dialogue, participation, and early intervention. But looking closer, there are some subtle but important differences. In the paper, *Stop Re-inventing the Wheel: Or How ELSA and RRI can Align*, Ryan and Blok (2023) clarify the strengths and weaknesses of ELSA and RRI approaches, and based on these insights, they propose integrating them into an updated version in which ELSA and RRI are aligned. As they explained, both ELSA and RRI emphasize stakeholder engagement and interdisciplinary collaboration, encouraging scientists and the public to co-design research agendas. But ELSA excels in embedding these considerations directly into science and engineering projects (Zwart and Nelis 2009). RRI supporters also advocate for this, but in practice, it often does not happen as effectively (Glerup, Davies, and Horst 2017). ELSA originally developed from ethical assessments in genomics research, so it has a more directed and goal-oriented approach. It focuses on analyzing ELSA aspects in real-life research and innovation, which makes it naturally well-integrated into scientific and engineering projects (Zwart and Nelis 2009). RRI, on the other hand, is often more about governance and socio-political perspectives, emphasizing issues of power, norms, and representations (Pellé and Reber 2015; Smolka and Fisher 2024). This is partly because RRI discussions have been largely shaped by social and political scientists who view responsibility as something that develops through social and institutional processes (Blok & Lemmens 2015; Ryan and Blok 2023). In contrast, ELSA brings in a wider range of researchers – most notably technical scientists – and is more embedded in scientific and technological practices (Zwart, Landeweerd, and Van Rooij 2014).

By comparing RRI and ELSA here, we want to show how integrating ELSA and RRI can reinforce responsible innovation: although both concepts of RRI and ELSA have their merits, the ELSA Lab methodology builds on the strengths of both concepts (including social labs and quadruple helix collaboration), and provides opportunities for RRI in practice. ELSA's strength lies in its proximity to scientific projects and its ability to pragmatically address ethical and social issues in specific cases (Ryan and Blok 2023; Zwart and Nelis 2009). However, this solutionist strategy may risk reducing ELSA only to an add-on or a servant to large-scale science and technology projects – evaluating ethical aspects without necessarily questioning the broader innovation process (Forsberg 2015). In contrast, the RRI debate suggests a negotiation strategy, acknowledging that heterogeneous values are often conflicting (Ludwig et al. 2022). It looks at the power dynamics involved in research and innovation and tries to ensure that all stakeholders, not just the dominant ones, have a voice. Therefore, by combining these two approaches, we can develop a more effective way of fostering responsible innovation.

Instead of focusing just on the ELSA concept itself, this paper therefore developed a concrete methodology to put ELSA-RRI (or ELSA 2.0) into practice by blending ELSA's practical, solution-driven strategies with RRI's broader focus on power dynamics and stakeholder negotiation. Yet, ELSA 2.0 as a governance framework does not offer a concrete methodology for implementing responsible AI. We believe that the extensive literature on Social Labs can provide a clear empirical lab setup to operationalize the ELSA 2.0.

Social labs as empirical settings in addressing grand social challenges

The concept of Social Labs (SL), introduced by Zaid Hassan (2014) and further developed by Timmermans et al. (2020), has been gaining popularity in large EU projects like NewHoRRIzon and RIconfigure. These empirical labs offer a socially driven, experimental, and systemic approach to tackling complex social challenges (Binder et al., 2015; Laurent 2017; Marschalek et al. 2022). Social labs are 'labs' because they are similar to a natural science laboratory in that potential solutions are developed and experimented with on a small scale before being implemented more broadly. However, they are 'social' because 'they address complex social challenges that are not to be solved by a technological fix approach', and they often need a socially transformative approach to reshape how society is imagined, designed and operated.[6] This process is thus an open context, involving diverse experts and stakeholders in negotiating their stakes, rather than scientific objects in a closed lab (Kieboom 2014; Timmermans et al. 2020). In short, the core idea of a social lab is its flexible and open approach to experiment and test different interventions on a small scale as a prototype, which allows stakeholders to gather valuable insights and lessons that can be applied to address the broader social challenge at a systemic level (Timmermans et al. 2020). Timmermans *et al.* provided a framework with six features of the Social Lab methodology (see Table 2).

Table 2. Six features of social lab methodology (Timmermans et al. 2020).

Six features of social labs
1. Experimentation: Not only develop but test interventions and solutions to address social challenges.
2. Social Context: Experimentation takes place in real-life social contexts, not in closed laboratories.
3. Inclusion: Invite different stakeholders to actively participate in addressing complex social challenges.
4. Interdisciplinarity: Involve experts from a wide range of background and disciplines.
5. Systemic Transition: Solutions and lessons developed and tested in small scale experiments can serve as prototypes for systemic change.
6. Iteration and Agility: Provide an iterative and agile approach, allowing for ongoing update and refinement of prototypes and solutions over time.

This social lab approach can provide an empirical setup to operationalize ELSA into responsible AI. As explained earlier, developing responsible AI is a complex, social-technical challenge. Building on the Social Lab approach, ELSA Labs can focus on the complexity right from the start, turning them into continuous, explorative processes aimed at the systemic transformation of responsible AI. Like social labs, ELSA Labs can also learn to break down complex ethical, legal and social challenges into manageable parts, addressing them piece by piece while acknowledging the limitations of this reduction, testing and validating the solutions, and dedicated to a systemic shift. The experimental and agile nature of social labs can help ELSA Labs bridge theory and practice, as well as complexity and concreteness. However, before diving into the details of the ELSA Lab approach, we need to outline more specifics about stakeholder engagement in this laboratory setup, as this is the driving force and cornerstone of the ELSA Lab.

Quadruple Helix collaboration: from stakeholder types to value co-creation processes

Stakeholders often refer to any groups or individuals who can affect or is affected by the Research and Innovation (R&I) process (Bryson 2004; Popa, Blok, and Renate Wesselink 2020). Their role in the R&I process could involve having legitimate interests, providing valuable resources, addressing potential risks, or a mix of all three (Bryson 2004). Many quantitative and qualitative studies show that involving stakeholders is crucial (Nutt 2002), but there is still ongoing debate about how best to engage stakeholders meaningfully. One widely recognized approach is the Triple Helix (TH) model by Etzkowitz and Leydesdorff (1995, 2000), which highlights the interactions between academia, industry, and government to drive innovation and economic growth. In the EU context, many now believe that the Quadruple Helix (QH) model, which adds the public as a fourth helix, is better suited to today's social needs (Carayannis and Campbell 2009, 2011).[7] Following the work of Popa *et al*., we adopt a Quadruple Helix model that does not see the helixes as fixed types of stakeholder identity, but conceives them as 'processes of value co-creation in which participants, regardless of their title or identity, collaborate and compete for the production of different types of value.' (Popa, Blok, and Renate Wesselink 2020, 880). Table 3 illustrates the integration of this QH model with stakeholder groups. This ensures that the perspectives, values, and concerns of those affected by AI from the four helixes are sufficiently considered in the ELSA Lab approach.

Table 3. An integration of QH stakeholders and value-creation processes.

Helix	Behaviors/Practices	Created values	Outputs
Academia (e.g. researchers, universities)	Conducting research, publishing, knowledge development	• Knowledge creation • Provides theoretical and empirical research insights on ethical, legal, and social standards	Publications, research grants
Civil Society (e.g. citizens/ AI users, NGOs, media)	Participating in AI (re)design and feedback sessions, using AI technologies	• Shares practical, real-world perspectives and concerns • Ensure AI meets actual user needs and social values	Improved user satisfaction, usability
Government (e.g. policymakers)	Developing and enforcing regulations, setting policies	• Ensure regulatory compliance • Shape policies that promote responsible AI development and use	Laws, policies, guidelines
Industry (e.g. AI developers, companies)	Developing and deploying AI solutions, managing businesses	• Business value creation • Bring technical expertise and resources • Help identify opportunities for responsible/trustworthy AI • Help translate ethical and social values into practical AI solutions	Return on investment, market share, responsible AI

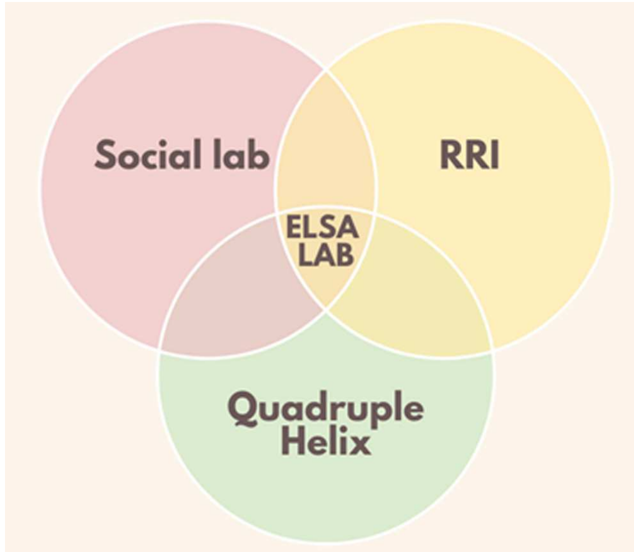

Figure 1. ELSA lab approach building up on social lab, RRI and QH.

Five features of ELSA Lab methodology

We propose to integrate the aforementioned theoretical fields – RRI, Social Lab, and the QH model – to develop and substantiate a concept of the ELSA Lab methodology.[8] RRI offers a general framework to develop responsible technologies. The Social Labs provides a concrete laboratory setting to address socio-ethical issues and implement responsible AI. The Processual QH model gives guidance whom to involve and how to effectively execute this lab process to achieve the intended values (see Figure 1). In this section, we articulate five requirements in building an ELSA Lab methodology.

(1) Systemic Diagnosis

ELSA Labs start with a systemic diagnosis of ethical, legal and social aspects in AI systems. This ELSA diagnosis is systemic in both a horizontal (broad) and vertical (deep) sense. Horizontally, the diagnosis covers not just ethical issues, as most ethical frameworks do in developing responsible AI, but important legal and social concerns as well, making a broader view of AI's impacts. Vertically, the diagnosis digs deep into multiple levels, identifying ELSA aspects not only at the individual artifact level (such as data privacy and transparency) but considering also structural social impacts (like sexism and colonialism) and ontological concerns, where for example AI-driven digitalization affects the human condition (e.g. deskilling) and disrupts traditional ways of living (Wang and Blok 2025). This systemic approach ensures that ELSA Labs address not just surface-level concerns but the deeper, often hidden, structural issues within the AI ecosystem, such as surveillance and deskilling, that have long-term impacts on individuals and society.

(2) QH Collaboration and Interdisciplinarity

A crucial part of the ELSA Lab process is its meaningful engagement with QH stakeholders in integrating complex ELSA aspects into AI (re)design. This QH engagement involves both stakeholder groups and the value co-creation process. On the one hand, it includes all four types of stakeholders – government, academia, industry, and the public – ensuring they are all represented and have an equal voice in the ELSA Lab process. On the other hand, it is not just about their titles or sectors, but about creating meaningful value that drives the redesign of AI processes in ethical acceptable and social desirable directions. This value co-creation process inherently requires interdisciplinary and transdisciplinary collaboration across various fields, bridging gaps among science, engineering, ethics, law, and social disciplines. For example, Ryan and Blok (2023) have shown that traditional RRI does not fully address the specific ELSA aspects engineers face, while one advantage of the ELSA approach is to put interdisciplinarity in a central position. While each partner plays different disciplinary roles in the lab, they learn from one another during the monitoring and evaluation of the (re)design process. This collaborative learning community shares knowledge and experiences, establishing common definitions and glossaries, and collectively designing AI responsibly.

(3) Experimentation and Testing

A key aspect of the ELSA Lab approach is to provide an experimental setup for QH stakeholders to experiment with and test alternative designs of AI. In this sense, the ELSA Lab serves a dual purpose: it provides a supportive environment to *experiment* with new insights or pilot actions in addressing complex ethical, legal, and social challenges. At the same time, it *evaluates* the actual efficacy of these experiments by monitoring the process to ensure that pilot actions are effectively integrated into the AI redesigns. For experimentation, the ELSA Lab enables the exploration of strategies to address complex ELSA challenges. These strategies are often new and untested, which means that they need to be further translated into concrete, actionable plans for real-world experimentation. This process of experimentation often embraces trial and error, allowing for creativity as well as failures and unexpected outcomes (Hassan 2014). As for testing, it is closely related to experimentation but can be seen as a distinct process in ELSA Labs. Once experimentation has been implemented, the testing phase rigorously evaluates the outcomes against specific criteria (Druckman 2022; McFadgen and Huitema 2018). So, unlike experimentation that focuses on creating and implementing new strategies, testing tends to be the process of evaluating the actual efficacy of those strategies to confirm they work as intended (Al-Ubaydli et al. 2021). After evaluating its overall efficacy, the decision can be made on whether to scale the strategies in other AI cases and broader institutional contexts (List 2024). This is related to the systemic transition we will discuss later.

(4) Iteration and Agility

An important feature of the ELSA Lab is its iteration and agility. Iteration in the ELSA Lab means that the process of identifying and experimenting with strategies or pilot actions is continuous and cyclical. Unlike a mere repetition, this iteration is a deliberate approach to improving AI's ELSA performance through continuous cycles of feedback, evaluation, and adjustment. Each cycle is informed by real-world experiences and insights of stakeholders, creating a continuous feedback loop. With this cycle, the ELSA Lab creates an actively learning structure where the Lab learns from what works and what does not, and this new knowledge is used to adjust the AI redesigns. This continuous adjustment builds a incremental improvement – each iteration builds on the previous one, gradually enhancing the AI's ethical alignment, legal compliance, and social acceptability. Agility is closely related to the idea of iteration, and it is widely used in the domain of software development (Fowler & Highsmith 2001). We use iteration and agility differently in this paper: iteration is more about the *process* of cycling and incremental improvement, while agility focuses on its *ability* to adjust and stay resilient to changes (Amugongo et al. 2025). So, in the context of the ELSA Lab, agility refers to its capacity to swiftly respond to new challenges, opportunities, and insights as they arise during the AI redesign process. This agility is crucial for dealing with the fast-paced and often unpredictable nature of AI innovation, making it resilient to, for instance, new regulatory requirements, shifts in public opinion, or technological advancements, etc.

(5) Systemic Transition

The last key feature of the ELSA Lab is its commitment to driving systemic transition. While the ELSA Lab works on individual AI cases, experimenting with strategies or pilot actions on a small scale, it often keeps the scalability and systemic transformation in mind. The Lab will evaluate the effectiveness of these solutions to determine whether they should be expanded to broader contexts and contribute to systemic change (List 2024). For example, the ELSA Lab for sustainable food systems at Wageningen University in the Netherlands is built not just for testing a new AI application for ELSA compliance, which is often located at the micro level, but dedicated to a broader shift of sustainable food eco-systems. The Lab, for instance, has defined what social goals AI should serve and conducts strict ELSA assessments to evaluate to what extend the AI systems really contribute to the social transition towards sustainable production and consumption.[9] This scalability towards systemic transition in the ELSA Lab can be achieved, for

Table 4. Five process requirements to develop ELSA labs.

Descriptions of five process requirements
1. **Systemic Diagnosis:** Systemically identify potential ethical, legal, and social aspects of AI at multi-level throughout its early development stages and the entire lifecycle.
2. **QH Collaboration and Interdisciplinarity:** Get QH stakeholders involved in co-creating pilot strategies to address relevant ELSA aspects of AI. Break down disciplinary boundaries to help researchers from different fields learn from each other and work together during the co-creation process.
3. **Experimentation and Testing:** Experiment with pilot strategies in real-world contexts and test their practical efficacy.
4. **Iteration and Agility:** Incrementally improve AI's ELSA performance through an iterative process, while using the ELSA Lab's agility to adapt to changing situations.
5. **Systemic Transition:** Provide tested pilot strategies for individual AI cases – along with lessons learned and best practices – as prototypes to support a systemic transformation of the AI ecosystem.

example, by documenting best practices and successful strategies from individual cases and sharing them widely to drive systemic transitions across the AI landscape (Long et al. 2020). By providing these best practices, the ELSA Lab can, for instance, influence policy-making and regulatory standards (McFadgen and Huitema 2018). By demonstrating what works in addressing ELSA challenges, the lab provides evidence-based recommendations that policymakers can use to shape new regulations or update existing ones. As these regulations are implemented, they enforce ELSA-integrated AI practices on a larger scale and contribute to a systemic transition towards responsible AI.

All in all, this section outlines the theory-based ELSA Lab approach by defining five key process requirements (Table 4). These requirements are *processual* because they are not isolated but interconnected in an iterative process. In the following section, we will depict how an ELSA Lab can actually work effectively by incorporating these five requirements.

The outline of ELSA Lab process

This section outlines an effective work process for implementing the ELSA Lab approach in the context of AI. The main goal of this ELSA Lab process is to develop responsible and trustworthy AI by integrating ELSA aspects in the (re)design of AI systems. The process enables R&D teams to enter the ELSA Lab, where the ethical, legal, and social aspects (ELSA) of specific AI cases are systematically identified and addressed with QH stakeholders at an early stage, when potential issues in AI systems can be proactively managed. By the time the AI cases leave the Lab, they are expected to have made fundamental progress regarding ELSA aspects and systemically integrated ELSA into their AI development. Specifically, the lab process follows some key steps. It starts with the identification of relevant ELSA challenges and opportunities, then moves to co-develop a series of redesign actions to address those challenges through QH stakeholder collaboration. After that, these ELSA redesign actions are tested in real-world settings and evaluated with their actual efficacy. Feedback and insights from these tests help refine the responsible redesign actions, fostering continuous learning and incremental improvement of AI's ELSA performance (see Figure 2). For the rest of this section, we will go into more detail about each step in the ELSA Lab process.

ELSA Lab's entry point: the ELSA scan

To enter the ELSA Lab, the first step of the process is an initial evaluation of the potential ELSA aspects for a specific AI technology in its context, we call this a use case. This 'ELSA Scan' acts as a preliminary diagnostic in the ELSA Lab process, because it mainly focuses on AI designers rather than a broader range of QH stakeholders. For example, in the Smart Food Tray case from our ELSA Lab – an AI-based system designed to monitor food intake and health conditions in elderly individuals and provide personalized nutrition recommendations – we first distributed a survey to the AI developers to collect information on the system's technical readiness and current stage of implementation. Following this, we conducted a 1.5-hour interview with the people who develop this Smart Tray on a separate day, using a set of ten ELSA questions as outlined by Van Hilten et al. (2024). Through this ELSA Scan, we aim to identify (a) the significance of the ELSA aspects involved (to legitimate whether a full ELSA Lab evaluation is needed), (b) the perception of these ELSA aspects related to the AI technology (like what actions AI developers have already taken to consider ELSA aspects), (c) the readiness to engage in an ELSA Lab process in which AI developers have to be

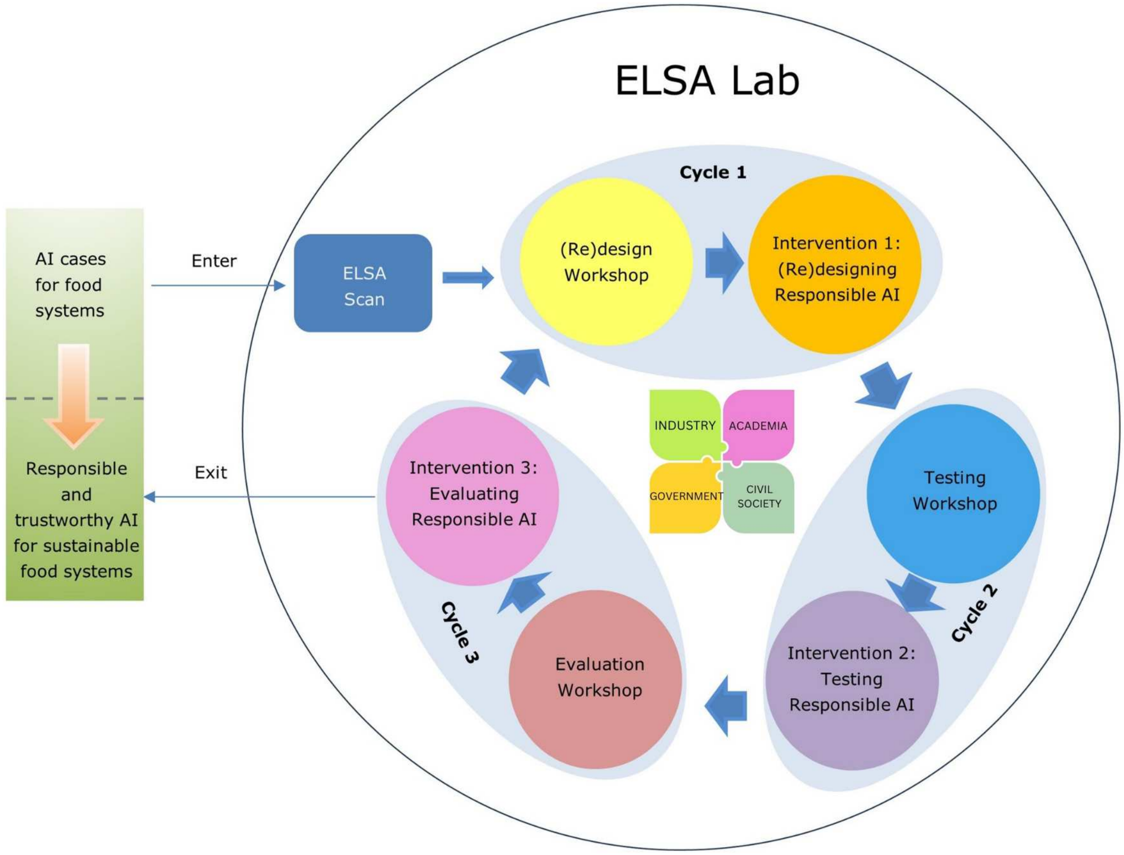

Figure 2. An overview of the ELSA lab process.

open for such a redesign with QH stakeholders, (d) if funding is available for the assessments in the ELSA Lab and/or the interventions that will be proposed, and (f) if AI developers have the basic governance structure that facilitates responsible redesign (checking for example if there are established protocols or guidelines for addressing ELSA aspects). If these criteria are met, the AI case can proceed to the next step the ELSA Lab process for a systemic identification and experimentation with ELSA Aspects.

The first ELSA-Lab cycle: (re)designing responsible AI

After the ELSA Scan, cases of AI-based systems will engage in a process of three ELSA-lab experiential learning cycles of redesigning, testing, and evaluating responsible AI. This learning cycle is experiential in the sense that it is a process of learning through experience or reflection on doing (Moon 2004). Each learning cycle includes two processes: (1) A QH stakeholders workshop (0.5 d); (2) An intervention phase where one or more ELSA pilot actions that engineers and QH stakeholders will experiment with (3–12 months). Pilot actions often refer to strategies that help change, adjust, or guide the AI development process to address ELSA aspects. For example, if AI technology is diagnosed to have the risk of violating user privacy, a pilot action could be redesigning data collection process to be more transparent and user-consent format. Its output might involve creating new privacy protocols or clearer user consent interfaces, etc. The ELSA Lab can create a wide range of pilot actions, such as providing legal advices, drafting ELSA guidelines, creating alternative designs, offering policy suggestions, developing scenarios, etc.

ELSA (re)design workshop

The first Lab cycle, at the core of the ELSA Lab, is the process of creating pilot actions to address ELSA aspects in AI. This cycle includes an ELSA (Re)design Workshop and an intervention phase of developing those pilot actions. The aim of the workshop is twofold: (1) to engage QH stakeholders to take a deep, systematic diagnosis of the ELSA aspects of AI, and (2) to work towards pilot actions in addressing them.

Unlike the ELSA Scan, this workshop brings together not only AI developers but QH stakeholders to identify the ELSA aspects they find most relevant. With a broader range of perspectives, the workshop is supposed to surface more complex issues – ones that may not have been recognized by AI developers alone during the ELSA Scan. In the Smart Food Tray case, for example, we invited nine QH stakeholders to a workshop (3hrs) held on September 27, 2024. Participants were divided into two groups, and facilitators guided them using the ELSA Impact Tool (Wang et al. 2025) to help each group identify a wide range of ELSA concerns. After generating a list of issues, each group was asked to select the top three they believed were most important to address. One group, for instance, prioritized the following: (1) camera surveillance; (2) liability (who is accountable if the Smart Tray malfunctions and causes harm); (3) distrust (elderly users who perceive the tray as impersonal, linked to government's budget cuts, or a tool of control rather than care). Following this prioritization, the workshop moved into planning concrete pilot actions and assigning responsibilities for the AI design teams to begin addressing these issues.

A pilot action in this context is not a detailed actionable plan but an initial action-oriented suggestion (or action point) that has not yet been fully developed or tested to address a specific ELSA aspect. For example, to address surveillance concern in the Smart Tray case, a pilot action could be: 'Involving ELSA experts and AI developers together to explore how to reduce both the actual surveillance risks and users' feelings of uneasiness'. This action point is short and is not a detailed implementation plan. The first workshop will end with three to six such pilot actions which the design team and ELSA experts will further work on.[10] Although no need to detail actions in the first workshop, it is important to specify 'action owners' who are volunteered to involve which action points. This way, participants can leave the first workshop with some general tasks in mind as well as create more commitment to the tasks in the coming months.

Intervention phase I: redesigning responsible AI

After the workshop, the process moves into the first phase of intervention: actually redesigning responsible AI. This phase will aim to not just create detailed redesign plans but seeks to implement the concrete plans in actual AI redesigns. During this phase – for example, in the Smart Tray case – the ELSA Lab manager engages with the action owners responsible for the camera surveillance issue, asking them to draft concrete plans outlining how and when the technology will be redesigned to address surveillance concerns and implement the proposed pilot actions.[11] Whereas it will be action owners (e.g. AI designers) who will take the lead in executing the detailed plan and actually integrating redesign actions into AI, the ELSA Lab and other QH stakeholders will provide the necessary support to help them successfully complete that plan..[12] For example, to address concerns around surveillance concern in that Smart Tray case, we organized a group meeting with AI developers and ELSA experts (N = 5) to explore the practical, economic, and technical feasibility of redesigning the camera setup. One proposed solution was to embed the camera discreetly within the tray, rather than placing it directly in front of users. Another idea was to add a button that allows users to turn off the camera if they feel uncomfortable.[13] The ELSA Lab manager will report on the process leading up to the first workshop and intervention, and document best practices from this first learning cycle. Generally, the first intervention phase will conclude with a list of AI redesign practices.

The second ELSA Lab cycle: testing responsible AI

After 3–6 months, the second cycle of ELSA-Lab experiential learning starts. In this cycle, pilot actions for AI redesign actions will be tested in real-world scenarios to see if they also work effectively in a more opened, institutional environment.[14] For example, in the Smart Tray case, the technology was initially tested in a restaurant setting to evaluate the effectiveness of its redesigned camera integration and to assess whether the opt-out option functions as intended. It is now planned to be tested in home care settings, where users are more vulnerable and the context is more complex. The testing in real-world situations is important for the ELSA Lab process because it tries to demonstrate how much the redesign actions, initially explored as 'pilot' interventions, can potentially make meaningful, real-world changes beyond Lab environment.[15] This second learning cycle includes an ELSA experimentation workshop and an intervention phase for real-world testing.

ELSA testing workshop

The aim of the workshop is to create a preliminary plan to test these redesign actions in real-world contexts. In the Smart Tray case, for example, QH stakeholders come back together in person in the workshop to assess how the earlier redesign actions (focused on the top three issues from each group, so six in total) have been carried out. They also look at what is worked well, what challenges came up (like technical issues or time pressure), and what new opportunities might help improve the design further. From there, they will figure out ways how to overcome those barriers and fine-tune the pilot redesign actions accordingly. After that, QH stakeholders will shift their focus to drafting a plan for real-world testing of these fine-tuned redesign actions in institutions or field settings. While not detailed, this preliminary plan should clarify which institutions will host the testing sites, who will volunteer to be involved in the testing, and what goals will be achieved. The workshop ends with one to three a draft plan to test those redesign actions in real-world institutions.

Intervention phase II: real-world testing

The second intervention phase focuses on a real-world testing.[16] The first thing for action owners is to detail the plan of real-world testing, specifying when and how the testing process will be executed in host institutions, such as the timeframe, tasks, roles assignment, setting milestones, and determining next steps. With this plan, it will be action owners and host institutions who are supposed to actually lead the real-world testing plan, but some other QH stakeholders will stay involved in monitoring the testing and giving feedback to improve the testing as it progresses. In the Smart Tray example, the trays were installed in the restaurant as planned, and redesigned features – such as the camera placement and opt-out option – were actively tested. It is important to track how these features perform in real time, either by engaging users during the testing phase or by collecting their feedback afterward through surveys or other tools. The second intervention ends with one to three redesign actions that have been well tested in real-world institutions.

The third ELSA Lab cycle: evaluating responsible AI

After 3–6 months, the third leaning cycle begins, which is mainly about evaluating how much the AI's ELSA aspects have improved and deciding if the case is ready to leave the Lab. This evaluation is a more systemic assessment of the AI case in the whole ELSA Lab process, which is different from the specific evaluations made in the last two cycles. In previous two cycles, evaluations are only used to assess if the interventions were working as intended to adjust and improve pilot redesign actions. These evaluations often look at criteria like stakeholder acceptance, technical feasibility, economic viability, and how these different things are properly balanced in actual redesign process.[17] However, for the third cycle, while its systemic evaluation is built up on the evaluations in these two cycles, it also needs to look at the bigger picture to assess responsible AI's overall success in achieving broader social goals. For instance, in the ELSA Lab for Sustainable Food Systems, this third cycle will assess to what extend the redesigned cases (e.g. Smart Trays) are really contributing to this sustainability goal. Based on this more systemic evaluations, the third learning cycle will end deciding if the AI case is ready to exit the lab.

ELSA evaluation workshop

The Evaluation Workshop (1/2 d) includes three main tasks. First, it engages QH stakeholders in evaluating how much efficacy revised redesign actions have been tested in real-world contexts. Second, it systemically assesses the entire ELSA Lab process to check how much the ELSA aspects have been improved in the AI case and if there are still any relevant ones that need attention. For example, in the Smart Tray case, the evaluation focuses on whether the six most relevant ELSA issues have been adequately addressed. Some issues, like legal concerns (e.g. liability), may be relatively straightforward to handle by involving professional legal experts from the Lab. However, challenges such as surveillance concerns and user distrust are more complex. To tackle these, QH stakeholders need to collaboratively define appropriate matrix of criteria and indicators for the systemic ELSA evaluation during the workshop. Even if issues like surveillance and distrust are difficult to fully resolve, it is important to establish a reasonable threshold – for instance, demonstrating that the redesign significantly mitigates surveillance risks or measurably improves trust among elderly users. Afterwards, the workshop will move on to the third task of defining the exit point for the ELSA Lab process. The exit point is where it is decided whether the AI case should stay in the ELSA

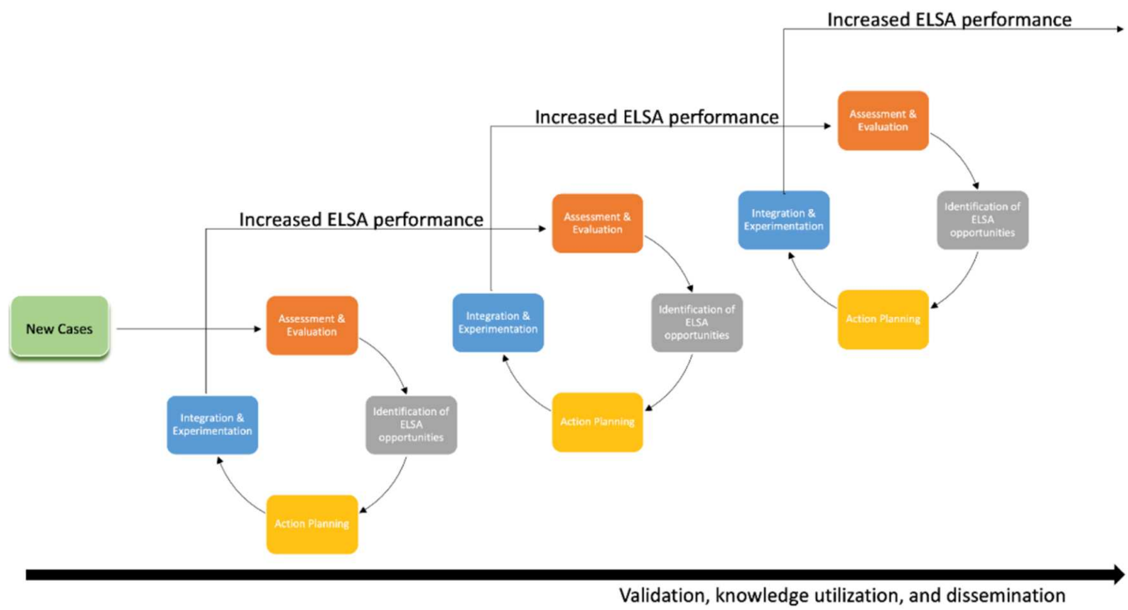

Figure 3. Incremental process of ELSA performance.

Lab or move out as a responsible AI system. This workshop will end up with a set of criteria for systemic evaluation and a clear definition of the exit point that all stakeholders agree on and will further work on.

Intervention phase III: performing a systemic evaluation

In the third intervention phase, systemic evaluations will be performed by applying the evaluation matrix defined during the third workshop to make a detailed assessment of whether the case (e.g. the Smart Tray) has made sufficient progress and whether there are still relevant ELSA aspects that engineers are willing and able to address. The ELSA Lab manager will then put together a final evaluation report and share this with all QH stakeholders. Based on the reports, the ELSA Lab has an advisory role in suggesting whether the AI case can leave the Lab. However, in the end it is the AI designers who have to decide whether to exit the lab or continue into a new lab cycle.

Hence, this entire ELSA Lab process is continuous and cycling, which contributes to the incremental improvements of AI's ELSA performance (see Figure 3 below). While the focus is on individual AI cases each time, the ELSA Lab process is often dedicated to wider institutional change. In the Smart Tray case, for instance, we encountered complex social issues such as distrust – not merely about privacy, transparency, or fairness, but a deeper, more systemic distrust rooted in past broken promises by government institutions, fear of societal and technological change, and the disruption of traditions among elderly populations. Addressing these concerns requires more than just redesigning technology; it calls for a shift in our underlying assumptions – questioning whether AI is always the answer, moving beyond a focus on efficiency, and placing greater value on emotional support and human connection. It also highlights the need for improvements in the broader care system itself. Documenting best practices for tackling such complex ELSA issues can offer valuable insights beyond the case at hand – for example, by informing evidence-based recommendations for policymakers to strengthen legal and regulatory frameworks. This kind of collective learning and knowledge-sharing can contribute to a systemic shift toward a more responsible and trustworthy AI ecosystem.

Discussion and conclusion

This paper develops an ELSA Lab approach to (re)designing, testing, and evaluating responsible AI. The ELSA Lab method consists in a systemic and experimental approach to integrating complex ethical, legal, and social aspects (ELSA) into AI development. We first explained why such an approach is needed (Section 2) and connected it to existing theories like RRI, Social Labs, and QH Collaboration (Section 3).

Then, we outlined the key features of this approach (Section 4) and laid out a step-by-step process to implement it (Section 5). Our aim was to develop a theory-informed ELSA Lab methodology, grounded in scientific theories and insights. By drawing from existing research in RRI, Social Labs, and QH stakeholder collaboration, we ensure that the ELSA Lab methodology is both theoretically sound and expect that it is practically viable to develop responsible AI.

We argue that the ELSA Lab approach may offer a unique opportunity to combine the strengths of design-based approaches (like VSD) and negotiation-based approaches (like RRI) to more effectively develop responsible AI. This approach is sensitive to complex structural issues and power dynamics, while also being effective at translating ELSA insights into real, actionable changes in technology. On the one hand, the ELSA Labs build on QH collaboration, involving scientists, companies, policymakers, and those directly affected or disadvantaged by technological developments. This inclusive approach allows ELSA Labs, like RRI, to consider not only artefact-level challenges but wider societal and institutional issues, including persistent power imbalances (Ryan and Blok 2023). On the other hand, similar to VSD, ELSA Labs have a strong solution-oriented and pragmatic focus, making it easier to integrate ELSA consideration into technological systems. ELSA approaches have historically maintained close 'proximity' to scientific practice, more actively involving engineers and technical experts throughout the development process (Zwart and Nelis 2009). Compared to RRI, as Ryan and Blok (2023) point out, ELSA Labs place a stronger emphasis on design practices and the acceptability of specific technological designs, rather than on broader science–society deliberation and democratic legitimation. The ELSA Lab approach fosters embedded, co-design processes, in which technical experts and ELSA researchers work collaboratively not only to identify and address complex ELSA issues (including power dynamics) but to translate these insights into concrete design tasks.

Although the potential advantages of the ELSA Lab framework became clear in this article, further empirical research is now needed to explore how the process unfolds in practice. Interdisciplinary and multi-stakeholder engagement is known to be difficult. Prior research in RRI, Post-ELSI, and STS has highlighted key challenges (Balmer et al., 2016; De Hoop, Pols, and Romijn 2016; Kuzma and Roberts 2018; Smolka and Böschen 2023). These challenges include the time-consuming nature of stakeholder engagement, the complexity of aligning interests across disciplines, and difficulties in involving citizens – especially those from marginalized groups (Lysen and Wyatt 2024). There is also the risk that participation becomes extractive (Sloane 2024) or merely symbolic, reduced to 'window dressing' rather than meaningful involvement (Arnstein, 2019; Gilman 2022).

Given these challenges, we need more empirical research on the dynamics of interdisciplinary collaboration within lab settings to validate the claimed advantages. Importantly, the ELSA Lab itself serves not only as a framework for responsible AI development but also as a research site for investigating these dynamics and conducting case studies. It enables experimentation with new interventions aimed at enhancing inclusion, interdisciplinarity, and sensitivity to power dynamics. For example, the multi-level ELSA Impact Tool developed in the Lab (Wang *et al.* 2025) helps reveal structural and often hidden challenges, encouraging designers and stakeholders to engage with responsibility at multiple levels of the AI system. It also helps bridge communication gaps by connecting technical scientists' focus on artefact-level concerns with citizens' interest in broader, structural ELSA issues. The ELSA Lab further strengthens QH collaboration through iterative cycles of case development, testing, mutual learning, and reflection. This ongoing process deepens shared understanding of challenges and opportunities, helping to align efforts across diverse stakeholders.

In summary, our paper presents a methodological approach to ELSA research for developing responsible AI. It is not meant as a rigid, one-size-fits-all approach but as a flexible model that researchers can adapt and refine.[18] Ideally, it could serve as a standardized way of studying responsible AI innovation, making it easier to compare findings across different projects – whether within ELSA Labs in the Netherlands or even broader ELSA and ELSI initiatives.

Notes

1. In 2021, the programme AiNed received €204.5 million from the National Growth Fund to advance the responsible AI development and application in Dutch companies and governments. See more information here: https://ained.nl/en/. As of Jan. 1, 2025, the Netherlands AI Coalition and AiNed have merged to form the AI Coalition 4 NL (AIC4NL): https://aic4nl.nl/.

2. An up-to-date list of established ELSA Labs in the Netherlands can be found here: https://nlaic.com/en/bouwsteen/human-centric-ai/elsa-concept/ .
3. For instance, in 2025, four new ELSA Labs have been funded through the AiNed programme: https://www.nwo.nl/en/news/four-research-projects-launched-for-ngf-call-ained-elsa-labs.
4. See: https://nlaic.com/en/bouwsteen/human-centric-ai/elsa-concept/
5. Thanks to the reviewer for pointing this out.
6. Quoted from the NewHoRRIzon project description: https://newhorrizon.eu/social-labs/.
7. Carayannis and Campbell (2010) added a fifth helix – the natural environment – to the innovation framework, creating the Quintuple Helix. This model highlights the ecological perspectives in 21st-century innovation, but it did not gain as much empirical traction as the Quadruple Helix, likely due to challenges in integrating the environmental aspect with the other elements (König et al. 2020, 8). This paper does not focus on the Quintuple Helix since we believe it has similar theoretical basis with the Quadruple Helix (Carayannis and Campbell 2021).
8. For example, understanding 'inclusion' as QH inherently involves interdisciplinary collaboration, so we merge these aspects to highlight stakeholder engagement. RRI and ELSA's concepts of reflexivity and the balance between solutionism and negotiation are fundamental features of systemic anticipation and other processes, such as stakeholder engagement, experimentation, and systemic transition. Responsiveness, or the change/redesign aspect, is specified in terms of experimentation and iterative agility, inspired by social labs. Experimentation permeates the entire ELSA lab process, with the lab setting dedicated to testing interventions and solutions within real social contexts. Lastly, we emphasize systemic change by specifying that developing responsible AI requires a systemic transition.
9. See the website for the ELSA Lab at Wageningen University & Research: https://ai4sfs.org/
10. For example, if data privacy is one of the top three ELSA aspects identified earlier, one action point is to figure out who will take the lead in addressing this concern. It is the same way for some more complicated issues, like the disruption to traditional lives. These action points should be discussed and agreed upon with the QH stakeholders during the workshop, and stakeholders can volunteer as action owners who will commit to leading efforts in executing the action point after the workshop.
11. The plan involves concrete details and a roadmap for implementation, such as outlining the time structure, identifying target groups, specifying redesign requirements and tasks, assigning roles, setting milestones, and determining next steps.
12. For example, ELSA experts will develop dedicated tools and services – like legal advices, ethical matrix, policy sandbox – that can help enable alternative designs. This exchange process can be supported by digital platforms.
13. The discussion also covered alternative design options and the content of caregiver training materials – specifically, what guidance should be included to support respectful and transparent communication with users.
14. Admittedly, the first cycle also tests pilot actions, but these tests are often more internal ones, performed in a relatively more controlled lab settings with *selected* QH stakeholders. In the second cycle, however, the pilot actions will be tested in a more open, real-life institutions or fields like in hospitals, companies, or farmlands. These more realistic contexts may involve unexpected AI users and interactions within complex economic and political circumstances.
15. In this sense, if the first learning cycle is more about experimenting with the co-design part of responsible AI, the second learning cycle would focus more on the testing whether the co-design is effective in real-world situations.
16. Before the real-world testing, the action owners should have identified the details about how to improve their previous pilot redesign actions (e.g., addressing time pressure and technical feasibility) based on the revised pilot actions developed in the workshop.
17. This type of evaluations for redesign actions can be seen throughout the ELSA Lab process, in workshops and during intervention phases, since the ELSA lab approach is featured as iterative and agile, which creates incremental improvement based on continual feedback and evaluation in each learning cycle.
18. The ELSA Lab approach is evolving and open to change. We plan to run 5 to 6 AI cases through the lab process, carefully documenting each step. Based on these cases, we will reflect on the effectiveness, strengths, and limitations of the approach and adjust the model if needed. Since the ELSA Lab is module-based, it also allows for flexibility – certain modules can be tested, adapted, or even replaced with other experimental methods if they prove more effective for responsible AI development.

Acknowledgments

We are grateful for the support from the ELSA Lab team and the Philosophy Group at Wageningen University. We clarify that Grammarly and, in limited cases, ChatGPT (3.5) were used to improve grammar after the full draft was written. The author(s) take(s) full responsibility for the content.

Author contributions

CRediT: **Hao Wang:** Conceptualization, Methodology, Visualization, Writing – original draft, Writing – review & editing; **Vincent Blok:** Conceptualization, Funding acquisition, Methodology, Supervision, Validation, Writing – review & editing; **Mireille van Hilten:** Funding acquisition, Methodology, Writing – review & editing.

Disclosure statement

No potential conflict of interest was reported by the author(s).

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 Nederlandse Organisatie voor Wetenschappelijk Onderzoek [grant number Nwa.1332.20.002].

References

AI HLEG. 2019. "European Commission's Ethics Guidelines for Trustworthy Artificial Intelligence (Issue May)." https://ec.europa.eu/futurium/en/ai-alliance-consultation/guidelines/1.

AI HLEG. 2020. *Assessment List for Trustworthy AI (ALTAI)*. https://digital-strategy.ec.europa.eu/en/library/assessment-list-trustworthy-artificial-intelligence-altai-self-assessment

Al-Ubaydli, O., M. S. Lee, J. A. List, C. L. Mackevicius, and D. Suskind. 2021. "How Can Experiments Play a Greater Role in Public Policy? Twelve Proposals from an Economic Model of Scaling." *Behavioural Public Policy* 5 (1): 2–49. doi:10.1017/bpp.2020.17.

Amugongo, L. M., A. Kriebitz, A. Boch, and C. Lütge. 2025. "Operationalising AI Ethics through the Agile Software Development Lifecycle: A Case Study of AI-Enabled Mobile Health Applications." *I Ethics* 5: 227–244. doi:10.1007/s43681-023-00331-3

Arnstein, S. R. 2019. "A Ladder of Citizen Participation." *Journal of the American Planning Association* 85 (1): 24-34.

Balayn, A., M. Yurrita, J. Yang, and U. Gadiraju. 2023. ""☑ Fairness Toolkits, A Checkbox Culture?" On the Factors That Fragment Developer Practices in Handling Algorithmic Harms." In *Proceedings of the 2023 AAAI/ACM Conference on AI, Ethics, and Society*, 482–495. San Jose, CA: The AAAI Press.

Balmer, A. S., J. Calvert, C. Marris, S. Molyneux-Hodgson, E. Frow, M. Kearnes, Kate Bulpin, Pablo Schyfter, Adrian Mackenzie, and P. Martin. 2016. "Five Rules of Thumb for Post-eLSI Interdisciplinary Collaborations." *Journal of Responsible Innovation* 3 (1): 73–80. doi:10.1080/23299460.2016.1177867.

Balmer, A. S., C. Marris, J. Calvert, S. Molyneux-Hodgson, M. Kearnes, K. Bulpin, E. Frow, P. Schyfter, A. Mackenzie, and P. Martin. 2015. "Taking Roles in Interdisciplinary Collaborations: Reflections on Working in Post-eLSI Spaces in the UK Synthetic Biology Community." *Science and Technology Studies* 28 (3): 3–25. doi:10.23987/sts.55340.

Barben, D., E. Fisher, C. Selin, and D. H. Guston. 2008. "Anticipatory Governance of Nanotechnology: Foresight, Engagement, and Integration." In *The Handbook of Science and Technology Studies*, edited by E. J. Hackett, O. Amsterdamska, M. Lynch, and J. Wajcman, 979–1000. Cambridge, MA: The MIT Press.

Binder, T., E. Brandt, P. Ehn, and J. Halse. 2015. "Democratic Design Experiments: Between Parliament and Laboratory." *CoDesign* 11 (3–4): 152–165. doi:10.1080/15710882.2015.1081248.

Birhane A, E. Ruane, T. Laurent, M. S. Brown, J. Flowers, A. Ventresqu, and C. L. Dancy. 2022. "The Forgotten Margins of AI Ethics." In *2022 ACM Conference on Fairness, Accountability, and Transparency*, 948–958.

Blok, V. 2023. *Philosophy of Technology in the Digital Age: The Datafication of the World, the Homo Virtualis, and the Capacity of Technological Innovations to set the World Free*. Wageningen: Wageningen University & Research.

Blok, Vincent. 2024. "Ecological Hermeneutic Phenomenology: A Method to Explore the Ontic and Ontological Structures of Technologies in the World." *Phenomenology and the Philosophy of Technology*. doi:10.11647/OBP.0421.01.

Blok, V., and P. Lemmens. 2015. "The Emerging Concept of Responsible Innovation. Three Reasons Why it is Questionable and Calls for a Radical Transformation of the Concept of Innovation." In *Responsible Innovation 2: Concepts, Approaches, and Applications*, edited by B. Koops, R. Oosterlaken, T. Swierstra, and J. van den Hoven, 19–35. Cham: Springer Internal Publishing

Bolte, L., and A. Van Wynsberghe. 2025. "Sustainable AI and the Third Wave of AI Ethics: A Structural Turn." *AI and Ethics* 5 (2): 1733–1742.

Bryson, J. M. 2004. "What to Do When Stakeholders Matter: Stakeholder Identification and Analysis Techniques." *Public Management Review* 6 (1): 21-53.

Brey, P., and B. Dainow. 2024. "Ethics by Design for Artificial Intelligence." *AI and Ethics* 4 (4): 1265–1277. https://doi.org/10.1007/s43681-023-00330-4.

Buijsman, S., M. Klenk, and J. Van den Hoven. 2025. "Ethics of AI: Toward a "Design for Values" Approach." In *The Cambridge Handbook of the Law, Ethics and Policy of Artificial Intelligence*, edited by N. Smuha, 59–78. Cambridge: Cambridge University Press.

Carayannis, E. G., and D. F. Campbell. 2009. "Mode 3'and'Quadruple Helix': Toward a 21st Century Fractal Innovation Ecosystem." *International Journal of Technology Management* 46 (3-4): 201–234.

Carayannis, E. G., and D. F. Campbell. 2010. "Triple Helix, Quadruple Helix and Quintuple Helix and How Do Knowledge, Innovation and the Environment Relate to Each Other?: A Proposed Framework for a Transdisciplinary Analysis of Sustainable Development and Social Ecology." *International Journal of Social Ecology and Sustainable Development (IJSESD)* 1 (1): 41–69. doi:10.4018/jsesd.2010010105.

Carayannis, E. G., and Campbell, D. F. 2021. "Democracy of Climate and Climate for Democracy: The Evolution of Quadruple and Quintuple Helix Innovation Systems." *Journal of the Knowledge Economy* 12 (4): 2050–2082.

Carayannis, E. G., and D. F. Campbell. 2011. "Open Innovation Diplomacy and a 21st Century Fractal Research, Education and Innovation (FREIE) Ecosystem: Building on the Quadruple and Quintuple Helix Innovation Concepts and the "Mode 3" Knowledge Production System." *Journal of the Knowledge Economy* 2 (3): 327–372. doi:10.1007/s13132-011-0058-3.

Conley, J. M., A. E. Prince, A. M. Davis, J. Cadigan, and G. Lazaro-Munoz. 2020. "Is Real-Time ELSI Realistic?" *AJOB Empirical Bioethics* 11 (2): 134–144. doi:10.1080/23294515.2020.1722289.

Corrêa, N. K., C. Galvão, J. W. Santos, C. Del Pino, E. P. Pinto, C. Barbosa, … N. de Oliveira. 2023. "Worldwide AI Ethics: A Review of 200 Guidelines and Recommendations for AI Governance." *Patterns* 4 (10): 1–14.

Davis, J., and L. P. Nathan. 2015. "Value Sensitive Design: Applications, Adaptations, and Critiques." In *Handbook of Ethics, Values, and Technological Design: Sources, Theory, Values and Application Domains*, 11–40. Dordrecht: Springer.

De Hoop, E., A. Pols, and H. Romijn. 2016. "Limits to Responsible Innovation." *Journal of Responsible Innovation* 3 (2): 110–134. doi:10.1080/23299460.2016.1231396.

de Laat, P. B. 2021. "Companies Committed to Responsible AI: From Principles Towards Implementation and Regulation?" *Philosophy and Technology* 34 (4): 1135–1193. https://doi.org/10.1007/s13347-021-00474-3.

Dignum, V. 2019. *Responsible Artificial Intelligence - How to Develop and Use AI in a Responsible Way. Artificial Intelligence: Foundations, Theory, and Algorithms*. Springer.

Druckman, J. N. 2022. *Experimental Thinking*. Cambridge University Press.

Etzkowitz, H., and L. Leydesdorff. 1995. "The Triple Helix--University-Industry-Government Relations: A laboratory for Knowledge Based Economic Development." *EASST Review* 14 (1): 14–19.

Etzkowitz, H., and L. Leydesdorff. 2000. "The Dynamics of Innovation: from National Systems and "Mode 2" to A Triple Helix of University-Industry-Government Relations." *Research Policy* 29 (2); 109–123.

Fisher, E. 2005. "Lessons Learned from the Ethical, Legal and Social Implications Program (ELSI): Planning Societal Implications Research for the National Nanotechnology Program." *Technology in Society* 27 (3): 321–328. doi:10.1016/j.techsoc.2005.04.006.

Fisher, E., and G. Maricle. 2015. "Higher-Level Responsiveness? Socio-Technical Integration within US and UK Nanotechnology Research Priority Setting." *Science and Public Policy* 42 (1): 72–85. doi:10.1093/scipol/scu017.

Fisher, E., and D. Schuurbiers. 2013. "Socio-Technical Integration Research: Collaborative Inquiry at the Midstream of Research and Development." In *Early Engagement and New Technologies: Opening up the Laboratory*, edited by N. Doorn, D. Schuurbiers, I. Van De Poel, and M. E. Gorman, 97–110. Dordrecht: Springer Netherlands.

Floridi, L. 2023. *The Ethics of Artificial Intelligence: Principles, Challenges, and Opportunities*. Oxford: Oxford University Press.

Forsberg, E. M. 2015. "ELSA and RRI–Editorial." *Life Sciences, Society and Policy* 11 (1): 1-3. doi:10.1186/s40504-014-0021-8.

Fowler, M., and J. Highsmith. 2001. "The Agile Manifesto." *Software Development* 9 (8): 28-35.

Gilman, M. E. 2022. "Beyond Window Dressing: Public Participation for Marginalized Communities in the Datafied Society." *Fordham Law Review* 91 (2): 503–555.

Glerup, C., S. R. Davies, and M. Horst. 2017. "'Nothing Really Responsible Goes on Here': Scientists' Experience and Practice of Responsibility." *Journal of Responsible Innovation* 4 (3): 319–336. doi:10.1080/23299460.2017.1378462.

Guston, D. H., and D. Sarewitz. 2002. "Real-Time Technology Assessment." *Technology in Society* 24 (1–2): 93–109. doi:10.1016/S0160-791X(01)00047-1.

Hagendorff, Thilo. 2020. "The Ethics of AI Ethics. An Evaluation of Guidelines." *Minds and Machines* 30 (3): 457–461. doi:10.1007/s11023-020-09526-7.

Hassan, Z. 2014. *The Social Labs Revolution: A New Approach to Solving Our Most Complex Challenges*. Oakland: Berrett-Koehler Publishers.

Jobin, A., M. Ienca, and E. Vayena. 2019. "The Global Landscape of AI Ethics Guidelines." *Nature Machine Intelligence* 1 (9): 389–399. https://doi.org/10.1038/s42256-019-0088-2.

Kieboom, M. 2014. *Lab Matters: Challenging the Practice of Social Innovation Laboratories*. Kennisland: Amsterdam.

Kijewski, S., E. Ronchi, and E. Vayena. 2024. "The Rise of Checkbox AI Ethics: A Review." *AI and Ethics* 5 (3): 1931–1940.

König, J., L. Suwala, and C. Delargy. 2020. "Helix Models of Innovation and Sustainable Development Goals." In *Industry, Innovation and Infrastructure*, edited by W. L. Filho, A. M. Azul, L. Brandli, A. L. Salvia, and T. Wall, 1–15. Cham: Springer.

Kuzma, J., and P. Roberts. 2018. "Cataloguing the Barriers Facing RRI in Innovation Pathways: A Response to the Dilemma of Societal Alignment." *Journal of Responsible Innovation* 5 (3): 338–346. doi:10.1080/23299460.2018.1511329.

Laurent, B. 2017. *Democratic Experiments: Problematizing Nanotechnology and Democracy in Europe and the United States*. Cambridge, MA: The MIT Press.

Leslie, D. 2019. *Understanding Artificial Intelligence Ethics and Safety: A Guide for the Responsible Design and Implementation of AI Systems in the Public Sector*. https://doi.org/10.5281/zenodo.3240529.
Long, T. B., V. Blok, S. Dorrestijn, and P. Macnaghten. 2020. "The Design and Testing of a Tool for Developing Responsible Innovation in Start-Up Enterprises." *Journal of Responsible Innovation* 7 (1): 45–75. doi:10.1080/23299460.2019.1608785.
Ludwig, D., V. Blok, M. Garnier, P. Macnaghten, and A. Pols. 2022. "What's Wrong with Global Challenges?" *Journal of Responsible Innovation* 9 (1): 6–27. doi:10.1080/23299460.2021.2000130.
Lysen, F., and S. Wyatt. 2024. "Refusing Participation: Hesitations about Designing Responsible Patient Engagement with Artificial Intelligence in Healthcare." *Journal of Responsible Innovation* 11 (1): 2300161. doi:10.1080/23299460.2023.2300161.
Marschalek, I., V. Blok, M. Bernstein, R. Braun, J. Cohen, M. Hofer, Lisa M. Seebacher, et al. 2022. "The Social Lab as a Method for Experimental Engagement in Participatory Research." *Journal of Responsible Innovation* 9 (3): 419–442. doi:10.1080/23299460.2022.2119003.
McFadgen, B., and D. Huitema. 2018. "Experimentation at the Interface of Science and Policy: A Multi-Case Analysis of How Policy Experiments Influence Political Decision-Makers." *Policy Sciences* 51 (2): 161–187. doi:10.1007/s11077-017-9276-2.
Menard, P., and G. J. Bott. 2025. "Artificial Intelligence Misuse and Concern for Information Privacy: New Construct Validation and Future Directions." *Information Systems Journal* 35 (1): 322-367. doi:10.1111/isj.12544322
Meslin Eric, M., E. J. Thomson, and J. T. Boyer. 1997. "Bioethics Inside the Beltway. The Ethical, Legal, and Social Implications Research Program at the National Human Genome Research Institute." *Kennedy Institute of Ethics Journal* 7 (3): 291–298.
Moon, Jennifer A. 2004. *A Handbook of Reflective and Experiential Learning: Theory and Practice*. London; New York: Routledge Falmer.
Munn, L. 2022. "The Uselessness of AI Ethics." *AI and Ethics* 3 (3): 869–877.
Myskja, B. K., R. Nydal, and A. I. Myhr. 2014. "We have Never been ELSI Researchers–There is No Need for A Post-ELSI Shift." *Life Sciences, Society and Policy* 10 (1): 1-17. http://www.lsspjournal.com/content/10/1/9
NLAIC. 2025. *ELSA Concept*. https://nlaic.com/en/bouwsteen/human-centric-ai/elsa-concept/.
NL AICoalition. 2023. *The Manifesto Human-Centric AI*. https://nlaic.com/wp-content/uploads/2024/02/Manifest_Mensgerichte_Artificiele_Intelligentie_augustus_2023.pdf.
Nutt, P. 2002. *Why Decisions Fail: Avoiding the Blunders and Traps That Lead to Debacles*. San Francisco, CA: Berrett-Koehler Publishers.
OECD. 2019. *Recommendation of the Council on Artificial Intelligence*. https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0449
OECD. 2022. *OECD Framework for the Classification of AI Systems*. https://doi.org/10.1787/cb6d9eca-en.
Pellé, S., and B. Reber. 2015. "Responsible Innovation in the Light of Moral Responsibility." *Journal on Chain and Network Science* 15 (2): 107–118. doi:10.3920/JCNS2014.x017.
Popa, Eugen Octav, Vincent Blok, and R. Renate Wesselink. 2020. "An Agonistic Approach to Technological Conflict." *Philosophy and Technology* 34 (34): 717–737.
Rabinow, P., and G. Bennett. 2012. *Designing Human Practices: An Experiment with Synthetic Biology*. Chicago, IL: The University of Chicago Press.
Rességuier, A., and R. Rodrigues. 2020. "AI Ethics should Not Remain Toothless! A Call to Bring Back the Teeth of Ethics." *Big Data & Society* 7 (2): 2053951720942541.
Rip, A. 2009. "Futures of ELSA." *EMBO Reports* 10 (7): 666–670. doi:10.1038/embor.2009.149.
Rittel, H. W. J., and M. M. Webber. 1973. "Dilemmas in A General theory of Planning." *Policy Sciences* 4 (2): 155–169.
Rose, D. C., J. Lyon, A. de Boon, M. Hanheide, and S. Pearson. 2021. "Responsible Development of Autonomous Robotics in Agriculture." *Nature Food* 2 (5): 306-309.
Ryan, M., and V. Blok. 2023. "Stop Re-Inventing the Wheel: Or How ELSA and RRI Can Align." *Journal of Responsible Innovation* 10 (1): 2196151. doi:10.1080/23299460.2023.2196151.
Ryan M., N. De Roo, H. Wang H, V. Block, and C. Atik. 2024. "AI through the Looking Glass: An Empirical Study of Structural Social and Ethical Challenges in AI." *AI & Society* 1–17. doi:10.1007/s00146-024-02146-0.
Sadek, M., R. A. Calvo, and C. Mougenot. 2023. "Designing Value-Sensitive AI: A Critical Review and Recommendations for Socio-Technical Design Processes." *AI and Ethics* 4 (4): 949–967.
Samuelson, P. 2023. "Generative AI Meets Copyright." *Science* 381 (6654): 158–161.
Sax, M., and H. Wang. 2025. "From Threat to Opportunity: Gaming the Aalgorithmic System as A Service." *Internet Policy Review* 14 (2): 1-27. doi:10.14763/2025.2.2007.
Scherer, M. U. 2015. "Regulating Artificial Intelligence Systems: Risks, Challenges, Competencies, and Strategies." *Harvard Journal of Law & Technology* 29 (2): 353-400.
Sloane, M. 2024. "Controversies, Contradiction, and "Participation" in AI." *Big Data & Society* 11 (1): 1–15. doi:10.1177/20539517241235862.
Smolka, M., and S. Böschen. 2023. "Responsible Innovation Ecosystem Governance: Socio-Technical Integration Research for Systems-Level Capacity Building." *Journal of Responsible Innovation* 10 (1): 2207937. doi:10.1080/23299460.2023.2207937.

Smolka, M., and E. Fisher. 2024. “Testing Reflexive Practitioner Dialogues: Capacities for Socio-Technical Integration in Meditation Research.” *NanoEthics* 18 (1): 1. doi:10.1007/s11569-023-00450-5.
Stilgoe, J., R. Owen, and P. Macnaghten. 2013. “Developing a Framework for Responsible Innovation.” *Research Policy* 42 (9): 1568–1580.
Susser, D., B. Roessler, and H. Nissenbaum. 2019. “Online Manipulation: Hidden Influences in a Digital World.” *Georgetown Law Technology Review* 4 (1): 1–45. https://doi.org/10.2139/ssrn.3306006.
Timmermans, Job, Vincent Blok, Robert Braun, Renate Wesselink, and Rasmus Øjvind Nielsen. 2020. “Social Labs as an Inclusive Methodology to Implement and Study Social Change : The Case of Responsible Research and Innovation.” *Journal of Responsible Innovation* 7 (3): 410–426. doi:10.1080/23299460.2020.1787751.
Tiwari, R. 2023. “The Impact of AI and Machine Learning on Job Displacement and Employment Opportunities.” *International Journal of Engineering Technologies and Management Research* 7 (1): 1–9.
Umbrello, S., and I. Van de Poel. 2021. “Mapping Value Sensitive Design onto AI for Social Good Principles.” *AI and Ethics* 1 (3): 283–296. doi:10.1007/s43681-021-00038-3.
UNESCO. 2021. *To Be Smart, the Digital Revolution Will Need to be Inclusive: Excerpt from the UNESCO Science Report*. https://unesdoc.unesco.org/ark:/48223/pf0000375429.
Van Hilten, M., M. Ryan, V. Blok, and N. de Roo. 2025. Ethical, Legal and Social Aspects (ELSA) for AI: An Assessment Tool for Agri-Food.” *Smart Agricultural Technology* 10: 100710.
Van Wynsberghe, A. 2013. “Designing Robots for Care: Care Centered Value-Sensitive Design.” *Science and Engineering Ethics* 19 (2): 407–433.
Veenstra, A. F., L. Zoonen, and N. Helberger. 2021. *ELSA Labs for Human Centric Innovation in AI*. https://nlaic.com/wp-content/uploads/2022/02/ELSA-Labs-for-Human-Centric-Innovation-in-aI.pdf.
von Schomberg, René. 2013. “A Vision of Responsible Research and Innovation.” In *Responsible Innovation: Managing the Responsible Emergence of Science and Innovation in Society*, edited by R. Owen, J. Bessant, and M. Heintz, 51–74. West Sussex: John Wiley & Sons, Ltd.
Wagner, Ben. 2018. “Ethics as an Escape from Regulation: From Ethics-Washing to Ethics-Shopping?” In *Bein Profiled. Cogitas ergo sum*, edited by Mireille Hildebrandt, 84–89. Amsterdam: Amsterdam University Press.
Wang, H. 2022a. “Transparency as Manipulation? Uncovering the Disciplinary Power of Algorithmic Transparency.” *Philosophy & Technology* 35 (3): 1–25. doi:10.1007/s13347-022-00564-w.
Wang, H. 2022b. *Algorithmic Colonisation: Automating Love and Trust in the Age of Big Data. UvA-DARE (Digital Academic Repository)*.
Wang, H. and Blok, V., 2025. “Why Putting Artificial Intelligence Ethics Into Practice is not Enough: Towards A Multi-Level Framework.” *Big Data & Society* 12 (2): 1-14. doi:10.1177/20539517251340620.
Wang, H., M. Ryan, and V. Blok. 2025. “Teaching Structural AI Ethics: The Design and Testing of an ELSA Education Intervention.” In *Oxford Intersections: AI in Society*, edited by Philipp Hacker. Oxford: Oxford Academic (online edn). doi:10.1093/9780198945215.003.0008.
Wilsdon, J., and R. Willis. 2004. *See-Through Science: Why Public Engagement Needs to Move Upstream*. London: Demos.
Winner, L. 2004. *Testimony to the Committee on Science of the U.S.S House of Representatives on the Societal Implications of Nanotechnology*. Wednesday, April 9, 2003. http://www.rpi.edu/~winner/testimony.htm.
Yeung, K. 2017. “Hypernudge: Big Data as a Mode of Regulation by Design.” *Information, Communication & Society* 20 (1): 118–136. https:// doi.org/10.10801369118X.2016.1186713.
Zuboff, S. 2019. *The Age of Surveillance Capitalism: The Fight for the Future at the New Frontier of Power*. London: Profile Books.
Zwart, H., L. Landeweerd, and A. Van Rooij. 2014. “Adapt or Perish? Assessing the Recent Shift in the European Research Funding Arena from ‘ELSA’ to ‘RRI’.” *Life Sciences, Society and Policy* 10 (1): 1–19.
Zwart, H., and A. Nelis. 2009. “What is ELSA Genomics?” *EMBO Reports* 10 (6): 540–544. doi:10.1038/embor.2009.115.