AI and Ethics (2026) 6:255
https://doi.org/10.1007/s43681-026-01084-5

RESEARCH

Scenario-based sociotechnical envisioning (SSE): an approach to enhance systemic risk assessments

Kimon Kieslich[1] · **Natali Helberger**[1] · **Nicholas Diakopoulos**[2]

Received: 18 July 2025 / Accepted: 10 March 2026 / Published online: 22 April 2026
© The Author(s) 2026

Abstract
Regulators and policy makers around the world have endorsed risk-based approaches for governing the challenges and opportunities from Artificial Intelligence (AI). In theory, these assessments should protect the public from the harmful impacts of AI, thus safeguarding important societal values such as the realization of fundamental rights. However, as an anticipatory decision-making tool, the success and legitimacy of risk-based regulatory approaches depend critically on methods that help policymakers and societal stakeholders anticipate how AI may impact a diverse citizenry, the many different ways in which technological change may unfold, and the factors that lead to desirable and less desirable visions of the future. In this paper, we first point out critical flaws in existing assessment methods and regulatory approaches and consequently identify the need to complement current approaches to risk assessment. We call on scholars, policy makers, and the technology industry to engage in anticipatory approaches that go beyond compliance and checklist exercises and acknowledge the responsibility they have for impacts that go beyond what is known. Second, we propose a concrete empirical approach to address our concerns: Scenario-Based Sociotechnical Envisioning (SSE). By using SSE, practitioners can make assessments more inclusive, more aware of the socio-technical context in which AI systems are deployed, and more prospective in terms of identifying impacts before they occur. As part of our article, we also provide examples and a concrete guidebook for scientists, policy makers and legislators wishing to use SSE as one method in their assessment toolbox. Third, we anticipate possible objections to our approach and proactively discuss how these can be met.

Keywords Risk regulation · Impact assessment · Anticipatory governance · Future scenarios · Method development

1 Introduction

"Risk is a choice, not a fate." (Bernstein, 1996) [10]

Regulators and policy makers around the world have endorsed risk-based approaches for governing the challenges and opportunities from Artificial Intelligence (AI). Most prominently, the European AI Act classifies AI systems into different risk categories and calls for risk and impact assessments among the core requirements for developers and deployers of AI systems. Risk-based approaches also feature prominently in the AI policies of a wide range of countries, including the US, Canada, Singapore, India, Mexico and Brazil [26]. The hope is that risk-based approaches can balance the need to protect important societal values, like health, safety, or the realization of fundamental rights, in a way that is agile, transparent, proportionate and leaves ample room for innovation [20, 70]. Yet, despite their proliferation, there are several key challenges to the effective implementation of such approaches.

Implementing risk-based governance approaches is ultimately a question of assigning power: the power to make choices and decide which values to prioritize and to effectively steer AI development in a particular direction. In many if not most of the existing risk and impact management frameworks the choice is with experts, regulators, technologists and academics, and citizen representatives [68]. The risks and consequences of technology adoption,

✉ Kimon Kieslich
k.kieslich@uva.nl

Natali Helberger
n.helberger@uva.nl

Nicholas Diakopoulos
nad@northwestern.edu

[1] Institute for Information Law, University of Amsterdam, Amsterdam, The Netherlands

[2] Northwestern University, Evanston, USA

however, will materialize as the fate of all citizens. The resulting discrepancy between those having the power to decide which risks are acceptable and futures desirable, and those having to live with those decisions has been heavily criticized [18]. Particularly with general purpose technologies such as generative AI, citizens will also not only be at the receiving end of AI-driven decisions, but will also play a crucial role in determining the direction and impact technology innovations will take, particularly for technologies that are as easily and cheaply available as generative AI. Risk assessment methods should therefore also offer the possibility to actively engage those stakeholders closer to the source of the actual risk into the act of governance [24]. Scholars and experts alike have therefore demanded that risk assessments should be more inclusive and involve diverse communities and people on the ground [18, 47].

Methodologically, however, it is difficult to develop and implement more qualitative, contextual and prospective methods of risk assessment for a number of reasons. First, current approaches focus on quantifiable risks and impacts—leaving out impacts that are inherently difficult to measure (e.g., human rights) or that need to be identified in the first place [31]. Second, risk and impact assessment need to reflect the lived experience of the affected population. Thus, approaches need to combine breadth (diversity of perspectives) with depth (contextual insights into a particular situation), while recognizing that both objective and subjective perceptions of risk are relevant to the regulatory process [54]. Third, involving citizens can be at odds with the very reason why governments embrace risked-based governance approaches, namely to make regulation more efficient, innovation-friendly and "simpler" [27, 52]. And finally, risk anticipation requires forward thinking that is not necessarily quantifiable [28]. Rather, it demands the development of future pathways or scenarios that illuminate how and when mitigation processes should be put in place. It also requires the articulation of values, such as desirable and undesirable futures.

To address these various challenges towards effective risk-based AI governance, we have developed a scenario-based approach to involve citizens and/or a diversity of perspectives in the anticipation of future risks from AI which we refer to as *Scenario-Based Sociotechnical Envisioning (SSE)*. Bringing together literatures on risk regulation, anticipatory governance, future studies and scenario writing, and AI impact assessment, the goal of this article is to describe the development and application of SSE and its theoretical underpinnings. We make a scholarly contribution to the risk and impact assessment literature in pinpointing a threefold critique (structural, prospection, and methodological) of the regulatory landscape related to risk and impact assessment practices. In addition, by introducing SSE, we provide a practical method for those involved in risk-based governance, be they regulators and policy makers, or private and public organizations and civil society, to address these shortcomings. Next to the conceptual contribution of this article, we also share a guidebook and practical examples of how SSE can be implemented and used (see supplementary material of this paper).

2 A critical view on the regulatory landscape and risk and impact assessment practices

Current regulatory approaches and risk and impact assessment practices have been criticized for a number of reasons. In this section, we dive deeper into these arguments and present our main criticisms. We assess the shortcomings of the current approach through two lenses: (1) a *legal lens*, which interrogates regulatory frameworks for AI systems, and (2) an *operational lens*, analyzing dominant risk and impact assessment practices. We will cluster our analysis into three main categories of criticism: (1) a structural critique, assessing power and accountability structures and the consequences thereof (Sect. 2.2), (2) a prospection critique, questioning the anticipatory element of assessments (Sect. 2.3), and (3) a methodological critique, highlighting the limitations of existing methods for assessing risks and impacts of AI (Sect. 2.4). Since Europe has developed one of the first and most comprehensive approaches to legal risk governance of AI so far, we will use the European example as a case study for the legal analysis.

We stress that we do not claim that commonly used expert-based approaches to risk assessment methods do not have their merits and an important function to play. On the contrary, by positioning our research within the risk assessment and risk management literature, we hope to make a constructive contribution to the risk management literature in general.[1] Further, we also acknowledge the importance of risk assessment approaches that rely on the technical knowledge of technology developers that are in possession of specialized knowledge and information that would otherwise be difficult to obtain, for instance, about technical specifications and capabilities of the AI systems at hand [46, 47, 56]. Yet, within the legal framework and risk assessment practices, we also see room for improvement of the current situation. Thus, the following subsections adopt a critical perspective with the concrete goal of identifying

[1] Although we also acknowledge that choosing risk-based approaches comes with its own shortcomings [52]. Yet, to make a practical contribution that can enrich the current practice in methodological approaches and legal frameworks such as the EU AI Act, we decided to align our approach to the concept of risk regulation.

shortcomings, and, based on that, propose ways to *enrich* and complement current approaches (Sect. 3).

2.1 Setting the scene: the regulatory framework and risk and impact assessments

In the legal sphere, risk-based approaches emerge as a common thread in many of the regulatory frameworks that govern the digital transformation. Europe is a case in point. Starting with the obligation for data controllers to carry out data protection impact assessments in the GDPR,[2] to mandatory risk monitoring and mitigation obligations for Providers of Very Large Online Platforms (VLOP) in the DSA,[3] up to the obligations for developers and deployers of high-risk AI systems and more generally the qualification of AI into several risk categories in the EU AI Act, the emerging regulatory framework from Brussels has embraced risk-based approaches in different variations and for a variety of actors [20]. The following analysis of the risk assessment obligations under the DSA and the AI Act follows the model of Gellert [28], who distinguishes risk governance in two phases: a risk assessment and a risk management phase. Thereby, he perceives risk as an "anticipative decision-making tool", where a risk is "an event of a certain magnitude (or severity) that has a number of chances of occurrence and which carries along with it a number of consequences, both positive and negative, which are themselves of a certain magnitude and have a probability of their own" [28:27].[4] Therefore, the assessment phase includes an analysis of the risk (in the sense of a risk event or a particular impact), the conditions or hazards that determine how probable it is that a risk materializes and with which severity (risk factors) and an assessment of the actual impact. The risk management phase consists of a cost/benefit analysis with the goal to decide whether a risk is acceptable or not, and the definition of mitigation measures [28:29]. In our analysis we focus primarily on *impacts*. We conceptualize an impact as the realization of a risk, i.e. an effect, consequence, or outcome of an AI system that positively or negatively affects individuals, organizations, or communities in society. By identifying impacts, we can show the real-world consequences that AI technologies can have on individuals, organizations, and society. These materializations can be understood and imagined by non-experts and, in a further step, can be traced back to the underlying risks that these technologies pose. On the practitioners side, numerous approaches to AI impact assessment have been established. An AI impact assessment can be defined as "a way of identifying positive and negative impacts early on to safeguard AI's benefits and avoid its downsides" [68]. According to Selbst, algorithmic impact assessments principally fulfill two goals: "The first is to get the people who build systems to think methodically about the details and potential impacts of a complex project before its implementation, thereby heading off risks before they become too costly to correct. […] The second goal is to create and provide documentation of the decisions made during development and their rationales, which in turn can lead to better accountability for those decisions and useful information for future policy interventions." [61:122] While Selbst puts emphasis primarily on the developers of technology, Moss and colleagues understand algorithmic impact assessments in a broader sense [49]. They state that "impact assessments mean many different things to different actors—they may be about compliance, justice, performance, obfuscation through bureaucracy, creation of administrative leverage and influence, documentation, and much more." [49:5] Its goal is produce accountability in highlighting potential risks and harms of technology and to outline mitigation strategies [47].

2.2 Structural critique

The overall goal of the emerging EU regulatory framework is to respond to (popular) concerns and narratives about the risks from AI technology and instill a sense of trust that those risks are under control, without hindering innovation and the realization of opportunities.[5] By outsourcing parts of the responsibility for identifying and mitigating risks to those actors that are closer to the source of risk or even control it (platforms, data controllers, technology developers), risk-based approaches also hold the promise of more efficiently using the limited resources of regulators [25]. In so doing, risk-based approaches result in new divisions of tasks and different degrees of co-regulation [28, 52]. Importantly, in the context of regulation, risk assessment frameworks are an element of establishing accountability [49], and therefore to be legitimate cannot be seen separately from an effective accountability framework with democratic checks, balances and responsibilities along the value chain. A closer analysis of the different risk assessment obligations in the DSA and the AI Act does therefore also reveal important information on who is accountable for which risks, as well as which risks are still unaccounted for.

[2] Art. 35 GDPR.

[3] Arts. 34 and 35 of the DSA.

[4] The components of this definition are shared by risk assessment bodies like NIST that define risk as follows: "risk refers to the composite measure of an event's probability (or likelihood) of occurring and the magnitude or degree of the consequences of the corresponding event." [50:2].

[5] See European Commission, AI Act Impact assessment, p. 36: "in the absence of an EU legislation on AI that addresses the new specific risks to safety and fundamental rights, businesses and citizens distrust the technology."

Neither the DSA nor AI Act require a measure of the actual impact of the identified risks. Mitigation measures are not made conditional upon the actual size or seriousness of an impact, or those impacted.[6] An exception is Article 9 (8) AI Act, which obliges developers to take into account the impact on persons under 18 and vulnerable groups. While the providers of high risk AI have to test their systems for all individual and societal risks, the risk monitoring obligations for VLOPS and for providers of Generative AI are limited to so called "systemic risks." Systemic risks are risks from systems that are "embedded in the larger context of societal processes" and therefore also impact not only individuals but society at large [34:5]. Unlike individual (idiosyncratic) risks, systemic risks are far more complex, potentially, also depending on the maturity and proliferation of the technology, involving a whole range of unknowns [23]. Identifying systemic risks can involve the identification of core and peripheral players, their relationships towards each other and the AI systems in question, and the contributions of individual players to the realization of impact [23]. And yet, neither the DSA nor the AI Act require engaging with stakeholders, diverse communities or actual citizens as end-users. In other words, it is enough to work with legal fictions like the "average user" [24:12 and 50]. The consequence is that differences between different users or groups of users in society, their particular situation, dynamics and potential individual harms remain invisible. In a similar way, the decisions of whether or not a risk is acceptable in the risk mitigation phase is left solely to the provider of the AI system (AI Act)[7] or the platforms (DSA),[8] and providers and developers are not required to engage stakeholders into this process. These regulatory shortcomings are largely mirrored in current risk and impact assessment practices. Moss and colleagues emphasize that impact assessments are power-laden practices in itself as the "ways in which various components are selected and various forms of expertise are assembled, are directly implicated in the distribution of power." [49:49]. Thus, impact assessments will only be effective if they can actively enable changes and repercussions, i.e. constitute accountability, for developers of technology [47]. Given the legal analysis, we can observe a structural hurdle to their enforceability, as the responsibility for risk and impact assessments is placed in the hands of corporations themselves, which are the same entities that also determine the metrics and measurements, or which risks to look into in the first place [52]. Thus, companies decide on value conflicts and the acceptability of risks for society, for instance, on how to calibrate generative AI tools [29]. This ultimately leads to the reinforcement of existing power relationships. Griffin argues that this could lead to the point "that corporations selectively focus on those risks and mitigation measures that are least disruptive for their business goals" [31]. Further, "[p]roviders or deployers who assess themselves can also be expected to look for ways to reduce risks and optimise existing models and systems only as far as is necessary to achieve the level of risk they themselves have determined to be acceptable rather than considering the option of not using models and systems at all." [52:21] Consequently, there is the real possibility that corporations will search for a cost-effective way to conduct these risk assessments and rely on established audits, checklists, and other forms of structured ways to comply with legal obligations [52, 71].

Thus, risk and impact assessments can be seen as prone to bias, and their "objective" character should be questioned in the first place [70] as these assessments require highly political and contextual decisions [52]. Who defines and assesses the risks, what evidence is relied on and how it is measured, what contributed to a risk being significant – all these steps involve decisions that are contestable.[9] Further, Schmitz et al. showed that aggregating risks in the risk management process is largely dependent on human interpretation as it is "to a large extent unclear how to calculate meaningful (real) probabilities for the occurrence of technical malfunction" [59]. Consequently, it is up to debate how trade-off situations will be solved, especially when organizational profits and economic values are weighted against societal benefits and public interests [68]. Considering past examples of the enforcement of privacy regulation and the allocation of decision power within assessment procedures [71], there is a clear risk that societal concerns can be outweighed by economic interests when it is companies that are required to the balancing decision. In a similar way, scholars pointed to this risk in the context of the DSA and the AI Act, which both require companies to identify and decide on the acceptability of (systemic) risk. Scholars point out that the range of expertise in companies is limited, which is why involving external parties such as NGOs, researchers, journalists or members of civil society organizations can offer domain expertise [34]. This is even more relevant since internal AI auditors are often not aware of the way technologies affect marginalized communities [34]. Orwat and colleagues describes the ethical challenges as follows: "it is doubtful

[6] Instead, the AI Act only requires an abstract estimate of the "severity of the possible harm and its probability of occurrence and they are used in a number of specifically pre-defined areas specified in the Regulation" (Recital 32 AI Act).

[7] Art. 9 (4) AI Act and 52d AI Act.

[8] Art. 34 and 35 DSA.

[9] For instance, the European Commission consultation process for the drafting of the first General-Purpose AI Code of Practice provided a pre-defined list in their consultation form without explaining how they derived the risks in the first place; this pre-selection already steered the consultation process in an overly technical direction.

whether providers or deployers will identify, investigate and assess the potential risks to fundamental rights from the perspective of the persons affected. It is less likely for instance that providers or deployers will consider risks to the common good, potential chilling effects or the distribution of risks among population groups" [52:21]. Similar concerns have been raised before in regard to ethical guidelines and their enforceability when meeting industry goals [48].

This is corroborated by the structural embedding of risk assessments as part of an innovation-oriented legal framework (e.g., EU AI Act) that turns impact assessments into "purely functional tools to ensure that AI's benefits can unfold" [68]. Importantly, both the AI Act and the DSA allocate the primary responsibility with identifying and mitigating systemic risks with technology companies themselves [34, 52, 69]. The resulting challenge is well described by Orwat and colleagues: "The fact that trade-offs are decided by actors who lack democratic legitimacy is another problem. Such trade-offs may also be based on an inadequate formal foundation (e.g., without any legal basis) or take place covertly despite the requirement for procedures to be conducted in public political processes." [52:23]. Yet, domain experts (e.g., NGOs, researchers, civil society members) possess domain knowledge that contains contextualized insights which may be ignored, or simply not on the radar of technological experts or first party auditors [34]. Thus, the inclusion of third parties (e.g., civil society, academics, journalists) enriches current assessment practices in providing oversight, questioning established standards, identifying new challenges of technologies, and highlighting the socio-technical nature of AI systems [34, 56]. Putting these theoretical observations into empirical terms, scholars have reported on limitations and biases in current risk assessment practices. Comparing a comprehensive set of first and third party evaluations of foundation models, a research consortium reported major gaps in first party assessments, highlighting that "first-party reporting is sparse, often superficial and has declined over time in key areas such as environmental impact and bias, while third-party evaluators including academic researchers, nonprofits, and independent organizations provide broader and more rigorous coverage of bias, harmful content, and performance disparities." [56:1]. Regarding technology foresight in general, researchers have reported on expert bias throughout the whole technology foresight process, beginning with the problem definition, through modelling, to the mapping of potential impacts [13]. Specifically in regard to AI risk assessment, Schiff and colleagues conducted interviews with AI ethics auditors and reported that there is a "significant gap" [58:8] in auditing practices when it comes to the inclusion of the general public, or vulnerable groups. As a consequence AI audits are primarily focusing on technical risk management, while largely neglecting socio-technical approaches [58]. Moreover, focusing on ethical behavior of technology developers, a survey study among AI developers found that – even if they were instructed to follow ethical standards – their practice didn't change [45]. This substantiates the claim that leaving technological, or first audit parties, in control of acting ethically, or identifying risks, is not sufficient to address potential risks holistically. Moreover, researchers also found that including lay stakeholders and domain experts in risk assessment practices for the use of GenAI in the news environment, for legal conflict resolution and for personal chatbots proved successful in identifying some novel risks that are not commonly mentioned in established risk frameworks [37–39].

Further, defining risks as something directly measurable creates the illusion of an objective, non-political discourse – a discourse that is then best handled by experts and bureaucrats rather than political actors, civil society groups or even citizens [31, 75]. In this sense, we reinvoke the critique of Gillespie, who states: "[c]ultural politics cannot be averaged to a satisfactory consensus. And ceding the power to manufacture a consensus to powerful, corporate intermediaries has had implications before" [29:10]. Indeed, we can observe this happening in AI risk and impact assessments, be it that the definition of risk and impact categories rely either on literature review on scholarly literature [62, 63], direct corporate authored reports [73, 74], or expert judgements [16, 65].

A summary of the structural critique is shown in Table 1.

2.3 Prospection critique

In the risk assessment literature a difference is made between risks and uncertainties, a distinction that is also useful to understand the kinds of risks both the DSA and the AI Act try to address. Ebers explains: 'risks are "known knowns" with statistical probabilities and quantifiable effects, uncertainties are "known unknowns" that cannot be quantified because we do not know what the effects of a particular technology might be. In addition, there are "unknown unknowns", where we are not even aware that things or activities may have adverse effects at all.' [24:7].

Some of the risks that both the AI Act and the DSA must address are "known knowns", in the sense that they can be measured with the help of empirical evidence and objectifiable criteria. Examples are the obligation to monitor risks from the dissemination of illegal content, gender-based violence, protection of minors or physical and mental well-being (DSA), respectively health and safety (AI Act). In contrast, risks to fundamental rights (DSA and AI Act), civic discourse or electoral processes are "known unknowns" and more difficult to express in technical standards or predefined

Table 1 Structural critique

Critique	Description
C1: Lack of democratic accountability	Risk-assessments are the cornerstone of a risked-based approach to the governance of AI and digital platforms. However, expert-centric and heavily institutionalised risk-assessments that do not sufficiently engage with, and give a voice and agency to affected citizens and societal stakeholders, create a democratic accountability gap.
C2: Depoliticization of assessments	The reliance of expert judgement and quantification leads to a depoliticization and bureaucratisation of assessment practices. Consequently, existing power structures are likely to be strengthened rather than challenged.
C3: Lack of inclusion	Most assessments do not require public input. Yet, the public / affected communities bring a special set of expertise and perspectives to the assessments of impacts that is currently largely ignored.
C4: Expert-bias	Assessments hugely lie in the responsibility of experts, whose judgements are subjective and can show crucial blindspots and/or are guided by corporate / academic / political interests.
C5: Limited guidance for trade-off decisions	Trade-off decisions and the decision of whether a particular risk is acceptable for society are mostly in the hands of (corporate) experts and suffer a lack of public input. Additionally, they mostly rely on a weighing of (subjective) metrics.

metrics and probability thresholds because of their deliberate conceptual vagueness, contextual nature and the need to balance different, potentially even conflicting values in their assessment [52]. This is similarly the case for risks that have not yet materialized (whether or how often is still uncertain), or cannot be simply pinpointed to a particular system but are the result of a complex interplay of actors, factors and the context in which an AI system is integrated.

And yet, both the AI Act and the DSA also frame uncertainties (known unknowns) in terms of risks that can be subjected to measurable criteria (or even automated tools). Gellert explains why: "It is in the context of the complex risk objects that the need for alternative assessment methods more qualitative in nature and more participatory emerged. Yet, because of the need to resort to risk management and its convenience as a ready to use tool […], there has been a tendency to keep treating events as risks. This has been achieved by subsuming them under objective and scientific risk assessment and methodologies even though it is, strictly speaking, impossible to do so" [28:39].

Concretely this means that reliance on objective methods to discern risks based on hard evidence becomes more untenable the more one moves into the area of still uncertain "reasonably foreseeable risks" and subjective risk perceptions. Here, more qualitative methods can be more time and cost intensive but potentially also more meaningful if they are able to provide the necessary context to reduce uncertainty in modeling foreseeable risks. This also corresponds to the standardization deliverables requested by the European Commission that, summarized by Soler Garrido et al., state that "risk management requirements defined in standards should provide strong assurance and documented evidence that all the relevant foreseeable risks of the AI system have been identified and addressed"[66].[10] It is detrimental to all subsequent quantitative work to identify these (foreseeable) risks and impacts in their breadth in the first place, but also to provide information on the acceptability of the (remaining) uncertainty.

The distinction between risks and uncertainties is also instructive for better understanding the risk factors that the DSA and AI Act do address, or rather: do not address. Interestingly, particularly in the AI Act there is a pronounced lack of engagement with "unknown unknowns" – developers are only obliged to identify and analyze risks from the "intended use" of their AI systems, with a focus on how the systems are developed and designed (9(2a) AI Act). Within the sphere of intended use, developers must also "estimate and evaluate" reasonable foreseeable misuses by deployers, but only to the extent that these "may be reasonably mitigated or eliminated through the development or design of the high-risk AI system, or the provision of adequate technical information" (Art. 9 (2a) AI Act), and therefore remain within the sphere of influence of the developer and are amendable to technological fixes. In a similar way, the DSA concentrates on risks within the sphere of influence of the providers of VLOPS and VLOS, namely the design and functioning of the automated content moderation and recommendation algorithms of the platforms (Article 35(2) DSA). The DSA, too, acknowledges that risk factors can also be the use of the platforms by third parties, notably the intentional manipulation, dissemination of illegal content or inauthentic uses. Again, there is the underlying assumption that there are intended (or "authentic") and unintended uses of platforms, and that the risk monitoring obligations do not extend to uses that platforms (or developers) did not anticipate in the first place or cannot address by adjusting their technologies. Insofar, both DSA and AI Act handle a rather narrow understanding of risks by excluding "unknown unknown's" from the scope of the obligatory risk assessment.

[10] Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonized rules on artificial intelligence and amending Regulations (EC) No 300/2008, (EU) No 167/2013, (EU) No 168/2013, (EU) 2018/858, (EU) 2018/1139 and (EU) 2019/2144 and Directives 2014/90/EU, (EU) 2016/797 and (EU) 2020/1828 (Artificial Intelligence Act). Regulation-EU-2024/1689.

Table 2 Prospection critique

Critique	Description
C6: Exclusion of unknown unknowns	Future impacts of AI might emerge that are not readily identified. Current assessments have limited guidance on how to anticipate those impacts (unknown unknows).
C7: Focus on technical risks	There is a focus on technological risks rather than aiming for structural changes (e.g., power balances).

The legal focus on known knowns can also be observed in the practical reality of existing risk and impact assessments. Oftentimes more qualitative and resource heavy forms of assessments that are required to detect "unknown unknowns" or risks for fundamental rights are sidelined [52]. This conclusion is supported by the findings of the meta-review of Stahl et al. who report that "a core component of the AI-IAs is typically a set of questions in the form of a questionnaire or checklist to which the AI-IA seeks responses"[68]. They further state that the criteria definition and grading scales for risks and impacts are highly subjective and, ultimately, rely on the assessor's perception of an issue [68].

Our prospection critique is summarized in Table 2.

2.4 Methodological critique

In terms of the actual risks to be measured, both the DSA and the AI Act make a distinction between actual or known and foreseeable (DSA), respectively "reasonably foreseeable", "likely" or "possible" risks (AI Act), thereby introducing elements of foresight and anticipation to risk assessment. There is, however, little, if any, guidance on how to measure the risks, though the AI Act shows a preference for objective measures[11] and limiting the burdens and costs of risk assessments for developers and deployers. Developers are only expected to monitor for "reasonable foreseeable risks," the European standardization bodies are asked to develop technical standards to facilitate compliance and for deployers that have to conduct a fundamental rights impact assessment the AI Act even suggests that the AI Office[12] shall develop a questionnaire, including an automated tool "to facilitate deployers to implement the obligations of the AI Act in a simplified manner".

However, scholars point out that to provide such metrics, existing data must be simplified and taken out of context, or the data may be incomplete to begin with [28, 52, 53] – and thus, "lose the richness, depth, and nuanced understandings that qualitative methods can provide" [72:301]. Or, in other words, common risk and impact assessment methods can fail to account for the complexity of the social world [53], which can lead, in the worst case, to a violation of fundamental human rights as Orwat and colleagues point out: "These data analyses do not take account of individuality of persons and their unique quality. Instead, they sort individuals into "prefabricated" categories of group characteristics based on single or multiple variables. A violation of human dignity can result from persons being treated as mere means, instruments or objects." [52:6] Consider, for example, the famous cases of discriminatory algorithms used to grant parole [3]. In this case, lives are at stake: whether in terms of individual freedom or in terms of having enough money to live on. This example also shows why relying on metrics is not enough: How, for example, can the price of freedom be quantified? And should it also be accepted in the first place that human rights of persons are endangered [52]? Stories describing these impacts are powerful tools for framing the real life impacts of these risks beyond reporting numbers, such as how many people have been mistreated by an AI.

The lack of the contextuality and qualitative nature of risks and impacts is further supported by the "convenience" of risk management [28] and the criticism of the understanding of risks in the legal documents as "simple-realist: meaning it assumes that risks have some objective existence, external to the institutions charged with risk management, whose task is simply to study these risks in order to manage them properly." [31]. Correspondingly, Stahl et al. found in their meta-review that although trade-offs between issues are recognized, they are mostly resolved by relying on metrics, be it "a quantitative scoring of issues and risks or the determination of key performance indicators" [68]. Thereby, the focus in risk assessments is often on technical-caused risks rather than human-caused risks. Based on an analysis of 43 risk taxonomies of academic and corporate funded frameworks, the MIT risk repository reports that risks that are primarily described as being caused by the actions of an AI system itself (e.g., generation of harmful content) rather than human actions (e.g., intentional manipulation) [63]. Further, in a comparison of first-party and third-party audits, another research team found that third-party, independent auditors provided the most thorough analyses of bias, performance disparities, and sensitive content, compensating for areas where developers report little or nothing." [56:10]. In other words, there is a tendency to neglect the human and societal element in risk and impact assessment and focus

[11] Article. 9 (7) AI Act: "Testing shall be made against prior defined metrics and probabilistic thresholds that are appropriate to the intended purpose of the high-risk AI system".

[12] The European AI office is a unit within the EU commission which should ensure trustworthy AI development and safety, and will be tasked with supporting the implementation of the EU AI Act.

Table 3 Methodological critique

Critique	Description
C8: Focus on narrow compliance	Structured ways of assessment (e.g., audits, checklists) establish a compliance structure, largely ignoring impacts out of the scope of these measurements. Complex impacts may not be captured.
C9: Focus on quantification	Events / issues are simplified as something objectively measurable. In many instances, this includes a massive reduction of complexity and introduces bias (e.g., measurement of human rights impacts).
C10: Lack of contextuality	Some impacts only emerge through a complex interplay of different actors and circumstances. Current assessments largely ignore the interplay of human and AI, or the situational embedding of AI deployment.
C11: Focus on the technology as primary source of risk	Many AI risk assessment frameworks and also the regulatory framework tend to focus narrowly on risks from a particular AI system, failing to acknowledge that risks can also be the result of the human, institutional, cultural, political and organisational context.

more on the technological features as the primary source of risks, especially if conducted by first party auditors. It is important to take into account the incentives for instance, developers tasked with assessing the risks that their technologies pose to prioritize risks that are amendable to technologically fixable risks, rather than on more complex risks like risks to abstract fundamental rights, or risks that are the result of a complex interplay of technological, organizational and societal factors that may lie outside their control. As a result, there is a danger that such more complex risks will be oversimplified or even not addressed at all: "Many fundamental rights and societal values are hardly comparable or commensurable. It is virtually impossible for instance to assess violations of human dignity or restrictions of freedoms in quantitative terms. […] Furthermore, it is inherent to the ethical nature of human dignity that humans cannot be evaluated in quantitative terms and should only be evaluated according to their unique individual qualities" [52:15]. A summary of the methodological critique is shown in Table 3.

2.5 Moving forward

Our critique of the existing legal approach to anticipatory AI governance is that the risk assessment obligations in the DSA and AI Act are not inclusive, unable to take into account broader systemic factors, unsuitable to account for individual (group) differences, and do not offer guidance on how to balance value conflicts. And also current assessment practices have a tendency to invoke the power and legitimacy of objectifiable criteria, do not leave room to navigate value conflicts, tend to ignore group-specific perceptions of harms and the socio-technical interplay of end-users with the technology, and invite bureaucratization and over-reliance on the perception of one particular stakeholder (developers), while generally suffering from a democratic deficit to the extent that they fail to actively involve and engage the societal stakeholders that are affected.

In response to the critique of existing approaches to risk and impact assessment practices, there have been calls for more inclusive approaches and democratized AI practices [46, 47]. Metcalf and colleagues state: "Expertise is […] not limited to professional capacities. Individuals and communities affected by algorithmic systems are often the foremost experts in the potential harms they regularly encounter, as well as the strategies they have developed to minimize or avoid such harms." [47:743] Thus, including lived-experiences and the socio-technical component is essential to understand risk and impacts caused by algorithmic systems [49]. This also implies that risks and impacts should not be understood solely in terms of quantifiable metrics, but that the identification and measurement thereof is highly political and contextual [49, 75]. Crucially, this call is echoed by the National Institute of Standards and Technology (NIST) of the United States[13] [50].

To achieve this, there is a need for alternative methods that are more qualitative in nature and do not rely exclusively on quantification [28]. That also implies acknowledging the limits of metric-driven risk assessment and with that the uncertainty of future developments [53]. Further, anticipatory approaches are needed as most of the current risk and impact assessment approaches focus on risks that emerged after deployment and not on risks during the development phase [63] – and thus are rather reactive then pro-active. But identifying risks early in the development process allows developers to engage in far less costly mitigation approaches [32, 61]. Thus, a shift to more anticipatory approaches could be effective in the sense that it is more feasible to actually bring about organizational and developmental changes. This is only possible, however, with approaches that allow to

[13] NIST advocate the following measures in their guidelines: "MEASURE 3.2: Risk tracking approaches are considered for settings where AI risks are difficult to assess using currently available measurement techniques or where metrics are not yet available" as well as in "MEASURE 3.3: Feedback processes for end users and impacted communities to report problems and appeal system outcomes are established and integrated into AI system evaluation metrics".

actively involve and engage diverse user groups, including laypeople.

Scenario-based methods can be a promising approach to anticipate risks, and doing so in a way that is able to also engage non-experts, and tap into the situated experience or range of domain-expertise of a diverse group of users. The goal of scenarios is to describe multiple, alternative, *plausible* futures [55]. Scenario-writing is used to engage participants in anticipatory thinking and outline a multitude of future developments [1, 14]. Thereby, scenario methods acknowledge the uncertainty that is inherent to the future and, thus, does not aim to predict future outcomes but rather describe plausible futures [55]. Scenarios can be developed in the form of creative description, for instance written stories or visual media [1, 14]. The International Organization for Standardization (ISO) acknowledges the merit of scenario analyses for risk management as they support "diversity of thinking", help "monitoring of lead indicators of change", and offering insights for situations where little to no knowledge about the future development exists [35].

In this paper, we introduce a methodological approach, *Scenario-Based Sociotechnical Envisioning (SSE)*, that responds to the range of critiques we have described in this section. In the next section we will elaborate on the approach and how it addresses the critiques.

3 Scenario-based sociotechnical envisioning (SSE) as a method for anticipatory risk governance

3.1 Terminology

SSE is a flexible method that can be used to assess impacts, for example in the context of an impact assessment or identification of risk mitigation strategies. SSE is an applied assessment tool which offers users the opportunity to adapt multiple elements (see Sect. 3.2.) to their aims. While current assessment practices are still of pivotal importance as they rely on primary information that might not be accessible to the public or regulatory bodies (e.g., technical system information), SSE is a complementary approach with the goal to help fill critical gaps left by expert-based risk assessment approaches that do not take into account the perspective of users. Combining the insights from current approaches with novel, context-sensitive, socio-technical approaches such as SSE could help risk assessors and regulators to get a more nuanced (and also representative) understanding of the risk landscape and the necessary mitigation measures.[14]

The elements of SSE are the following:

- *Scenario-based*: Narratives as the core of scenarios make complex issues concrete, for instance through describing interactions of actors with technology and its associated impacts in an explicit context and setting, and, thus, makes these issues more accessible and salient for a broader audience (e.g., political decision-makers, or the public).
- *Sociotechnical*: The *socio*-technical aspect entails the human interplay with technology and brings situated knowledge to the surface. Especially in the case of general purpose technology the possible intended uses are virtually endless—gauging the insights from different user groups that use the technology in different ways is essential to understanding the range of risks and impacts that can emerge.
- *Envisioning*: SSE is about anticipating risks and impacts of technology on individuals and society. Grounded in anticipatory governance [32], this aspect highlights the necessity to illuminate future pathways of technological interplay with individuals and society to inform decision-makers—whether regulatory or developer driver—early on in technology development.

SSE is, in principle, applicable for every AI technology. However, SSE is most useful for technologies,

- Where impacts or types of impact are potentially diverse and difficult to anticipate, i.e. unpredictable due to the reach or take up of the technology.
- Where performance of the technology is mostly the result of its interaction with people or socio-institutional contexts.
- In situations where impacts are difficult to quantify, e.g. understanding impact on human rights.
- Where technologies have potentially severe or transformative implications for users, or different groups of users.

All of these assessment types require societal input and a form of anticipation that exceed current practices.

SSE allows its users to tap into different application domains. In the case of Generative AI, we utilized SSE in the domains of the news environment, general purpose chatbots, the future of academic knowledge production, and legal conflict resolution [6, 7, 37–39, 41].

SSE was developed for multiple user groups. While we, as an author team, used SSE for academic studies in the AI risk assessment field, we highlight its potential for other user groups that aim to conduct socio-technical analyses of risks (and benefits) of new technologies. For instance,

[14] We also do not conceive SSE as the sole solution to the systemic critique we raise. On the contrary: We encourage other scholars to build on our critique and develop more approaches that tackle several, or all of the shortcomings of current risk assessment approaches.

for the use case of generative AI in the news environment, we encourage news organizations to engage in SSE to brainstorm potential risks and benefits for their operations. In this case, media organizations might want to tap into risk and benefit anticipations from different perspectives. From the journalistic side media organizations might want to explore which practical (e.g. technical capabilities and limitations), or normative (e.g., journalistic values) issues occur in a potential implementation process. On the other hand, scenarios from news readers are also important as they are the consumers of journalistic content. SSE can help media organizations to narrativize those perspectives and create an easy accessible resource as basis for further strategic decision making. Further, SSE can provide a viable method for NGOs who want to engage with technology impacts, and develop policy ideas on how to mitigate specific risks.

We also emphasize the potential of SSE as part of risk assessment practices with non-expert and non-corporate voices. To give a concrete example: under the AI Act, providers of general purpose models with systemic risks are obliged to identify and mitigate any systemic risks arising from their models (Art. 55 (1) AI Act). This obligation is further concretized and operationalized in the Code of Practice for General Purpose AI Models[15] that instructs providers how to perform these risk assessments. The Code of Practice instructs AI providers to provide relevant information inter alia via forecasting methods, but also community consultations or "other participatory research methods investigating, e.g. the effects of models on natural persons, including vulnerable groups" (Measure 3.1 Code of Practice). As a participatory forecasting method, SSE can be usefully employed in that context. In a similar way, SSE can be a useful method to use for providers of Very Large Online Platforms and Very Large Online Search Engines in the context of the systemic risk assessment obligations under Art. 34 and 35 DSA. Having said so, SSE can also be employed by regulators, auditors, researchers and civil society organizations to identify systemic risks that were not identified by AI providers and platforms, and in this way contribute to the effective enforcement of risked-based regulations, like the DSA and the AI Act.

Alongside this paper, we published a guidebook that discusses methodological considerations for applying SSE in different settings and presents example scenarios from our previous studies. We encourage potential users of SSE to thoroughly engage with the guidebook (see supplementary file to this article) and previous studies of the author team to get practical insights in how to conduct surveys or facilitate workshops using SSE. The guidebook also entails helpful pointers on sampling considerations, how instruction information should be conveyed to respondents or participants, and how reflexivity of the users should be integrated in the application process. In the following, we provide an *overview* of the methodology and discuss how it addresses the critical observations outlined in Sect. 2.

3.2 Data collection

3.2.1 Data collection considerations

Methodologically, SSE can be applied in a survey or a workshop setting. The core of SSE is a writing task that participants undertake individually. Participants first get introduced to a specific technology including a description of what a story entails[16], information about the capabilities, limitations and trends of the technology at hand[17], and quality criteria of a well-written story[18]. Afterwards, participants write a story that outlines their projection of the future impacts of a particular AI application and context, such as generative AI chatbots in the domain of the online information ecosystem. While scenario-writing is the core element of SSE, users of SSE can complement the task with further survey questions or workshop tasks—depending on the purpose of the assessment that nudge them to reflect upon the scenario that they or others have written. For instance, after engaging in scenario-writing, participants can be tasked to reflect on the story they just wrote in terms of a variety of factors, including:

- The underlying values that guided their writing process,

[15] Chapter *Safety and Security*: https://digital-strategy.ec.europa.eu/en/policies/contents-code-gpai.

[16] In our studies, we used the following description: "A scenario is a short fictional story that includes: (1) a setting of time and place, (2) characters with particular motivations and goals, and (3) a plot that includes character actions and events that lead to some impact of interest." [37].

[17] For scenario-writing to be successful (in the sense of producing meaningful and plausible stories), participants must have a basic knowledge about the technology. Thus, we deem this step necessary for respondents who do not necessarily have had contact with the respective technology (this step may be left out in expert samples). However, we acknowledge that providing information about capabilities, limitations, and trends includes a deployer bias as the ones deploying this method decide on which capabilities, limitations, and trends are made salient—and which are not.

[18] In our studies, we used the criteria creativity, specificity, believability, and plausibility, which are common scenario-quality criteria in the scenario literature [37–39]. Though, depending on the goal of scenario-writing, scholars who apply this method can also define other criteria that are, for instance, more aligned to the assessment task, e.g., severity, likelihood, magnitude.

- Potential mitigation strategies,
- The evaluation of existing policies in relation to the scenario,
- The assignment of accountability for identified risks and impacts,
- The subjective judgment of whether a risk is acceptable or not.

In principle, SSE allows its users to include a variety of complementary questions that are used to gain deeper insight into why respondents wrote their specific stories and what (real-world) implications those stories might have. A crucial step in applying SSE is also whose prospection to gauge. In principle, we recommend to identify potential stakeholder groups beforehand and draw an appropriate sample (for more detailed reflections about sampling, see our guidebook). Table 4 shows a summary of data collection considerations that needs to be reflected on when using SSE.

3.2.2 Positioning SSE among participatory frameworks

SSE is conceptualized as a participatory and inclusive method that is explicitly targeted to gather stakeholder perspectives that are commonly neglected in impact assessment practices. In particular, research on participatory auditing [34] and participatory foresight [51] are central inspirations for SSE. Studies in this field explicitly aim at leveraging cognitive diversity, i.e. engaging diverse stakeholder groups, to anticipate future impacts of technology [15, 51]. In this way, perspectives on AI impact assessment can be enriched, and a more comprehensive dialogue on desirable futures can evolve that grounds expert positions into everyday social life [47, 51]. Thus, participatory approaches in technology assessment can have a democratizing function, and aligning technology development with societal values [9, 60]. There are numerous data collection approaches on how to gather stakeholder input for participatory research, including citizen juries, participatory action research, technology assessment panels, workshops, focus groups, or interviews [2]. SSE is proposing a methodological approach that is aimed to increase accessibility of technology impacts, as well as inclusion through the incorporation of stakeholder voices, including lay stakeholders [8, 17]. As such, SSE can generally be implemented in various data collection methods. For instance, workshops offer the possibility to foster in-person follow-up conversations about the scenarios. Citizen juries can include scenario-writing elements [67], but also technology assessment panels can incorporate innovative concepts [21, 36]; SSE can thus be applied for deliberating about potential technology impacts and follow-up discussions on policy reactions for either citizen juries or technology assessment panels.

In the field of technology impact assessment and scenario writing there is also a debate on different engagement forms and the reach of participatory methods [5, 21]. Delgado and colleagues differentiate four levels of participation in participatory research: (1) consultation studies aim primarily at getting feedback on technology or policy options; (2) inclusion studies are working with pre-defined scenarios, or policy options, but give participants the ability to interrogate, and alter those to shape technology development; (3) collaboration studies include participants even in the creating process of scenarios or technology development, but leave decision authority in the hands of researchers and technology developers; (4) owning studies define participants as

Table 4 Data collection considerations

Summary of data collection considerations	
Procedure	Individual scenario-writing task after introduction to (1) scenario-writing exercise, (2) technology in focus, (3) quality criteria.
Time frame	Flexible and dependent on the anticipatory endeavor: Quantitative assessments (e.g., impact assessments) are more useful for shorter time frames, while qualitative assessments are more useful for longer time frames (e.g., formulating (un)desirable scenarios).
Quality criteria	Emphasis on creativity, specificity, believability, and plausibility of scenarios. Scenarios should be based on expertise and lived-experiences of participants.
Data collection method (Examples)	*Workshop*: Allows for deliberative forms of participant engagement, such as having participants discuss the scenarios in a group and brainstorming on mitigation strategies and responsibilities. *Survey*: Allows for collection of a larger number of scenarios. Collecting a larger sample allows researchers to use SSE to sample ideas from a potentially more diverse array of participants and to use that larger scenario corpus as a basis for conducting risk or impact assessment.
Sampling	The sample size depends on the study's objectives, with a recommendation to collect scenarios iteratively until no new impacts emerge (for surveys). SSE can also be used to compare stakeholder perspectives, trace impact enumerations to socio-demographic factors, and amplify the voices of underrepresented groups in discussions about AI risks and impacts.
Additional questions	The scenario-writing task can be complemented with further survey questions or workshop tasks to stimulate additional brainstorming activities. Participants can be invited to reflect upon the scenario that they or others have written. For instance, respondents can be tasked to develop mitigation strategies. In workshop settings, it is also possible to include a participant-review stage in which respondents evaluate scenarios and present interpretations of the material.

equal partners and aim at involving these in a co-creation process in every step of the project [21]. Depending on the goal of users of SSE, the level of participation can occur at any of these four levels. When using a scenario-writing task, i.e. tasking respondents to write scenarios themselves, we conceptualize SSE as a collaboration design as also indicated by our previous studies [6, 37–39]. Yet, especially when working in partnership with applied partners like NGOs, SSE can also be used on the owning level; and as such used as a method to complement participatory action research [19]. Further, researchers might want to take predefined scenarios – either written by the research team or sampled from a previous study – to evaluate specific scenarios, or policies [see 7, 40 for example studies]. When taking this approach, the participation level falls either in the category of consultation or inclusion.

3.2.3 Example application of SSE

To provide a more concrete example on how an example study can be performed, we briefly outline the methodological decisions we made for our study on the impacts of GenAI on the news environment [37]. In this study, our scenario was scoped towards the question "how generative AI technology could create risks or negative impacts for the news environment". We chose to set the time frame in the near future, in five years from now, as we aimed to inform current risk assessment practices. A five year time frame aimed to engage respondents in future thinking without falling too much into speculative science-fiction visions. We defined creativity, specificity, believability, and plausibility as quality criteria, and emphasized that scenarios should be informed by personal, or professional experiences.

We sampled three different stakeholder groups to gauge different prospections: technology developers as technical experts, content creators (e.g., journalists) as domain experts, and news consumers (citizens) as lay stakeholders. We also aimed for a cohesive mapping of risks, and thus aimed for a substantial sample size. We collected scenarios iteratively to test saturation of our sample. In consideration of our financial, and personnel resources, we decided on a target sample size of 50 respondents per group. Due to the comparable large sample size for a qualitative study, we decided to collect data via an online survey using the survey provider Prolific. Ethical approval was obtained from the IRB of the first author prior to data collection. We ensured an equal gender balance to acknowledge one diversity dimension. Yet, due to limited sample size and the primary focus of this study to identify risks and propose risk mitigation strategies for the three different stakeholder groups, including citizens, we were not able to specifically sample for other diversity criteria. However, in the context of expert-driven risk assessment practices including different user perspectives, especially public voices, contributes to diversifying voices[19]. We also added two additional open questions that (1) asked respondents to brainstorm mitigation strategies concerning the risks they described in their scenarios, and (2) evaluated a concrete policy proposal of the EU AI Act, namely the transparency obligation (article 50 (4) AI Act). We identified ten impact themes, detailing fifty specific impacts, as well as four mitigation themes with twenty specific mitigation proposals in the scenarios. Our study provided novel insights that demonstrate the value of our approach to enriching risk assessment practices. For instance, we demonstrated that the harms identified in the scenarios placed significant emphasis on societal harms, which are less regulated than individual harms [64]. Furthermore, SSE allowed us to identify and describe moral trade-offs, as well as the rationale and motivations of stakeholders engaging with GenAI in the news ecosystem. The contextual richness of SSE also helped lower the abstraction of potential impacts, grounding them in relatable examples. These descriptions convey meaningful information and provide alternative narratives that could resonate with regulators and assessors more than abstract, quantified numbers. Additionally, by comparing three stakeholder groups, we identified that positionality and expertise influence what is perceived as an important risk and how to mitigate it. These findings strengthen our claim that risk assessment should not rely solely on expert assessors.

We highlight that this is only one example for a particular use of SSE, and we recommend consulting the guidebook and/or our other empirical studies for examples of practical implementations of SSE.

3.3 Data analysis

At its core, the stories that respondents write are qualitative in nature. The stories describe events, actors and their relationships, causal connections, and contextual contingencies. The narrators (respondents) write their stories from their own perspective, which may reflect their lived experiences and cultural embedding—as well as their individual expertise, which brings informed perspectives based on intentional sampling of that expertise in the process. The resulting explorative stories can be used stand-alone as future projections of human-machine interactions and their implications.

[19] We also highlight the need to apply SSE with targeted sampling of vulnerable communities (see our recommendations in Section *Sampling* of the methods guidebook). The referenced study was our first study that provided a proof-of concept of the method and aimed for sampling and comparing three different stakeholder groups while ensuring gender parity.

Table 5 Overview of the SSE approach to address shortcoming of risk and impact assessments

Shortcoming of assessments	SSE approach	Description
Structural		
C1: Lack of democratic accountability	Strengthening of democratic accountability	Inclusion of multiple stakeholders (incl. laypeople) creates more democratic accountability.
C2: Depoliticization of assessments	Politicization of assessments	Acknowledgement of the political character of assessments. Politicization leads to a stronger demand for accountability.
C3: Lack of inclusion	Inclusive sampling	Inclusive sampling leads to the amplification and visibility of different voices. Formerly ignored perceptions are made visible.
C4: Expert-biased	Facilitating diversity	Gauging the subjectivity of many affected populations and acknowledging expertise of different crowds.
C5: Limited guidance for trade-off decisions	Contextual insights to enrich decision basis	The description of different perspectives provides more contextual insights for decision-makers that can further guide trade-off decisions.
Prospection		
C6: Exclusion of unknown unknowns	Identification of unknown unknowns	AI technologies might entail novel risks that are yet unknown in common assessment frameworks. Through the anticipatory approach these can be identified.
C7: Focus on fixing risks	Focus on exploring future scenarios	Identification of and distinction between more and less desirable futures.
Methodological		
C8: Focus on compliance	Focus on societal values	Societal values are the guiding principle. Illuminating what constitutes positive and negative futures guide the way on how society as a whole wants to deal with AI.
C9: Focus on quantification	Keeping qualitative character of impacts	Impacts are identified and qualitatively described. Narratives convey meaning and make compelling arguments that cannot be expressed in numbers.
C10: Lack of contextuality	Assessing contextuality of impacts	Describing the contextuality helps in identifying new risks and in explaining why risks do or don't emerge.
C11: Focus on the technology as primary source of risk	Inclusion of societal, organizational, and human factors as sources of risk	Widening up the view on risk sources to include human or organizationally caused risks.

By deploying SSE with a larger sample size, SSE can be used to develop a socio-technical impact assessment that complements the mostly technical and/or expert-led assessment frameworks. Using qualitative thematic analysis and axial coding [30, 42], users can develop impact classification frameworks that enumerate the impacts outlined in the stories. These impacts can also be analyzed quantitatively (e.g. how many respondents with which demographic characteristics mentioned these factors) – while still retaining qualitative and illustrative examples of these impacts by excerpting material from the scenarios. Depending on the granularity of the impact framework and the sample size, our methods also allows users to statistically trace specific risk perceptions back to sociodemographic characteristics or attitudes of the respondents. It is also possible to sample for different expertise and then compare the emerging impact frameworks with the lived experience and expertise profiles of these groups.

When coupled with a reflection task, even more elaborate questions can be tackled. For instance, the stories can be linked to mitigation strategies that show how anticipated impacts might be addressed. Depending on the research goal and the available resources, users might also want to conduct a participant-review, in which scenario(s) are evaluated by the workshop participants themselves in terms of quality criteria (e.g., plausibility), or implications for individuals, organizations, or society. Such a research stage strengthens the participatory core of the method, and can help center stakeholder voices around their own interpretations, reducing the analytical lens of the user team.

The data analysis approach is dependent on how the users of SSE gather data. It can be qualitative, and incorporate, for instance, in-depth discussion on possible mitigation strategies. But it can also be related to accountability and gather importance ratings on which entities should be responsible for mitigation of impacts. In this way, the strength of the deeply contextualized qualitative description captured by the scenario can also be used as stimulus to collect additional subjective perceptions or quantified ratings to facilitate downstream analysis.

3.4 SSE approach to our critique of risk and impact assessment practices

We developed SSE to address the shortcomings of current risk and impact assessment practices as we elaborated in

Sect. 2. Here, we detail how SSE concretely addresses our critical analysis (for a summary, see Table 5).

3.4.1 SSE approach to address the structural critique

The emerging stories are informative for scientists, policy makers, or auditors, as they contain readily accessible and vivid imaginings of possible future paths [17]. These narratives can be powerful because they transport values, reflect identities, and trigger agency and provide relatable frames, which make a heuristic processing of the information easier for the audience [43]. They can also be seen as a source of counter-power in the form of counter-narratives [43, 57]; this is particularly relevant when considering the inclusion of marginalized groups (Critique 3). Making some scenarios publicly available gives these groups a voice, and can be a vehicle to draw awareness to risk factors that are particularly relevant to these groups but may not yet be on the radar of policy makers and technology developers (Critique 4). Thus, gauging the subjectivity of many affected populations acknowledges the expertism of different crowds [49]. As such, SSE can also strengthen democratic accountability in explicitly advocating for a plurality of voices and empowering groups to call for more accountability of responsible entities (Critique 1). Enriching the current metric-driven impact assessment with qualitative data also gives decision-makers more contextual insights that can help them in making informed decisions, e.g., for deciding on trade-offs (Critique 5). Relying on lived experiences can describe in a real-world context, how impacts actually are perceived by the people affected. This connotation, while subjective, adds an understanding of what an impact actually means, for instance, what consequence a false negative decision from an AI system might have on an individual's life and can politicize the broader discourse (Critique 2).[20]

3.4.2 SSE approach to address the prospection critique

In some use cases of AI it might still be unclear what potential impacts the technology might have. Scenarios as an anticipatory and creative tool can help to detect novel issues and can as such contribute to the identification and description of "unknown unknowns" or "known unknowns" (Critique 6). They can also help to prioritize and identify issues that are, from a societal perspective, more urgent than others, which is particularly relevant e.g. in the context of a general purpose technology like Generative AI where part of the risk is the unforeseeability of how different users in society will employ it and to which ends. Consequently, SSE can be used to identify novel impacts that are currently not present in existing checklist and assessment frameworks.[21] This information can then be used by assessors to add to already known impacts and get a better understanding of the breadth of potential impacts. Also, SSE takes an anticipatory step that is not focused on fixing known risks (Critique 7). Following the research tradition of futures studies [1, 14], some scenarios can also be selected that represent different versions of the future under predefined criteria (e.g. desirable/undesirable) and can, thus, be used to tap into normative prospections (how can a specific target be reached?) or facilitate a discussion of which scenarios or risks are acceptable to the societal stakeholders involved and under which conditions. Presenting these "high-level" scenarios can provide guidance to stakeholders on what actions to take to pursue or avoid a particular future outcome. As such, SSE also responds to calls from scholars for alternative narratives that challenge the techno-optimist mindset that is currently dominating the AI discourse in exploring different futures [4, 57].

3.4.3 SSE approach to address the methodological critique

While current assessments mostly rely on quantification (e.g., metrics) and/or compliance criteria, SSE opens up a debate about how society actually wants to deal with AI technology, i.e. how it relates to societal values (Critique 8). Illuminating what constitutes positive and negative futures guides the way on how society as a whole wants to engage with AI. Therefore, a huge emphasis lies also on keeping the qualitative character of impacts visible as narratives to convey meaning and make compelling arguments that cannot be expressed in numbers (Critique 9). Furthermore, SSE also adds a contextual component to the assessment (Critique 10). SSE can help to understand the conditions under which a technology can have impacts and how they are related to actions of involved entities. Thus, users of SSE can gain a better understanding of why impacts might (not) emerge, and what actions can be taken to mitigate these beforehand. Importantly, this process is not solely focused on the technology itself. It also includes societal, organizational, and human factors as sources of risks and impacts (Critique 11).

[20] As an example, think of the description of a person who got denied financial aid by the state because they were falsely classified. A description of the material and mental consequences this person might have experienced adds a human touch to a metric (e.g., a count of false negatives which is arguable quite abstract).

[21] The efficiency of SSE's ability to identify novel impacts is also related to the sampling approach taken e.g. diversity of perspectives and expertise.

4 Anticipating and meeting objections

While we stress the benefits of SSE to enrich the current landscape of risk and impact assessments, we anticipate some objections from scholars and regulators. In the following, we discuss these and take a prospective look at what needs to be done to meet these objections in a constructive way.

4.1 Objection 1: qualitative nature of evidence

While academics argue that there is a need for more qualitative methods to enrich risk assessment and risk management, industry and standardization bodies prefer to rely on fixed and quantifiable metrics [18, 27]. SSE doesn't produce metrics that can be standardized; hence, critics may object to the method on the grounds of its reliability and representativeness. We reply that risk and impact do not only materialize in metrics, but need socio-cultural contextualization. Scholars have argued that narratives are performative insofar as they have real-world consequences [44]. For instance, utopian futures and the ongoing AI-hype have sparked rapid development of AI technologies and increased funding in tech companies. Especially legal frameworks and political decision-making are dependent on the narratives that are present around a technology. Berresheim argues "[t]he claim to correctness raised by a legal statement is not a claim that this statement is correct in a universal way, but only that it can be reasonably justified with regard to the specific legal framework in which it claims to be correct" [11:77]. In other words, in legal reasoning, which often undergirds regulatory and policy efforts, it is not so much whether or not a claim can be empirically verified, but whether it can be justified, for example with lived experiences. Telling stories of affected people and visualizing concerns and perceptions can inform policy decisions on how technology should be regulated.

As such, the way people act with technologies warrants a more qualitative way of evidence, such as narratives [57]. Especially when working with communities on the margins, the stories emerging from SSE can offer new insights that might challenge dominating narratives, for instance those by Big Tech companies [44]. These counterstories can recenter discourses about AI and contribute to justice for affected people [22].

4.2 Objection 2: resources

The regulatory framework of risk and impact assessment practices is set out to be innovation friendly and cost-effective [27, 52]. But, depending on its purpose, implementing SSE can be costly. Especially when used in a survey design, users of SSE will need to pay participants. In large-scale designs, such as those required to conduct risk assessments, the sample size must be large enough to ensure a holistic assessment of risks. In addition, users must account for the time spent on analyzing the scenarios.

Against this objection, we argue that the cost of risk and impact assessments cannot be the sole deciding factor, particularly for companies whose systems have so much capacity or reach that they are classified as models or services with systemic risks. True, more qualitative methods such as SSE can incur additional costs for technology companies, but the failure to achieve an adequate picture of the potential risks of a technology can have costs for society that by far exceed the initial costs for the company. As AI systems can have detrimental impacts on the public, we argue that safeguarding AI in the public interest (e.g., respecting human rights) should be a priority when implementing systems. This also includes that the use of AI should be justified in the first place [76]. Ultimately, a (political) trade-off decision has to be made on how much needs to be invested in order to protect affected communities. We have shown that producing qualitative evidence with SSE offers valuable insights that current assessment methods are lacking; hence, we argue that the resources for conducting SSE are well spent and a necessity.

4.3 Objection 3: inclusiveness

Socio-technical scholars emphasize the relevance of including all affected people to co-create the future of technology. Different narratives and perceptions can affect how AI technology will be developed [57]. SSE offers an approach to do so, but it also requires its users to make a variety of decisions prior to application. For instance, whose narratives should be assessed with SSE? Highlighting voices from the margins is particularly important given the detrimental effects that generative AI can have on these groups. However, when conducting SSE with marginalized groups, it is important to engage in equitable participatory design, that is, to fully acknowledge the lived experiences and contributions of these communities, understand their historical context and allow for alternative solutions or mitigation strategies [21, 33]. Therefore, when explicitly applying SSE with these groups, it is recommended that multiple researchers and collaborators with different cultural backgrounds are involved in the study design and data analysis. Further, while SSE follows the idea of participatory AI to make the lived experiences of communities visible and to bring them into the broader discussion of AI impact assessment, the approach can still leave communities out: For example, some marginalized groups may not be able (e.g.,

time constraints, or literacy issues) or willing (e.g., refusal) to engage in workshops or surveys [12].

Additionally, on a methodological level, the emerging scenarios are commonly biased toward the instructions that SSE users give to respondents (e.g., what information will be given to describe the technology). Subsequently, also the analysis is tied to the judgment of the research team analyzing the scenarios. This can introduce an expert-bias and a lack of inclusiveness. To counter this, it is recommended to involve people with different backgrounds (e.g., cultural, socio-economic, gender, etc.) in the study design and data analysis, which will also allow for cognitive diversity in the analysis process – or taking a co-creation approach as suggested in the participation action research literature [19, 21]. If that is not possible, users of SSE can also run workshops or surveys to culturally validate their findings.

5 Conclusion

The regulatory landscape, as well as most risk and impact assessment frameworks for AI, are not inclusive, focus on quantifiable risks while ignoring more qualitative risk factors (unknown unknowns), and provide little guidance on how to balance value conflicts. Thus, there is a need to develop approaches to identify (new) risks and impacts of AI technology that take into account the socio-technical relationships users have with the technology. This is all the more urgent given the vast possibilities offered by general-purpose AI tools (e.g. LLMs). How these technologies "behave in the wild" depends largely on how they are used, with what motivations, and with what values; ultimately, all of these factors contribute to emerging risks and impacts that are arguably outside the scope of the "reasonably foreseeable" risks or impacts that developers are required by EU regulations to assess. Moreover, some of these risks may be impossible to quantify due to their qualitative nature. In this context, we have developed the Scenario-Based Sociotechnical Envisioning (SSE) approach.

SSE describes a novel method for envisioning the socio-technical interactions between humans and emerging AI applications that complements current risk assessment approaches. The use of scenario writing as a creative tool allows for the exploration of different future pathways and is particularly suited to tap into the values, concerns, and beliefs of those who will be confronted with the technology. The stories collected as data represent a sharp contrast to current metrics-driven approaches to regulation and assessment. Previous research has shown the power that narratives can have on decision makers. SSE provides a way to make these impacts salient and to amplify the lived experiences that enrich expert-driven evaluation approaches. SSE also enables its users to conduct human-driven risk and impact assessments at a larger scale, which can lead to the identification of risk and impact categories that would otherwise go unrecognized.

With this paper, we make the following contributions: First, we point out critical flaws in existing assessment methods and regulatory approaches. Consequently, we identify the need for a shift in current approaches to assessment approaches. We call on scholars, policy makers, and the technology industry to engage in anticipatory approaches that go beyond compliance and checklist exercises and acknowledge the responsibility they have for impacts that go beyond what is known. General purpose technologies will have a decisive impact on individual and social life, and there is still time to navigate this process responsibly. Second, we propose a concrete alternative empirical approach to address our concerns. By using SSE, practitioners can make assessments more inclusive, more aware of the socio-technical context in which AI systems are deployed, and more prospective in terms of identifying impacts before they occur. We encourage scholars, policy makers and legislators to use SSE as one method in their assessment toolbox. Furthermore, we encourage future research to expand SSE to other, non AI-related, general-purpose technologies (e.g., quantum technology). Third, we anticipate possible objections to our approach and pro-actively discuss how these can be met.

We began this paper by quoting Bernstein: "Risk is a choice, not a fate." Similarly, future pathways are choices. SSE can help illuminate those pathways and make choices that put people and society first.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s43681-026-01084-5.

Author contributions Conceptualization: KK, NH, ND; Theoretical Background: KK, NH; Methodological Approach Development: KK, NH, ND; Discussion & Conclusion: KK, NH, ND; First Draft: KK, NH; Review & Editing: KK, NH, ND.

Funding The funding for this research was provided by UL Research Institutes through the Center for Advancing Safety of Machine Intelligence.

Data availability We uploaded a methods guidebook as supplementary material to this article.

Declarations

Conflict of interest The authors declare no competing interests.

Ethical approval This paper didn't involve human subjects; thus no ethics approval was required. However, we highlight that—if used by scholars—we strongly recommend obtaining ethical approval from the respective institution. All our empirical studies received ethical ap-

proval from the first author's IRB.
Additionally, we discuss potential ethical objections towards our approach in Sect. 4 (Anticipating and Meeting Objections). Yet, we want to add the following: SSE is an approach that responds to calls for new perspectives and methodologies that go beyond quantifiable impact metrics to identify previously overlooked issues and amplify the perspectives of typically underrepresented populations. We hope to spark an academic discussion about the limitations of current impact assessments, and to highlight the need for more qualitative impact assessments for generative AI technologies. We note that we don't aim to replace current impact assessment, but rather to enrich the current landscape with alternative approaches that aim to capture previously overlooked impacts and illuminate the contextual nature of impacts.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

References

1. Amer, M., Daim, T.U., Jetter, A.: A review of scenario planning. Futures **46**, 23–40 (2013). https://doi.org/10.1016/j.futures.2012.10.003
2. Andersen, P.D., Hansen, M., Selin, C.: Stakeholder inclusion in scenario planning—a review of European projects. Technol. Forecast. Soc. Change **169**, 120802 (2021). https://doi.org/10.1016/j.techfore.2021.120802
3. Angwin, J., Larson, J., Mattu, S., Kirchner, L.: Machine bias. In *Ethics of data and analytics*, 254–264. Auerbach Publications (2022)
4. Bareis, J.: The trustification of AI. Disclosing the bridging pillars that tie trust and AI together. Big Data Soc. **11**(2), 20539517241249430 (2024). https://doi.org/10.1177/20539517241249430
5. Barnett, J., Kieslich, K., Sinchai, J., Diakopoulos, N.: Scenarios in Computing Research: A Systematic Review of the Use of Scenario Methods for Exploring the Future of Computing Technologies in Society. *Proc. AAAIACM Conf. AI Ethics Soc.* 8, 1 (2025), 316–329. https://doi.org/10.1609/aies.v8i1.36551
6. Barnett, J., Kieslich, K., Diakopoulos, N.: Simulating Policy Impacts: Developing a Generative Scenario Writing Method to Evaluate the Perceived Effects of Regulation. In: Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society. pp 82–93 (2024) https://doi.org/10.1609/aies.v7i1.31620
7. Barnett, J., Kieslich, K., Diakopoulos, N.: Envisioning Stakeholder-Action Pairs to Mitigate Negative Impacts of AI: A Participatory Approach to Inform Policy Making. In: Proceedings of the 2025 ACM Conference on Fairness, Accountability, and Transparency. ACM, Athens Greece, pp 1424–1449 (2025) https://doi.org/10.1145/3715275.3732096
8. Barreteau, O., Bots, P., Daniell, K., Etienne, M., Perez, P., Barnaud, C., Bazile, D., Becu, N., Castella, J.-C., Daré, W., Trebuil, G.: Participatory approaches. In: Edmonds, B., Meyer, R. (eds.) Simulating Social Complexity, pp. 253–292. Springer International Publishing, Cham (2017). https://doi.org/10.1007/978-3-319-66948-9_12
9. Baum, S.D.: Social choice ethics in artificial intelligence. AI Soc. **35**(1), 165–176 (2020). https://doi.org/10.1007/s00146-017-0760-1
10. Bernstein, P.L.: Against the gods: The remarkable story of risk. Wiley, New York (1996)
11. Berresheim, L.H.M: The right to privacy. Universiteit van Amsterdam. Retrieved from https://pure.uva.nl/ws/files/197406328/Thesis.pdf(2024)
12. Birhane, A., Isaac, W., Prabhakaran, V., Diaz, M., Elish, M., C., Gabriel, L., Mohamed, S.: Power to the People? Opportunities and Challenges for Participatory AI. In: *Equity and Access in Algorithms, Mechanisms, and Optimization*, 2022. ACM, Arlington VA USA, 1–8 https://doi.org/10.1145/3551624.3555290 (2022)
13. Bonaccorsi, A., Apreda, R., Fantoni, G.: Expert biases in technology foresight. Why they are a problem and how to mitigate them. Technol. Forecast. Soc. Change (2020). https://doi.org/10.1016/j.techfore.2019.119855
14. Börjeson, L., Höjer, M., Dreborg, K.-H., Ekvall, T., Finnveden, G.: Scenario types and techniques: towards a user's guide. Futures **38**(7), 723–739 (2006). https://doi.org/10.1016/j.futures.2005.12.002
15. Brey, P.: Ethics of emerging technology. Ethics Technol. Methods Approaches **2017**, 175–191 (2017)
16. Buçinca, Z., Pham, C., M., Jakesch, M., Ribeiro, M, T., Olteanu, A., Amershi, S.: AHA!: Facilitating AI Impact Assessment by Generating Examples of Harms. Retrieved June 8, 2023 from http://arxiv.org/abs/2306.03280 (2023)
17. Burnam-Fink, M.: Creating narrative scenarios: science fiction prototyping at Emerge. Futures **70**, 48–55 (2015). https://doi.org/10.1016/j.futures.2014.12.005
18. Cohen, J.E., Waldman, A.A.: Introduction: Framing Regulatory Managerialism as an Object of Study and Strategic Displacement. Law Contemp Probs **86**(2023)
19. Cornish, F., Breton, N., Moreno-Tabarez, U., Delgado, J., Rua, M., de-Graft Aikins, A., Hodgetts, D.: Participatory action research. Nat. Rev. Methods Primers **3**(1), 34 (2023). https://doi.org/10.1038/s43586-023-00214-1
20. De Gregorio, G., Dunn, P.: The European risk-based approaches: Connecting constitutional dots in the digital age. Common Mark. Law Rev. 59, 2. Retrieved July 25, 2024 https://doi.org/10.54648/COLA2022032(2022)
21. Delgado, F., Yang, S., Madaio, M., Yang, Q.: The Participatory Turn in AI Design: Theoretical Foundations and the Current State of Practice. In: *Equity and Access in Algorithms, Mechanisms, and Optimization*, ACM, Boston MA USA, 1–23. https://doi.org/10.1145/3617694.3623261 2023
22. Dutta, U., Azad, A.K., Hussain, S.M.: Counterstorytelling as epistemic justice: decolonial community-based praxis from the Global South. Am. J. Community Psychol. **69**(1–2), 59–70 (2022). https://doi.org/10.1002/ajcp.12545
23. Broughton Micova, S., Calef, A.: Elements for Effective Systemic Risk Assessment Under the DSA. *SSRN Electron. J.* https://doi.org/10.2139/ssrn.4512640 (2023)
24. Ebers, M.: Truly Risk-Based Regulation of Artificial Intelligence - How to Implement the EU's AI Act. https://doi.org/10.2139/ssrn.4870387 2024
25. Efroni, Z.: The Digital Services Act: risk-based regulation of online platforms. *Internet Policy Rev.* Retrieved from https://policyreview.info/articles/news/digital-services-act-risk-based-regulation-online-platforms/1606 (2021)
26. Ezeani, G., Koene, A., Kumar, R, Santiago, N., Wright, D.: A survey of artificial intelligence risk assessment methodologies. The

global state of play and leading practices identified. EY, Trilateral Research (2021)
27. Floridi, L., Strait, A.: Ethical foresight analysis: What it is and why it is needed? Minds Mach. **30**(1), 77–97 (2020). https://doi.org/10.1007/s11023-020-09521-y
28. Gellert, R.: Fundamental notions. Risk and regulation. In: The Risk-Based Approach to Data Protection, pp. 26–53. Oxford University Press, Oxford (2020)
29. Gillespie, T.: Generative AI and the politics of visibility. Big Data Soc. **11**(2), 20539517241252131 (2024). https://doi.org/10.1177/20539517241252131
30. Glaser, B., Strauss, A.: *Discovery of grounded theory: Strategies for qualitative research*. Routledge (2017)
31. Griffin, R.: What do we talk about when we talk about risk? Risk politics in the EU's Digital Services Act. *DSA Observatory*. Retrieved from https://dsa-observatory.eu/2024/07/31/what-do-we-talk-about-when-we-talk-about-risk-risk-politics-in-the-eus-digital-services-act/ (2024)
32. Guston, D.H.: Understanding 'anticipatory governance.' Soc. Stud. Sci. **44**(2), 218–242 (2013). https://doi.org/10.1177/0306312713508669
33. Harrington, C., Erete, S., Piper, A.M.: Deconstructing community-based collaborative design: Towards more equitable participatory design engagements. Proc. ACM Hum.-Comput. Interact. **3**, 1–25 (2019). https://doi.org/10.1145/3359318
34. Hartmann, D., De Pereira, J.R.L., Streitbörger, C., Berendt, B.: Addressing the regulatory gap: Moving towards an EU AI audit ecosystem beyond the AI Act by including civil society. AI Ethics (4),(2024). https://doi.org/10.1007/s43681-024-00595-3
35. International Organization for Standardization: Risk Management: Risk Assessment Techniques, Edition 2.0 ed. IEC Central Office, Commission Electrotechnique Internationale, Geneva (2019)
36. Kaplan, L.R., Farooque, M., Sarewitz, D., Tomblin, D.: Designing participatory technology assessments: A reflexive method for advancing the public role in science policy decision-making. Technol. Forecast. Soc. Change **171**, 120974 (2021). https://doi.org/10.1016/j.techfore.2021.120974
37. Kieslich K, Diakopoulos N, Helberger N.: Anticipating impacts: using large-scale scenario-writing to explore diverse implications of generative AI in the news environment. AI Ethics. (2024). https://doi.org/10.1007/s43681-024-00497-4
38. Kieslich, K., Helberger, N., Diakopoulos, N.: My Future with My Chatbot: A Scenario-Driven, User-Centric Approach to Anticipating AI Impacts. In The 2024 ACM Conference on Fairness, Accountability, and Transparency, June 03, 2024. ACM, Rio de Janeiro Brazil, 2071–2085 (2024). https://doi.org/10.1145/3630106.3659026
39. Kieslich, K., Helberger, N., Diakopoulos, N.: Make It Sound Like a Lawyer Wrote It": Scenarios of Potential Impacts of Generative AI for Legal Conflict Resolution. (2026) https://doi.org/10.48550/ARXIV.2602.24130
40. Kieslich, K., Morosoli, S., Diakopoulos, N., Helberger, N.: Trade-Offs in Deploying Legal AI: Insights from a Public Opinion Study to Guide AI Risk Management (2026) https://doi.org/10.48550/arXiv.2602.09636
41. Kieslich, K., Schmitt, J. B.: "But what is the alternative?!" - The impact of generative AI on academic knowledge production in times of science under pressure. In: Internet Policy Review. (2025). https://policyreview.info/articles/news/what-alternative-impact-generative-ai-academic-knowledge-production-times-science
42. Lofland, J., Snow, D., Anderson, L., Lofland, L., H.: Analyzing social settings: A guide to qualitative observation and analysis, (2022) Waveland Press.
43. Lowndes, V.: Narrative and storytelling. In: Evidence-Based Policy Making in the Social Sciences (1st ed.), Gerry Stoker and Mark Evans (eds.) 103–122. Bristol University Press. https://doi.org/10.46692/9781447329381.007 (2016)
44. Mager, A., Katzenbach, C.: Future imaginaries in the making and governing of digital technology: Multiple, contested, commodified. New Media Soc. **23**(2), 223–236 (2021). https://doi.org/10.1177/1461444820929321
45. McNamara, A., Smith, J., Murphy-Hill, E.: Does ACM's code of ethics change ethical decision making in software development? In: Proceedings of the 2018 26th ACM Joint Meeting on European Software Engineering Conference and Symposium on the Foundations of Software Engineering, 2018. ACM, Lake Buena Vista FL USA, 729–733. https://doi.org/10.1145/3236024.3264833 2018
46. Meßmer, A. K., Degeling, M.: *Auditing Recommender Systems. Putting the DSA into practice wit a risk-scenario-based approach*. Stiftung Neue Verantwortung. Retrieved from https://www.stiftung-nv.de/sites/default/files/auditing.recommender.systems.pdf2023
47. Metcalf, J., Moss, E., Watkins, E., A., Singh, R., Elish, M., C.: Algorithmic impact assessments and accountability: the co-construction of impacts. Proc. 2021 ACM Conf. Fairness Account. Transpar. (2021). https://doi.org/10.1145/3442188.3445935
48. Mittelstadt, B.: Principles alone cannot guarantee ethical AI. Nat. Mach. Intell. **1**(11), 501–507 (2019). https://doi.org/10.1038/s42256-019-0114-4
49. Moss, E., Watkins, E., Singh, R., Elish, M., C., Metcalf, J.: Assembling Accountability: Algorithmic Impact Assessment for the Public Interest. SSRN Electron. J. (2021). https://doi.org/10.2139/ssrn.3877437
50. National Institute of Standards and Technology. Artificial Intelligence Risk Management Framework: Generative Artificial Intelligence Profile. National Institute of Standards and Technology, Gaithersburg, MDhttps://doi.org/10.6028/NIST.AI.600-1 2024
51. Nikolova, B.: The rise and promise of participatory foresight. Eur. J. Futures Res. (2014). https://doi.org/10.1007/s40309-013-0033-2
52. Orwat, C., Bareis, J., Folberth, A., Jahnel, J., Wadephul, C.: Normative challenges of risk regulation of artificial intelligence. NanoEthics **18**(2), 11 (2024). https://doi.org/10.1007/s11569-024-00454-9
53. Pasquale, F.: Power and Knowledge in Policy Evaluation: From Managing Budgets to Analyzing Scenarios. *Law Contemp. Probl.* 86, 3 (2023)
54. Poortvliet, P.M., Duineveld, M., Purnhagen, K.: Performativity in action: How risk communication interacts in risk regulation. Eur. J. Risk Regul. **7**(1), 213–217 (2016)
55. Ramírez, R., Selin, C.: Plausibility and probability in scenario planning. Foresight **16**(1), 54–74 (2014). https://doi.org/10.1108/FS-08-2012-0061
56. Reuel, A., Ghosh, A., Chim, J., Tran, A., Long, Y., Mickel, J., Gohar, U., Yadav, S., Ammanamanchi, P., S., Allaham, M., Rahmani, H., A., Akhtar, M., Friedrich, F., Scholz, R., Riegler, M., R., Batzner, J., Habba, E., Saxena, A., Kornilova, A., Wei, K., Soni, P., Mathew, Y., Klyman, K., Sania, J., Sahoo, S., Bruvik, O., B., Sadeghi, P., Goswami, S, Wang, A., Jernite, Y., Talat, Z., Biderman, S., Kochenderfer, M., Koyejo, S., Solaiman, I.: Who Evaluates AI's Social Impacts? Mapping Coverage and Gaps in First and Third Party Evaluations. https://doi.org/10.48550/ARXIV.2511.05613 (2025)
57. Sartori, L., Theodorou, A.: A sociotechnical perspective for the future of AI: narratives, inequalities, and human control. Ethics Inf. Technol. (2022). https://doi.org/10.1007/s10676-022-09624-3

58. Schiff, D.S., Kelley, S., Ibáñez, J.C.: The emergence of artificial intelligence ethics auditing. Big Data Soc. **11**(4), 20539517241299732 (2024). https://doi.org/10.1177/20539517241299732
59. Schmitz, A., Mock, M., Görge, R., Cremers, A.B., Poretschkin, M.: A global scale comparison of risk aggregation in AI assessment frameworks. AI Ethics (2),(2024). https://doi.org/10.1007/s43681-024-00479-6
60. Seger, E., Ovadya, A., Garfinkel, B., Siddarth, D., Dafoe, A.: Democratising AI: Multiple Meanings, Goals, and Methods. Retrieved August 10, 2023 from http://arxiv.org/abs/2303.12642 2023
61. Selbst, A., D.: An institutional view of algorithmic impact. Harv. J. Law Technol. 35, 1 (2021)
62. Shelby, R., Rismani, S., Henne, K., Moon, A., Rostamzadeh, N., Nicholas, P., Yilla, N., Gallegos, J., Smart, J., Garcia, E., Virk, G.: Sociotechnical Harms of Algorithmic Systems: Scoping a Taxonomy for Harm Reduction. Retrieved August 1, 2023 from http://arxiv.org/abs/2210.057912023
63. Slattery, P., Saeri, A., K., Grundy, A., C., Graham, J., Noetel, M., Uuk, R., Dao, J., Pour, S., Casper, S., Thompson, N.: The AI Risk Repository: A Comprehensive Meta-Review, Database, and Taxonomy of Risks From Artificial Intelligence. https://doi.org/10.13140/RG.2.2.28850.00968 (2024)
64. Smuha, N.A.: Beyond the individual: governing AI's societal harm. Int. Policy Rev. (2021). https://doi.org/10.14763/2021.3.1574
65. Solaiman, I., Talat, Z., Agnew, W., Ahmad, L., Baker, D., Blodgett, S., L., Daumé, H., Dodge, J., Evans, E., Hooker, S., Jernite, Y., Luccioni, A., S., Lusoli, A., Mitchell, A., Newman, J., Png, M.-T., Strait, A., Vassilev, A.: Evaluating the Social Impact of Generative AI Systems in Systems and Society. Retrieved June 14, 2023 from http://arxiv.org/abs/2306.059492023
66. Soler Garrido, J., De Nigris, S., Bassani, E., Sanchez, I., Evas, T., André, A.-A., Boulangé, T.: Harmonised Standards for the European AI Act. European Commission, Seville (Spain) (2024)
67. Soneryd, L., Amelung, N.: Translating participation: scenario workshops and citizens' juries across situations and contexts. In: Voß, J.-P., Freeman, R. (eds.) Knowing Governance, pp. 155–174. Palgrave Macmillan UK, London (2016). https://doi.org/10.1057/9781137514509_7
68. Stahl, B.C., Antoniou, J., Bhalla, N., Brooks, L., Jansen, P., Lindqvist, B., Kirichenko, A., Marchal, S., Rodrigues, R., Santiago, N., Warso, Z., Wright, D.: A systematic review of artificial intelligence impact assessments. Artif. Intell. Rev. **56**(11), 12799–12831 (2023). https://doi.org/10.1007/s10462-023-10420-8
69. Ullstein, C., Jarvers, S., Hohendanner, M., Papakyriakopoulos, O., Grossklags, J.: Participatory AI and the EU AI Act. Proc. AAAIACM Conf. AI Ethics Soc. 8, 3:2550–2562 (2025). https://doi.org/10.1609/aies.v8i3.36737
70. Van der Heijden, J.: Risk as an approach to regulatory governance: An evidence synthesis and research agenda. Sage Open 11, 3 (2021), 21582440211032202. https://doi.org/10.1177/21582440211032202
71. Waldman, A.E.: Privacy law's false promise. Wash UL Rev. **97**, 773 (2019)
72. Walker, H., Pope, J., Morrison-Saunders, A., Bond, A., Diduck, A.P., Sinclair, A.J., Middel, B., Retief, F.: Identifying and promoting qualitative methods for impact assessment. Impact Assess. Proj. Apprais. **42**(3), 294–305 (2024). https://doi.org/10.1080/14615517.2024.2369454
73. Weidinger, L., Mellor, J., Rauh, Griffin, C., Uesato, J., Huang, P.-S., Cheng, M., Glaese, M., Balle, B., Kasirzadeh, A., Kenton, Z., Brown, S., Hawkins, W., Stepleton, T., Biles, C., Birhane, A., Haas, J., Rimell, L., Hendricks, L., A., Isaac, W., Legassick, S., Irving, G., Gabriel, I.: Ethical and social risks of harm from Language Models. Retrieved from http://arxiv.org/pdf/2112.04359v1 2021
74. Weidinger, L., Rauh, M., Marchal, N., Manzini, A., Hendricks, L., A., Mateos-Garcia, J., Bergman, S., Kay, J., Griffin, C., Bariach, B., Gabriel, I., Rieser, V., Isaac, W.: Sociotechnical Safety Evaluation of Generative AI Systems. Retrieved October 19, 2023 from http://arxiv.org/abs/2310.119862023
75. Wynne, B.: Risk and environment as legitimatory discourses of technology: Reflexivity inside out? Curr. Sociol. **50**(3), 459–477 (2002). https://doi.org/10.1177/0011392102050003010
76. Züger, T., Asghari, H.: AI for the public. How public interest theory shifts the discourse on AI. AI Soc. **38**(2), 815–828 (2023). https://doi.org/10.1007/s00146-022-01480-5

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.