Co-designing an AI Impact Assessment Report Template with AI Practitioners and AI Compliance Experts

Edyta Bogucka[1], **Marios Constantinides**[1], **Sanja Šćepanović**[1], **Daniele Quercia**[1,2]
[1]Nokia Bell Labs, Cambridge, UK
[2]Kings College London, London UK
edyta.bogucka@nokia-bell-labs.com.com, marios.constantinides@nokia-bell-labs.com,
sanja.scepanovic@nokia-bell-labs.com, daniele.quercia@nokia-bell-labs.com

Abstract

In the evolving landscape of AI regulation, it is crucial for companies to conduct impact assessments and document their compliance through comprehensive reports. However, current reports lack grounding in regulations and often focus on specific aspects like privacy in relation to AI systems, without addressing the real-world uses of these systems. Moreover, there is no systematic effort to design and evaluate these reports with both AI practitioners and AI compliance experts. To address this gap, we conducted an iterative co-design process with 14 AI practitioners and 6 AI compliance experts and proposed a template for impact assessment reports grounded in the EU AI Act, NIST's AI Risk Management Framework, and ISO 42001 AI Management System. We evaluated the template by producing an impact assessment report for an AI-based meeting companion at a major tech company. A user study with 8 AI practitioners from the same company and 5 AI compliance experts from industry and academia revealed that our template effectively provides necessary information for impact assessments and documents the broad impacts of AI systems. Participants envisioned using the template not only at the pre-deployment stage for compliance but also as a tool to guide the design stage of AI uses.

Introduction

The potential of AI to bring about significant societal shifts requires a careful examination of the associated risks and benefits (United Nations Educational, Scientific and Cultural Organization 2023). This has prompted scholars to investigate established impact assessments processes in fields such as environmental protection (Selbst 2021; Metcalf et al. 2021) and human rights (Gaumond and Régis 2023; United Nations Educational, Scientific and Cultural Organization 2023), as well as to suggest algorithmic impact assessments (AIAs) as initial self-regulatory approaches for recognizing and mitigating algorithmic harms (Metcalf et al. 2021; Golbin 2021). These assessments, guided by ethical principles such as responsibility and fairness, were recommended to be conducted at different stages, such as system's design, pre-launch, and post-launch, and to be publicly shared as impact statements (Diakopoulos et al. 2016).

As it has become evident that AI system impacts involve not only the underlying algorithms (Watkins et al. 2021; Barnett and Diakopoulos 2022) but also their systemic impacts (Ehsan et al. 2022; Shelby et al. 2023), there has been a growing demand for comprehensive AI impact assessments (AIIAs) (Selbst 2021). In the absence of regulatory requirements and standardized frameworks, initially, AIIAs have been regarded as an extension of existing AI governance processes, such as assessments of privacy (Wright and Wadhwa 2012), data protection (Janssen 2020), and social consequences of AI systems, models and services (Selbst 2021; Raji et al. 2020). However, over time, AIIAs developed into an independent component of AI governance (Skoric 2023). By 2021, a total of 38 distinct impact assessment methods have been introduced (Stahl et al. 2023) serving various goals such as prompting companies to proactively address social consequences (Sherman and Eisenberg 2024), aligning system behavior with organization's responsible AI principles (Raji et al. 2020; Fujitsu Research Center for AI Ethics 2022; Microsoft 2022b), increasing awareness of potential harms among development teams (Johnson and Heidari 2023), and documenting decisions to facilitate learning and future governance development (Selbst 2021).

However, current AIIAs reports are marked by a lack of comprehensiveness and insufficient grounding in established frameworks such as the EU AI Act (European Comission 2024), the NIST's AI Risk Management Framework (AI RMF) (National Institute of Standards and Technology 2023a), and international standards such as the ISO 42001 AI Management System (ISO/IEC 2023), that provide complementary angles for impact assessment: regulatory compliance (EU AI Act), risk management (AI RMF), and organizational best practices (ISO 42001). The lack of alignment with regulations makes it harder for stakeholders to understand the broad societal, environmental, and economic impacts of AI uses, which are essential for risk assessments (European Comission 2024; Hupont et al. 2024; Stahl et al. 2023). Instead, they must repeatedly gather information across disparate assessments of various AI components (e.g., models, systems, and algorithms) rather than focusing on specific uses. Developer teams, in particular, often encounter difficulties in initiating AI impact assessments (Buçinca et al. 2023) and require additional guidance throughout this process (Wang et al. 2023).

Project Website: social-dynamics.net/impact-assessment
Appendix: social-dynamics.net/impact-assessment/appendix.pdf

In response to these issues, our aim is to examine recent regulatory frameworks and existing AIIAs to devise together with AI practitioners and AI compliance experts a reporting template focusing on the intended use of the system. This template aims to provide the necessary information for conducting impact assessments based on regulations across various AI system uses and to be adaptable to different roles. In so doing, we made two contributions:

1. We designed a comprehensive template for an impact assessment report that is grounded in regulations, and did so by conducting two studies. First, we elicited an initial set of design requirements through literature review and semi-structured interviews with 2 AI compliance experts. Second, we co-designed the impact assessment report template with 14 AI practitioners and 6 AI compliance experts, grounded it in regulatory requirements of the EU AI Act, AI RMF, and ISO 42001.
2. We populated the template with a real-world use case of an AI system (a companion app designed to improve the meeting experience) and evaluated it with 8 AI practitioners and 5 AI compliance experts from academia and industry. We compared the final report to a baseline report derived from the typical structure of existing impact assessment reports. Participants found that the final report provided more complete information for impact assessments and addressed all AI system components and impacts more broadly than the baseline. Both the final report and the baseline were rated as highly adaptable to different AI uses and adaptable to different roles.

In light of these results, we discuss the implications of our work for designing impact assessment reports and conducting impact assessment.

Related Work

We surveyed various lines of research that our work draws upon and grouped them into three main areas: *(1)* eliciting requirements for designing a comprehensive template for an impact assessment report that is grounded in regulations, *(2)* co-designing the template, and *(3)* evaluating the template.

Requirements for Designing a Comprehensive Template That Is Grounded in Regulations

AI impact assessments are defined as structured processes for understanding the implications of proposed AI systems (Stahl et al. 2023). They have been proposed as both law-agnostic self-regulating processes (National Institute of Standards and Technology 2023a,b) and official processes in regulations (European Comission 2024; The Danish Institute for Human Rights 2023; The Government of Canada 2023) and organizational standards (ISO/IEC 2023). For example, the NIST AI RMF, a voluntary guidance framework for organizations that design, develop, deploy, or use AI systems, suggests assessing the beneficial and harmful impacts of AI on individuals, groups, organizations, and society. The EU AI Act (European Comission 2024), the first comprehensive AI regulation in the European Union, requires fundamental rights impact assessment for high-risk AI uses deployed by bodies governed by public law, private operators providing public services, and operators assessing creditworthiness and conducting risk assessments for life and health insurance. This impact assessment, conducted before deployment and updated as needed, should cover affected groups, risks of harm, human oversight, and risk management strategies. Similarly, the world's first ISO 42001 standard for AI management system (ISO/IEC 2023) requires organizations to assess the potential consequences of AI systems on individuals and societies, including human rights, legal positions, and life opportunities.

Despite ongoing efforts to standardize AIIAs (ISO/IEC 2025), a consensus on their content and reporting methods for stakeholders and the public is still lacking (Watkins et al. 2021; Sherman and Eisenberg 2024). Stahl et al. (2023) identified and studied 38 AIIAs, showing that they vary widely in topics, focus, and formats, primarily covering human rights (of Europe The Committee of Experts on Internet Intermediares MSI-NET; Mantelero 2022) and ethics (Gebru et al. 2021). The focus of AIIAs varies, assessing the impacts of AI models, systems, or services, typically mentioning use cases only briefly in final reports. However, there is growing agreement that impact assessments should be conducted on specific uses (European Comission 2024; Hupont et al. 2024). In the EU AI Act (European Comission 2024), the specific use determines one of four risk categories (unacceptable, high, limited, minimal), each with its own legal requirements. Similarly, in the NIST framework (National Institute of Standards and Technology 2023a), use-case profiles are required to describe the current state and the desired state of the system, allowing for risk management at different stages of the AI lifecycle.

Collaborative Design of an Impact Assessment Report Template

To achieve consensus on AIIAs, their report templates should adequately document AI uses, reflect the sociotechnical nature of their impacts (Metcalf et al. 2021; National Institute of Standards and Technology 2023b), and accommodate different roles (e.g., developers, researchers, managers, compliance experts) to surface these impacts (Moss et al. 2021; Selbst 2021; The High-Level Expert Group on Artificial Intelligence 2020; Constantinides et al. 2024a). However, existing AIIAs often lack publicly available documentation on their design processes (Johnson and Heidari 2023), raising concerns about whether the current reports' templates meet the diverse needs of stakeholders.

When available, the documentation typically only specifies the nature of the consultations with stakeholders, lacking depth about the types of stakeholders involved and how their input influenced the final design. For example, Microsoft's Responsible AI Impact Assessment template (2022b) was developed following discussions with internal collaborators, while the Canadian ADM template (The Government of Canada 2023) was revised through consultations with relevant stakeholders. In contrast, the Ada Lovelace Institute's Algorithmic Impact Assessment template in healthcare (2022) was developed through interviews with potential users, researchers, and impact assessment experts.

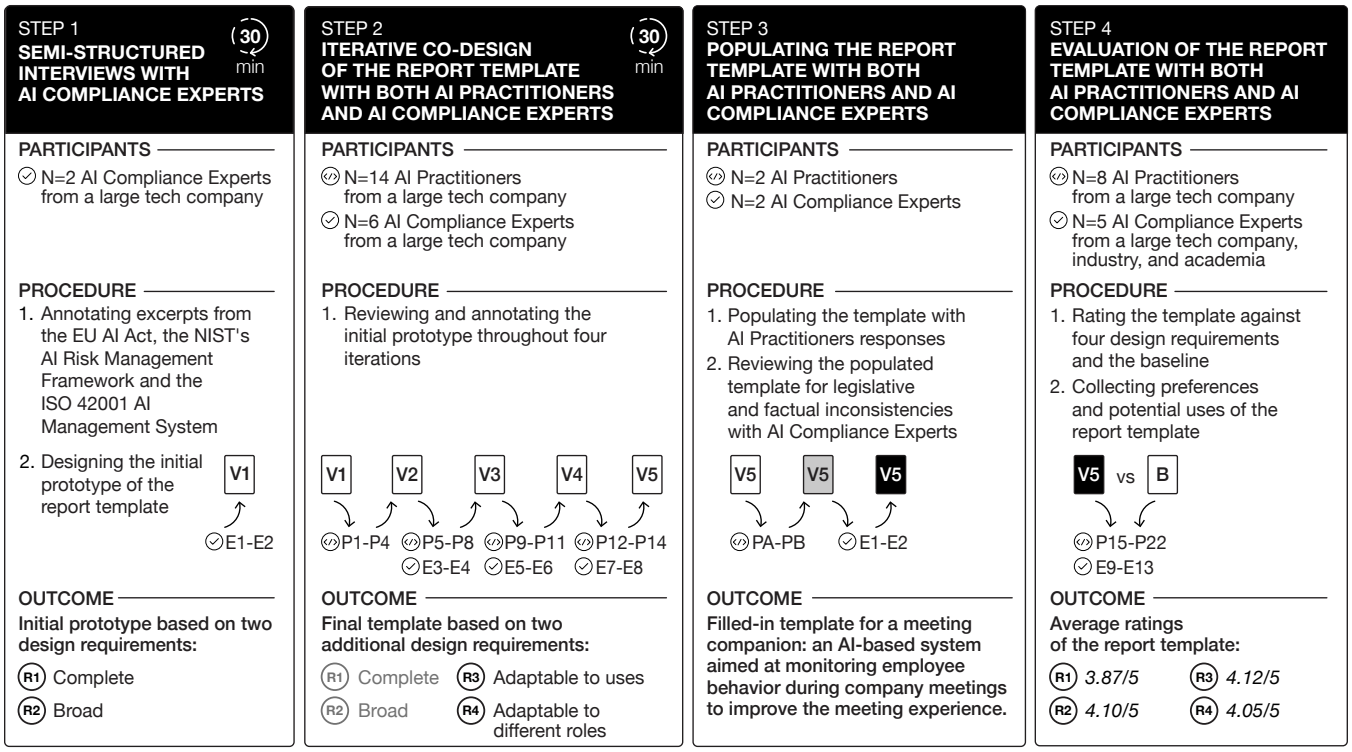

Figure 1: Overview of our four-step method for designing a comprehensive template for an impact assessment report grounded in the EU AI Act, NIST's AI RMF, and ISO 42001. In the first step, we interviewed AI compliance experts to elicit two design requirements and design the initial prototype of the template (V1). In the second step, we engaged both AI practitioners and AI experts to iteratively elicit two additional design requirements and co-design four iterations of the template (V2-V5). In the third step, we populated the final version of the template with AI practitioners' responses. In the fourth step, we evaluated it against our four design requirements and the baseline template identified in the literature review of AIIAs (Stahl et al. 2023).

Researchers highlight the need for more inclusive, value-sensitive co-design processes (Johnson and Heidari 2023; Sadek et al. 2024) that explicitly incorporate compliance experts (Hupont et al. 2024; Constantinides et al. 2024a). For example, OpenLoop's policy prototyping experiment involved developers, researchers, policymakers, and regulators collaborating to create a new policy and documentation of potential harms and mitigations for automated decision-making systems (Gomes de Andrade and Kontschieder 2021). Similarly, the creators of use case cards detailing AI systems' intended uses engaged in co-design workshops with Unified Modeling Language (UML) experts, and EU policy experts (Hupont et al. 2024).

Evaluation of Impact Assessment Report Template

Even less common than documenting template design processes is the public release of their evaluations (Johnson and Heidari 2023), resulting in a lack of evidence for their effectiveness and contributing to the lack of consensus on what constitutes a comprehensive impact assessment (Schiff et al. 2020; Stahl et al. 2023). User studies evaluating three impact assessment report templates, with students role-playing organizational roles, show that template design influences perceptions of AI risks and stakeholder responsibility for potential harm (Johnson and Heidari 2023). The report's content can also affect stakeholders' perceptions and trust (Gaba et al. 2024), especially when it incorporates best practices from information visualization (Franconeri et al. 2021).

More broadly, AIIA evaluations have identified three key pitfalls of current templates. First, AIIAs are fragmented and templates fail to encompass the full range of impacts, including environmental concerns, democratic, safety, human agency, and economic factors (Stahl et al. 2023). Second, AIIAs are not tailored to multiple stakeholders, with templates featuring unfamiliar terms understandable only by those with technical knowledge or terms with multiple legal meanings (Schiff et al. 2020). Third, AIIAs lack supporting questionnaires derived from legal frameworks (Skoric 2023) to ensure templates meet two requirements: completeness of necessary information for risk assessment and broad overage of system components to identify socio-technical risks.

Research Gaps. Current impact assessment reports lack grounding in established regulations and often focus on specific aspects like human rights or data protection, without fully addressing all relevant compliance issues related to specific uses of AI systems. Moreover, there is no systematic effort to design and evaluate these reports with both AI practitioners and AI compliance experts.

Methods

To design a comprehensive template for an impact assessment report that is grounded in regulations, we conducted a series of four studies (Figure 1). These studies included literature review and interviews with 2 AI compliance experts, an iterative co-design process with 14 AI practitioners and 6 AI compliance experts, review of the populated template, and a user study to evaluate the populated template with 8 AI practitioners working in industry research and 5 AI compliance experts from industry and academia.

All sessions were conducted via video conferencing over three months, and were recorded and transcribed. We ensured anonymity of data by excluding personal identifiers, and maintained exclusive access to the data for the research team only. All sessions were approved by our organization.

Eliciting Design Requirements From Literature Review and Semi-structured Interviews With AI Compliance Experts

To elicit requirements for designing a comprehensive template for an impact assessment report that is grounded in regulations, we resorted to previous literature and did so in two steps. First, we reviewed prior academic literature investigating the use of risk assessments, impact reports, and AI documentation for compliance (Selbst 2021; Yeung 2021; Skoric 2023; Gaumond and Régis 2023; National Institute of Standards and Technology 2023a). We also drew upon five existing AI impact assessment templates (Microsoft 2022b; Fujitsu Research Center for AI Ethics 2022; National Institute of Standards and Technology 2023b; Ada Lovelace Institute 2022; The Government of Canada 2023) and value-sensitive games and cards (Ballard, Chappell, and Kennedy 2019; Vakkuri et al. 2021; Sadek et al. 2024). Second, we identified the necessary minimum information to conduct impact assessment in line with three frameworks: the EU AI Act (European Comission 2024), the NIST's AI RMF (National Institute of Standards and Technology 2023a), and the ISO 42001 (ISO/IEC 2023). We selected these frameworks because they have undergone a consensus-driven, transparent process with wide consultations involving various roles such as developers, researchers, managers, and compliance experts. Together, they provide three complementary angles for impact assessment: the EU AI Act offers regulatory compliance, setting a precedent for other AI regulations; the AI RMF provides practical insights for risk management in corporate environments; and ISO 42001 emphasizes organizational best practices. In this step, we started from reviewing the EU AI Act (European Comission 2024) and identifying relevant excerpts that pertain to risk management (Article 9), data governance (Article 10), system monitoring (Articles 14, 15, 72), technical documentation (Articles 11, 12, 18, Annex IV), system transparency (Articles 13, 50), impact assessment (Article 27), and provider obligations (Articles 16, 17, 53, 55). We then reviewed the NIST's AI RMF (National Institute of Standards and Technology 2023a) and its Algorithmic Impact Assessment template (National Institute of Standards and Technology 2023b). From these two sources, we identified relevant excerpts related to mapping the risks of a system's use, minimizing its negative impacts while maximizing its benefits, and communicating these aspects to various stakeholders. Finally, from the ISO 42001, we identified excerpts on documenting resources for AI systems (data, tools, systems, computing, and human resources), assessing impacts on individuals, groups of individuals, and society, and reporting AI system details to stakeholders.

Having these excerpts from previous literature at hand, we conducted 30-minute semi-structured interviews with 2 AI compliance experts to jointly annotate the collected excerpts (i.e., the EU AI Act, the NIST's AI RMF, and the ISO 42001) with high-level topics. We then jointly organized the topics in a thematic and hierarchical manner, resulting in the creation of the first version of the template (Figure 1, Step 1, V1). Through this joint exercise, we surfaced two main design requirements for an impact assessment report template:

R1: Complete. The report template should provide the necessary minimum information to complete an impact assessment in line with selected frameworks (Selbst 2021; Skoric 2023). The scope of this information must be carefully selected to enable consistent reporting, evaluation, and comparison of various AI systems.

R2: Broad. The report template should address all AI system components to identify, evaluate, and mitigate broad socio-technical impacts associated with this system's use. This approach allows for, firstly, identifying specific impacts related to individual system components, such as data sources or algorithms; and secondly, systematically broadening the scope to include other risks.

Co-designing the Impact Assessment Report Template With AI Practitioners and AI Compliance Experts Through an Iterative Process

After designing the first version of the template, we conducted a series of individual, 30-minute co-design sessions with 14 AI practitioners and 6 AI compliance experts (Figure 1, Step 2). These sessions aimed to ensure the template provided sufficient information for impact assessment per the EU AI Act, NIST's AI RMF, and ISO 42001, and was usable by various roles in the assessment process.

Participants. We aimed to achieve a diverse participant sample using snowball sampling, a method where existing study participants recruit future participants from among their acquaintances. We began by identifying 6 initial participants through an internal mailing list at a large tech company, seeking individuals familiar with the EU AI Act and actively involved in developing or evaluating at least one ongoing AI project during the time of the planned interviews. These participants were then asked to refer additional participants from their networks, thus expanding the sample size through successive referrals. We recruited a total of 20 participants (13 male, 7 female, with a median age of 34 years old) representing a variety of AI practitioners such as researchers (9), engineers (2), managers (3), designers (1), and AI compliance experts (6), and potential end users of the template. Their expertise span across various areas including generative AI, deep learning, AI standardization, and human rights. AI practitioners offered practi-

cal examples on navigating regulatory challenges in their projects while maintaining innovation. Compliance experts contributed their knowledge of existing regulations, insights from ongoing regulatory works, and lessons learned from reviewing AI system uses. Table 1 provides an overview of the participants' demographics and relevant experience.

Template version	ID	Gender	Age	Education	Expertise	Yrs in AI	Role
V1	E1	F	39	M.L.	AI governance	4	compliance expert
	E2	M	43	MSc	AI procurement	5	compliance expert
	P1	F	33	PhD	deep learning	5	researcher
V2	P2	M	30	PhD	machine learning	5	researcher
	P3	M	33	PhD	machine learning	5	researcher
	P4	M	32	PhD	machine learning, NLP	7	researcher
	P5	M	35	PhD	embedded machine learning	5	manager
	P6	M	28	PhD	mobile sensing	6	researcher
	P7	M	37	PhD	NLP	10	researcher
V3	P8	F	33	PhD	machine learning	2	engineer
	E3	F	47	MA	human rights impact assessment	1	compliance expert
	E4	F	33	MSc	standardization	7	compliance expert
	P9	M	59	BSc	generative AI	2	designer
	P10	M	32	MSc	machine learning, NLP	6	researcher
	P11	M	27	MSc	machine learning	4	researcher
V3	E5	M	43	M.L.	licensing	1	compliance expert
	E6	F	34	PhD	standardization	6	compliance expert
	P12	M	43	PhD	machine learning, computer vision	5	researcher
V5	P13	M	40	PhD	computer vision, bioinformatics	5	researcher
	P14	F	26	MSc	AI UX design	3	designer
	E7	F	42	M.L	AI case approval	3	compliance expert
	E8	M	50	MA	AI procurement	5	compliance expert

Table 1: Demographics of AI compliance experts (E1-E2) who participated in semi-structured interviews and demographics of AI practitioners (P1-P14) and AI compliance experts (E3-E8) who participated in co-design sessions of the impact assessment report template, including gender (F: Female, M: Male). AI compliance experts highlighted that the current templates lacks scaffolding elements understandable to various roles and often miss the effects on individuals outside the direct users or subjects. AI practitioners recognized the need for these elements but struggle to define impacts on stakeholders beyond direct users, such as organizations and societies, which are frequently referenced in regulations.

Procedure. Before the session, we emailed participants with a demographic survey, a brief description of the session goals, the current iteration of the template, and summaries of the EU AI Act, NIST AI RMF, and ISO 42001 (Appendix A). We asked participants to familiarize themselves with these materials and optionally provide preliminary feedback by reviewing and annotating the template. This ensured participants understood the key regulatory frameworks, leading to more informed discussions and template revisions.

During the session, we first asked participants to evaluate the overall structure and design of the template. We then encouraged them to provide detailed comments on each section, focusing on any critical information they felt was missing or underrepresented. For compliance experts, we specifically sought to identify potential discrepancies in the template for missing information required for impact assessment in current AI regulations. We then asked all participants if they would use the template for their AI systems, and if so, how they would populate it. Finally, we discussed what features could be added to make the template more accessible to users with different roles.

We conducted individual sessions with 3 to 6 participants each, based on research suggesting that the first five participants can identify over 80% of usability issues (Nielsen and Landauer 1993). After each session, we summarized the prevalent issues and revised the template accordingly. The revised template was then tested in co-design sessions with the next batch of participants. Initially, we started the sessions exclusively with AI practitioners and included both AI practitioners and compliance experts after the first iteration.

By the time we produced the fifth version of the template (see iteration descriptions in Appendix B), we had gathered enough information from the co-design activities with AI practitioners and AI compliance experts that no new significant insights or usability issues were emerging. This indicates that the template had been refined to the point where additional testing was unlikely to yield further meaningful improvements. Following the approach of prior co-design studies (Hupont et al. 2024) and cyclical action research (Vakkuri et al. 2021), we concluded the template iterations at this stage and moved on to conducting a user study.

Analysis. After each co-design session, two authors conducted an inductive thematic analysis (bottom-up) of the session's transcripts, following established coding methodologies (Saldaña 2015; Miles and Huberman 1994; McDonald, Schoenebeck, and Forte 2019). The authors used the Figma platform (Figma 2024) to capture participants' feedback in sticky notes, and collaboratively created themes based on these notes. They discussed and resolved any disagreements that arose during the analysis to derive a list of usability issues to be addressed in the next iteration of the template.

These co-design sessions with AI practitioners and compliance experts surfaced two additional design requirements, which led to the design of a template in which no further refinements were deemed necessary (Figure 1, Step 2, V5):

R3: Adaptable to uses. The report template should apply to a wide range of AI systems, their possible uses, and their application domains. It should also offer the flexibility to modify, or omit sections that may not apply to specific systems or their development stages, allowing for a more accurate and relevant evaluation.

R4: Adaptable to different roles. The report template should have a clear, self-contained structure with understandable sections, headings and subheadings. This allows users with different roles and levels of expertise with AI systems to effectively use it and participate in the impact assessment process.

Evaluating the Impact Assessment Report Template

To then evaluate the final template, we first populated it with a real-world use of an AI system that was developed in the same large tech company and then conducted a user study to evaluate the populated report produced for this system's use with 8 AI practitioners and 5 AI compliance experts.

Populating the Impact Assessment Report Template. To populate the template, we employed a three-step semi-automatic method, which included soliciting AI practitioners responses and reviewing these responses with AI compliance experts (Figure 1, Step 3).

In the first step, we compiled a list of statements to gather responses from AI practitioners. To do so, we sourced statements from responsible AI guidelines (The High-Level Expert Group on Artificial Intelligence 2020; Constantinides et al. 2024a), documentation standards (Selbst 2021; Gebru et al. 2021; Holland et al. 2020; Bender and Friedman 2018; Mitchell et al. 2019; Sokol and Flach 2020; Raji et al. 2020), checklists and impact assessment questionnaires (Madaio et al. 2020; Golpayegani, Pandit, and Lewis 2023; Skoric 2023; National Institute of Standards and Technology 2023b). Next, we reviewed these statements, linking them to the relevant excerpts from the EU AI Act, the NIST AI RMF, and the ISO 42001, and grouping similar ones together. This process resulted in a list of 32 statements grounded in regulations and best responsible AI practices, designed to systematically gather information about the system's use, components, and data, team involvement, and the associated risks, mitigations, and benefits (Appendix C). They also directly map to the sections of our template.

In the second step, we reached out to 2 AI practitioners—a researcher and a designer—who had contributed to the development of an AI-based meeting companion app. We asked them to provide responses to these 32 statements.

In the third step, two authors manually parsed the practitioners' responses, placing them in the template to complete the report (Figure 2). We then consulted 2 AI compliance experts, whom we had interviewed during the requirement elicitation phase, to review the report. They marked any legislative inconsistencies in the sections reporting risks and any factual inconsistencies in the sections reporting mitigation strategies and benefits. Upon inspecting the marked report, we found that the experts agreed on their assessments, and the report did not contain any inconsistencies.

In addition to the report, we included a baseline condition to compare against our final template. For the baseline, we derived a generic impact assessment template for AI system use, based on the typical structure of existing impact assessment reports identified by Stahl et al. (2023). This generic template consisted of three sections: the first, "Intended Use", included a description of the use (covering the technology and application area) and stakeholders affected by the use; the second,"Risks", covered human rights risks, ethical risks, data protection and privacy risks, safety risks, security risks, and environmental risks; and the third, "Mitigation measures", detailed technical and organizational measures for mitigating risks.

The baseline content was the same as the final report except for describing intended use through full sentences, reorganizing risks and mitigations under different categories, and lacking the enumeration of benefits. Figures 5 and Figure 6 in Appendix D present the populated final impact assessment report and the baseline used in the user study.

Conducting a User Study for Evaluating the Populated Impact Assessment Report. We conducted a user study in the form of semi-structured 30-minute interviews with 8 AI practitioners from the same large tech company and 5 AI compliance experts from both industry and academia (Figure 1, Step 4). Specifically, our evaluation ought to answer four questions:

Q1: To what extent does the report's content contain all necessary minimum information for conducting impact assessments in line with AI regulations (R1)?

Q2: To what extent does the report's content address all AI system components to identify, evaluate, and mitigate socio-technical risks of the system's use (R2)?

Q3: To what extent is the report's template adaptable to different AI system uses (R3)?

Q4: To what extent is the report's template adaptable to different roles (R4)?

Participants. We recruited 8 new AI practitioners working in the tech industry, with various roles and expertise, including computer vision engineers and researchers in deep learning, different from those who participated in the co-design sessions. Additionally, 5 new AI compliance experts took part in the study, who work in industry (3) and academia (2). Table 2 summarizes participants' demographics.

Procedure. Before conducting the interviews, we sent an email to all participants. This email included a brief description of the study and a short survey focusing on demographics. The survey solicited details about the participants' age, area of expertise, professional role, and years of experience in developing AI systems (Appendix E). Furthermore, we attached brief overviews of the EU AI Act, the NIST AI RMF and the ISO 42001 (Appendix A). We requested the participants familiarize themselves with these documents to ensure a thorough evaluation of the template for alignment and more focused discussions during the interviews. Our organization gave its approval for the study. Again, we maintained data anonymity, removed personal identifiers, and restricted data access to the research team only.

During the interviews, we presented participants with both the populated final impact assessment report (Figure 5, Appendix C) and the baseline (Figure 6, Appendix C) for the AI-based meeting companion. Participants read each report for up to 15 minutes, alternating between them to avoid learning effects. After each reading, participants rated 12

statements (Figure 3 S1-S12) related to the four design requirements identified from the co-design sessions, using a Likert scale from 1 (strong disagreement) to 5 (strong agreement). We then asked participants about their preferences, dislikes, and how they would adapt both report templates to their own work. Finally, two authors transcribed the interviews and conducted an inductive thematic analysis.

Analysis. First, we computed the average Likert scale score for each statement (Figure 3) for both the final report and the baseline. Second, two authors conducted an inductive thematic analysis (bottom-up) of the interview transcripts, following established coding methodologies (Saldaña 2015; Miles and Huberman 1994; McDonald, Schoenebeck, and Forte 2019). The transcripts covered how the report's content supports impact assessment, how usable each template is, and any other preferences or dislikes.

The authors used the Figma platform (Figma 2024) to capture participants' answers in sticky notes, and collaboratively created themes based on these notes. They discussed and resolved any disagreements that arose during the analysis process. Each theme included quotes from at least two participants, signifying that data saturation was reached (Guest, Bunce, and Johnson 2006).

The Impact Assessment Report Template

The final template organizes information into five modular sections (Figure 2A-E), allowing to add, remove or rearrange sections as needed (addressing R3):

Section 1: Information on the System's Use and Teams. The section begins with a clear description of the system's intended use (Figure 2A), including its intended purpose, the technical capabilities that enable this use, the area or sector chosen for its application (including its geographic and temporal extent), any natural persons and groups likely to be affected by this use, and any natural or legal persons who have authority over it. This description mirrors the five risk assessment components identified by Golpayegani, Pandit, and Lewis (2023): *purpose*, *capability*, *domain*, *user*, and *subject* that provide information necessary for conducting the risk assessment based on the EU AI Act (addressing R1). Users could specify these components using external dictionaries, such as the Vocabulary of AI Risks (Golpaygani, Pandit, and Lewis 2024), which contains pre-defined descriptions of these concepts directly sourced from the Act.

Following that, the section contains two subsections related to system's underlying components (including its machine learning models, third-party technologies, and mechanisms for system's monitoring) and data (addressing R2). Afterwards, it contains a subsection related to system's evaluation at different stages of the system's lifecycle (e.g., development, deployment, and use as per AI RMF) including description of the system's performance, accuracy, reliability, and limitations (e.g., the reasonably foreseeable misuse). The section concludes by disclosing the diversity of the teams overseeing the system's use at different stages.

Additional subsections could be included to address specific regulatory requirements of different regions or to provide more contextual information about the system's use,

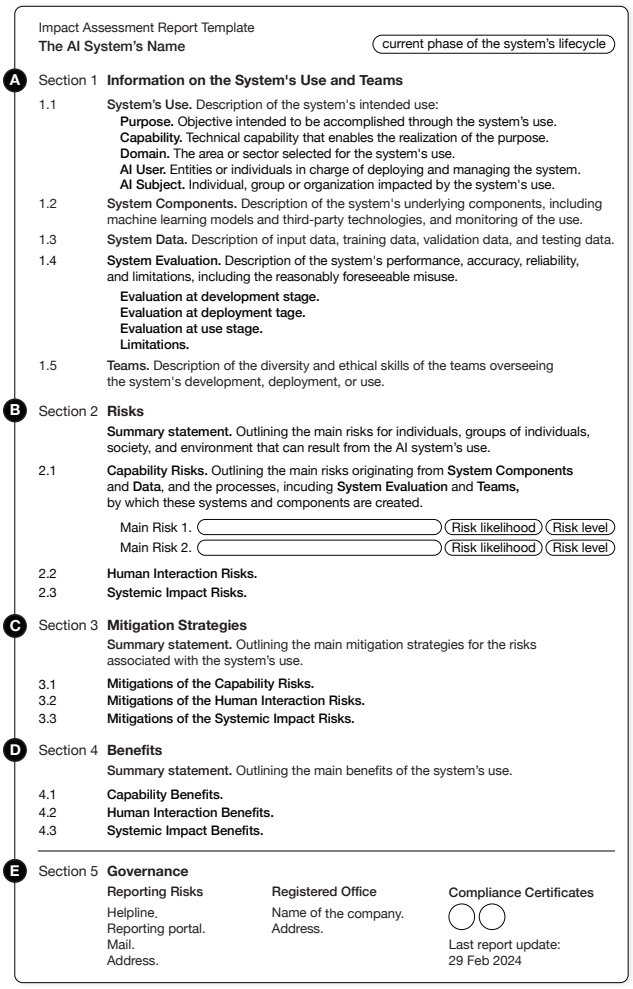

Figure 2: The final template for an Impact Assessment Report. Section 1 provides information on the system's use, components, data, evaluation, and teams; Section 2 lists potential risks; Section 3 lists mitigation strategies; Section 4 outlines the anticipated benefits from the system's use; and Section 5 outlines information about reporting mechanisms and who is responsible for the governance of the use.

such as previous experiences with the deployment of similar systems (addressing R3 and R4).

Section 2: Risks. This section lists potential risks (Figure 2B) associated with putting the system into use (addressing R2). To begin with, it presents a brief summary statement outlining the main risks that may undermine safety, rule of law, fundamental rights, health, environment, and democracy. It then lists all the specific risks of system's use grouped into risks stemming from the *system's capability* (i.e., the technical components of the AI system), *human interaction* (i.e., experiences of people interacting with the AI system), and *systemic impact* (i.e., societal, economic, and environmental impacts of the AI system's use). This grouping aligns with a three-layered framework for evaluating sociotechnical harms by Google DeepMind (Weidinger et al.

2023). It was selected because AI practitioners found it more approachable (addressing R4), based on feedback indicating difficulties in envisioning realistic risks across stakeholders (i.e., organizations, individuals, groups of individuals, and societies) listed in the EU AI Act, AI RMF, and ISO 42001. For each risk listed in the section, we have added to the template two placeholders specifically for 'Risk likelihood' and 'Risk level'. This encourages users of the template to categorize each risk's impact by estimating its probability and potential impact as high, medium, low, respectively, and record these assessments in the provided placeholders. This facilitates impact assessment per ISO 42001 (addressing R1) and allows estimation of the overall risk level score of the use as a combination of the risk likelihoods and risk levels.

Section 3: Mitigation Strategies. This section lists mitigation strategies (Figure 2C) designed to address and minimize the previously identified risks of system's use (addressing R2). The section provides a brief summary statement highlighting the main mitigation strategies for the system's use, and then a list of mitigation strategies for each risk group identified in Section 2 (Figure 2B).

Section 4: Benefits. This section outlines the anticipated benefits and positive impacts resulting from the system's use (Figure 2D). First, the section provides a brief summary statement covering direct and indirect advantages for users, organizations, and society, emphasizing potential long-term positive effects. It then lists all the benefits of the system's use grouped into themes, such as improving access to quality education. By doing so, the section balances the discussion of risks and fosters a comprehensive understanding of the overall impact of the system's use.

Section 5: Governance. This final section (Figure 2E) outlines information about reporting mechanisms (e.g., dedicated email, phone number, the registered office) and who is responsible for the governance of the use (e.g., European conformity marking and other compliance certifications issued so far for the use; addressing R2).

The template provides clear section and subsection titles to help users quickly identify the content they are looking for (addressing R4). Having public interpretability in view, we also applied best practices from Information Visualization such as consistent formatting style throughout the template, including headings, sans-serif fonts, and ample white spacing (Franconeri et al. 2021).

The resulting template aligns with regulatory guidelines (European Comission 2024; National Institute of Standards and Technology 2023a) while offering an alternative approach to current reporting practices (Stahl et al. 2023). Instead of broadly considering the AI models or systems, we have focused our template on the system's intended use for which prevalent documentation is typically limited to a brief textual description without a standardized format (World Economic Forum 2022).

Group	ID	Gender	Age	Education	Expertise	Yrs in AI	Role
AI Practitioners	P1	F	33	PhD	deep learning	5	researcher
	P2	M	31	PhD	machine learning	5	researcher
	P3	M	33	PhD	machine learning	5	researcher
	P4	M	32	PhD	machine learning	7	researcher
	P5	M	36	PhD	embedded machine learning	5	manager
	P6	M	27	PhD	mobile sensing	6	researcher
	P7	M	37	PhD	NLP	10	researcher
	P8	F	33	PhD	machine learning	2	engineer
AI Compliance Experts	E9	F	47	MA	human rights law	1	manager
	E10	F	33	MSc	standardization	7	executive advisor
	E11*	F	25	PhD	impact assessment	5	researcher
	E12*	F	32	MSc	policy making, content moderation	2	researcher
	E13	M	44	MSc	audit, risk assessment	2	manager

Table 2: Demographics of AI practitioners (P15-P22) and AI compliance experts (E9-E13) from industry and academia (marked with *), including gender (F: Female, M: Male), who participated in the template evaluation user study.

User Study of the Impact Assessment Report Template

We report the quantitative results with the qualitative results, which add the participants' perspectives on what could be improved in the next co-design iterations.

Quantitative results. Participants found that the final report provided more complete information for conducting impact assessments than the baseline across all three frameworks (Figure 3, R1). The report received the highest ratings for alignment with the NIST AI RMF and ISO 42001, and the lowest for the EU AI Act. Participants also found that the final report provided a broader scope in addressing all AI system components and the impacts of their use compared to the baseline (Figure 3, R2). In terms of adaptability to different uses (Figure 3, R3), participants found the final report slightly more applicable to a wide range of AI systems than the baseline. Regarding adaptability to different roles (Figure 3, R4), all participants rated both the final template and the baseline as very straightforward and intuitive to complete, easy to navigate with clear headings and subheadings, and suitable for users with varying roles.

Qualitative results. AI practitioners found the final template and its sections helpful. For example, P6 stated that *"I really like it [the template] covers various aspects of the system like its components and data"*. Practitioners also praised the template's simplicity. P3 stressed they *"like its simplicity as it helps me deal with the complexity of the AI system"*, for example by scoping well the intended use of the system, which P7 summarized as *"you see it [five-format component], and it sticks with you"*. All 8 AI practitioners praised the idea of grounding the report's content in the EU AI Act articles. For example, P8 stated that *"these risks are citing the specific sections of the Act [...] It's definitely enough in-*

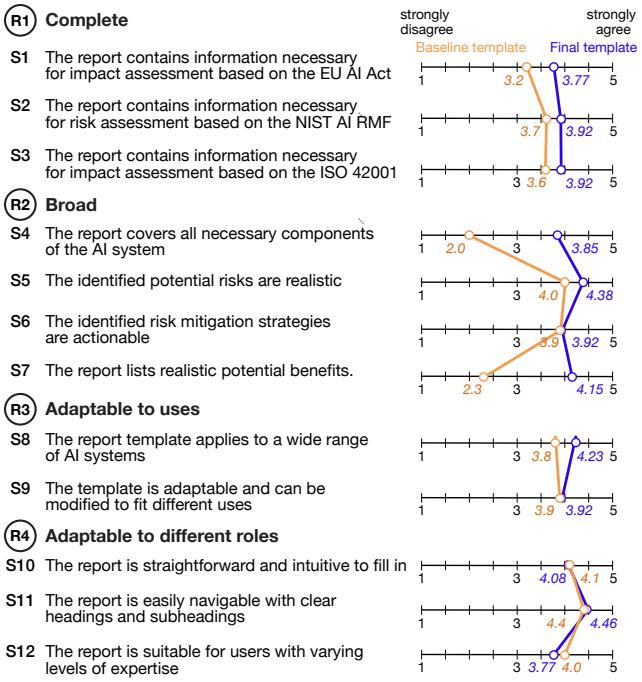

Figure 3: Ratings on twelve statements regarding the four requirements – R1: Complete, R2: Broad, R3: Adaptable to uses, and R4: Adaptable to different roles – for both the baseline template and the final template. Participants found that the final report provided more complete information for impact assessments (R1) and addressed all AI system components and impacts more broadly than the baseline (R2).

formation to convince me that it [the meeting companion] is high risk". This was more evident when our participants commented about the risks section of the report. Specifically, P1 mentioned that *"the risk section is a good start for understanding how it [the meeting companion] can be misused"*. However, AI practitioners also saw a number of improvements. Six respondents highlighted the need for automated tools to populate the report. E10 stated that *"it is s very much about discovering risks and I strongly believe that we should use automated tools and GenAI to aid in that process to prompt with the right kind of responses"*.

AI compliance experts have confirmed that the final template effectively addresses two aspects of integrating and managing AI systems within organizational processes. First, the final template promotes the integration of AI-driven processes with existing, well-established organizational procedures. As stated by E13, *"the approach with the evaluation during development, deployment and use is quite familiar to anyone really involved in product development. This feedback must have been given by developers and I agree - completing the template should not be an additional burdensome process, but something that is integrated into what we are already used to doing"*. Second, by advocating for the blending of multiple taxonomies and frameworks, the template ensures a thorough and comprehensive approach to risk assessment. This method helps prevent the oversight of potential risks by discouraging a narrow focus on single frameworks. As summarized by E10, *"I see your intention not to focus on only one risk taxonomy. If you want to make sure that you are covering everything, it is very good to start with one taxonomy and blend in taxonomies from other frameworks. There are no clear boundaries between risks; they overlap and create different categories across frameworks. In reality, these frameworks interact and mix to some extent."* To see how, E12 explained how the same term "systemic risk" has multiple meaning based on the regulation: *"very large online platforms and search engines, e.g., providing AI-based recommender systems, need to perform a systemic risk assessment under the Digital Services Act, which specifies four categories of such risks [the dissemination of illegal content, negative effects for the fundamental rights; negative effects for the civic discourse, electoral processes, and public security; and negative effects in relation to gender-based violence, the protection of public health and minors and serious negative consequences to the person's physical and mental well-being]. However, systemic risk in the EU AI Act relates also to the general purpose AI models and refers to negative effects on public health, safety, public security, fundamental rights, or the society"*.

Both AI practitioners and AI compliance experts stressed the importance of the inclusion of expert oversight to elevate the quality and reliability of the assessments carried out jointly by these stakeholders. E9, AI compliance expert, echoed AI practitioners' perspectives: *"the biggest challenge is ensuring teams have the necessary skills and knowledge to complete the assessment. Good written guidelines and preferably some expert guidance and oversight would lead to better results"*. Participants also envisioned using the template not only at the pre-deployment stage right before the compliance but also as a tool to guide the design stage of AI uses. P7, researcher, noted that the template *"suits all roles and summarizes the design decisions made so far. It has a different structure than typical compliance reports, where each section has a topic-specific focus and is relevant only for certain experts"*. E12, a compliance expert, suggested that *"the template can be completed not only at the development stage but also every time the socio-technical context changes"*. To illustrate this point, they gave the example of the national system that evaluates the chances of specific groups in the labor market, which was not usable during the COVID-19 pandemic, and for which similar risks had not been foreseen in the limitations of the system.

Discussion

We contextualize our template within prior literature, then discuss its implications, limitations, and future research.

Differences from Existing Templates

We compared our template with the Algorithmic Impact Assessment template from NIST (National Institute of Standards and Technology 2023b), Microsoft's Responsible AI Impact Assessment Template (Microsoft 2022b), Credo AI's Standardized Risk Profile (Sherman and Eisenberg 2024), and the Algorithmic Impact Assessment from the Ada

Lovelace Institute (Ada Lovelace Institute 2022), and identified four key differences. First, our template enhances cross-company comparisons by aligning closely with legislation rather than relying on framework-specific biases or company-specific risks. Second, unlike existing templates which focus heavily on risks, ours equally emphasizes both risks and benefits. Third, there is variation in the extent of guidance provided for completing the templates. NIST's (National Institute of Standards and Technology 2023b), Microsoft's (Microsoft 2022a), and the Ada Lovelace Institute's (Ada Lovelace Institute 2022) offer guidebooks on how to produce a report. Rather than leaving stakeholders with a blank report and a guide, we offer a guidebook through our 32 simple statements that alleviate potential anxiety associated with an empty page and facilitate the identification of risks and benefits (Bogucka et al. 2024b). Finally, whereas other templates have separate sections for legal and compliance issues, ours integrates these aspects throughout.

Theoretical and Practical Implications

Embedding AI governance in impact assessment reports. We introduced a standardized template that not only facilitates risk assessment in accordance with the EU AI Act, the NIST's AI RMF, the ISO 42001 but also serves as a model for integrating regulatory considerations into AIIAs. Our work contributes to co-designing and validating improved AIIAs and the processes and workflows surrounding them (Skoric 2023). This includes developing new tools and methodologies that are practical and adaptable to enhance the quality and effectiveness of impact assessments in the AI context. They provide a comprehensive and practical framework for companies to navigate the complexities of AI regulation, ensuring both compliance and ethical responsibility in AI development and implementation.

Facilitating contextual evaluation. Finding the right balance between making impact assessment template general enough to apply to various AI systems and specific enough to provide meaningful assessments for each unique system's use is a significant challenge (Stahl et al. 2023). We partly addressed this challenge by adding specific subsections into the template. These subsections cover the five components comprehensively describing the system's use, the three key stakeholders, and the three stages of the system's lifecycle. By doing so, we assist stakeholders in systematically documenting the intended use for which the system was built, ensuring that it meets performance and safety criteria across different situations (Johnson and Heidari 2023).

Improving stakeholder engagement. Our report template can be utilized by individual team members or as a group. For teams that are new to impact assessments, working through this template can be an educational experience that helps in building understanding and skills related to responsible AI practices. Future work on improving stakeholder engagement should explore alternative interactive tools that effectively balance various methods of eliciting system information and envisioning system impacts, such as divergent and convergent thinking styles (Selbst 2021).

Limitations and Future Work

Generalizability. Our results are based on a pool of study participants who were familiar with the EU AI Act, the NIST framework and ISO standards. Including a broader range of roles beyond managers, designers, and researchers may yield different results. Future studies should include participants with different levels of knowledge about AI regulations. They should also evaluate how effectively impact assessment reports explain AI risks and benefits to everyone and find ways to help people understand important AI regulations and laws.

Propagating biases in impact assessment. AIIAs, like other responsible AI tools, have inherent biases from their design choices, such as excluding potential users (Moss et al. 2021). The quality of our report template depends on the accuracy and completeness of user-provided information. Using biased or incomplete data can lead to incorrect assessments, making problems seem smaller than they really are because people might be afraid to report negative impacts. To fix this, future research should include external perspectives by involving independent experts and marginalized groups, and regularly checking the results with new team members (Raji et al. 2020; Ada Lovelace Institute 2022).

Automated tools pre-populating the template. Automated tools can help make gathering data for templates easier and less prone to mistakes. Large Language Models (LLMs) can help fill out impact assessment reports by generating lists of AI users and subjects (Buçinca et al. 2023), identifying intended and unintended uses (Wang et al. 2024; Herdel et al. 2024), and listing potential risks and benefits (Constantinides et al. 2024b; De Miguel Velazquez et al. 2024). Bogucka et al. (2024a) have recently proposed a semi-automatic system that collects input from stakeholders about an AI system's use , uses LLMs to find additional risks, mitigation strategies, and benefits, and pre-fills reports for experts to review.

Responsible by Design. AIIAs are usually conducted at the pre-deployment stage for legal compliance. However, after our co-design process and user studies which began with compliance as a focal point, we learned that reports should be updated not only when lifecycle stages change (e.g., from design to development), but also when the socio-technical context shifts. This underscores the challenge of balancing a fixed impact assessment template with the significant learning benefits users gain from engaging with it to "think differently" about AI. Without clear updates, users may lack the critical thinking and ethical deliberation needed, ultimately reducing empathy for those impacted by AI uses.

Conclusion

We developed an impact assessment template with input from 16 AI practitioners and 6 compliance experts, designed to align with standards such as the EU AI Act, NIST AI RMF, and ISO 42001. A user study involving other 8 company AI practitioners and 5 compliance experts confirmed that our template effectively captures AI system impacts, serving as a starting point for navigating regulatory compliance and fostering responsible design.

Researcher Positionality Statement

In this study, we, the authors, are based in United Kingdom and engage predominantly in industry and academic research during the 21st century. Our team includes two women and two men from Southern, Eastern, and Central Europe, representing a variety of ethnic backgrounds. Our collective expertise spans several fields such as human-computer interaction, ubiquitous computing, software engineering, artificial intelligence, natural language processing, data visualization, and digital humanities. Our positionality may influence the inherent subjectivity in formulating our research, choosing our methods, structuring our co-design and evaluation sessions, interpreting and analyzing data, and addressing the needs of study participants in future iterations of the template (Frluckaj et al. 2022).

As researchers in a predominantly Western setting, we understand the critical importance of broadening the scope of perspectives in our research, particularly to include voices from outside academic and industry spheres. We are committed to promoting future research that is conducted by and with individuals from a wide range of backgrounds, especially those with personal experiences of the impacts of AI systems. This inclusive approach will deepen our understanding and help us develop research methods that are truly responsive to the needs of diverse roles and often underrepresented communities.

References

Ada Lovelace Institute. 2022. Algorithmic Impact Assessment: AIA Template. https://www.adalovelaceinstitute.org/resource/aia-template. Accessed: 2024-08-20.

Ballard, S.; Chappell, K. M.; and Kennedy, K. 2019. Judgment Call the Game: Using Value Sensitive Design and Design Fiction to Surface Ethical Concerns Related to Technology. In *Proceedings of the ACM on Designing Interactive Systems Conference*, 421–433.

Barnett, J.; and Diakopoulos, N. 2022. Crowdsourcing Impacts: Exploring the Utility of Crowds for Anticipating Societal Impacts of Algorithmic Decision Making. In *Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society*, 56–67.

Bender, E. M.; and Friedman, B. 2018. Data Statements for Natural Language Processing: Toward Mitigating System Bias and Enabling Better Science. *Transactions of the Association for Computational Linguistics*, 6: 587–604.

Bogucka, E.; Constantinides, M.; Šćepanović, S.; and Quercia, D. 2024a. AI Design: A Responsible AI Framework for Pre-filling Impact Assessment Reports. *IEEE Internet Computing*.

Bogucka, E.; Constantinides, M.; Šćepanović, S.; and Quercia, D. 2024b. Co-designing an AI Impact Assessment Report Template with AI Practitioners and AI Compliance Experts - Supplementary Materials. https://social-dynamics.net/impact-assessment. Accessed: 2024-08-20.

Buçinca, Z.; Pham, C. M.; Jakesch, M.; Ribeiro, M. T.; Olteanu, A.; and Amershi, S. 2023. AHA!: Facilitating AI Impact Assessment by Generating Examples of Harms. arXiv:2306.03280.

Constantinides, M.; Bogucka, E.; Quercia, D.; Kallio, S.; and Tahaei, M. 2024a. RAI Guidelines: Method for Generating Responsible AI Guidelines Grounded in Regulations and Usable by (Non-)Technical Roles. In *Proceedings of the ACM on Human-Computer Interaction*, CSCW, 1–28. ACM New York, NY, USA.

Constantinides, M.; Bogucka, E.; Scepanovic, S.; and Quercia, D. 2024b. Good Intentions, Risky Inventions: A Method for Assessing the Risks and Benefits of AI in Mobile and Wearable Uses. *Proceedings of the ACM on Human-Computer Interaction*, 8: 1–30.

De Miguel Velazquez, J.; Šćepanović, S.; Gvirtz, A.; and Quercia, D. 2024. Decoding Real-World AI Incidents. *IEEE Computer*.

Diakopoulos, N.; Friedler, S.; Arenas, M.; Barocas, S.; Hay, M.; Howe, B.; Jagadish, H. V.; Unsworth, K.; Sahuguet, A.; Venkatasubramanian, S.; Wilson, C.; Yu, C.; and Zevenbergen, B. 2016. Principles for Accountable Algorithms and a Social Impact Statement for Algorithms. www.fatml.org/resources/principles-for-accountable-algorithms. Accessed: 2024-08-20.

Ehsan, U.; Singh, R.; Metcalf, J.; and Riedl, M. 2022. The Algorithmic Imprint. In *Proceedings of the ACM Conference on Fairness, Accountability, and Transparency*, 1305–1317.

European Comission. 2024. Regulation of the European Parliament and of the Council on Laying Down Harmonised 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). https://www.europarl.europa.eu/doceo/document/TA-9-2024-0138-FNL-COR01_EN.pdf. Accessed: 2024-08-20.

Figma. 2024. Figma: The Collaborative Interface Design Tool. https://www.figma.com. Accessed: 2024-08-20.

Franconeri, S. L.; Padilla, L. M.; Shah, P.; Zacks, J. M.; and Hullman, J. 2021. The Science of Visual Data Communication: What Works. *Psychological Science in the Public Interest*, 22(3): 110–161.

Frluckaj, H.; Dabbish, L.; Widder, D. G.; Qiu, H. S.; and Herbsleb, J. 2022. Gender and Participation in Open Source Software Development. *Proceedings of the ACM on Human-Computer Interaction*, 6(CSCW2): 1–31.

Fujitsu Research Center for AI Ethics. 2022. AI Ethics Impact Assessment. Technical report, Fujitsu. Available at: www.fujitsu.com/global/about/research/technology/aiethics.

Gaba, A.; Kaufman, Z.; Chueng, J.; Shvakel, M.; Hall, K. W.; Brun, Y.; and Bearfield, C. X. 2024. My Model is Unfair, Do People Even Care? Visual Design Affects Trust and Perceived Bias in Machine Learning. *IEEE Transactions on Visualization and Computer Graphics*, 30(1): 327–337.

Gaumond, E.; and Régis, C. 2023. Assessing Impacts of AI on Human Rights: It's Not Solely About Privacy

and Nondiscrimination. https://www.lawfaremedia.org/article/assessing-impacts-of-ai-on-human-rights-it-s-not-solely-about-privacy-and-nondiscrimination. Accessed: 2024-08-20.

Gebru, T.; Morgenstern, J.; Vecchione, B.; Vaughan, J. W.; Wallach, H.; Daumé III, H.; and Crawford, K. 2021. Datasheets for Datasets. *Communications of the ACM*, 64(12): 86–92.

Golbin, I. 2021. Algorithmic Impact Assessments: What Are They and Why Do You Need Them? https://www.pwc.com/us/en/tech-effect/ai-analytics/algorithmic-impact-assessments.html. Accessed: 2024-08-20.

Golpayegani, D.; Pandit, H. J.; and Lewis, D. 2023. To Be High-Risk, or Not To Be–Semantic Specifications and Implications of the AI Act's High-Risk AI Applications and Harmonised Standards. In *Proceedings of the ACM Conference on Fairness, Accountability, and Transparency*, 905–915.

Golpaygani, D.; Pandit, H. J.; and Lewis, D. 2024. VAIR: Vocabulary of AI Risks. https://delaramglp.github.io/vair. Accessed: 2024-08-20.

Gomes de Andrade, N. N.; and Kontschieder, V. 2021. AI Impact Assessment: A Policy Prototyping Experiment. *SSRN Electronic Journal*.

Guest, G.; Bunce, A.; and Johnson, L. 2006. How many interviews are enough? An experiment with data saturation and variability. *Field Methods*, 18(1): 59–82.

Herdel, V.; Šćepanović, S.; Bogucka, E.; and Quercia, D. 2024. ExploreGen: Large Language Models for Envisioning the Uses and Risks of AI Technologies. arXiv:2407.12454.

Holland, S.; Hosny, A.; Newman, S.; Joseph, J.; and Chmielinski, K. 2020. The Dataset Nutrition Label: A Framework to Drive Higher Data Quality Standards. In Hallinan, D.; Leenes, R.; Gutwirth, S.; and De Hert, P., eds., *Data Protection and Privacy*, chapter 1, 1–26. Hart Publishing.

Hupont, I.; Fernández-Llorca, D.; Baldassarri, S.; and Gómez, E. 2024. Use Case Cards: A Use Case Reporting Framework Inspired by the European AI Act. *Ethics and Information Technology*, 26(2).

ISO/IEC. 2023. Information Technology – Artificial Intelligence – Management System. Standard ISO/IEC 42001:2023, International Organization for Standardization.

ISO/IEC. 2025. Information Technology – Artificial Intelligence – AI System Impact Assessment. Standard ISO/IEC DIS 42005, International Organization for Standardization. Status: Under development.

Janssen, H. L. 2020. An Approach for a Fundamental Rights Impact Assessment to Automated Decision-Making. *International Data Privacy Law*, 10(1): 76–106.

Johnson, N.; and Heidari, H. 2023. Assessing AI Impact Assessments: A Classroom Study. arXiv:2311.11193.

Madaio, M. A.; Stark, L.; Wortman Vaughan, J.; and Wallach, H. 2020. Co-designing Checklists to Understand Organizational Challenges and Opportunities Around Fairness in AI. In *Proceedings of the ACM Conference on Human Factors in Computing Systems*, 1–14.

Mantelero, A. 2022. *Beyond Data: Human Rights, Ethical and Social Impact Assessment in AI*. T.M.C. Asser Press.

McDonald, N.; Schoenebeck, S.; and Forte, A. 2019. Reliability and Inter-Rater Reliability in Qualitative Research: Norms and Guidelines for CSCW and HCI Practice. *Proceedings of the ACM on Human-Computer Interaction*, 3(CSCW).

Metcalf, J.; Moss, E.; Watkins, E. A.; Singh, R.; and Elish, M. C. 2021. Algorithmic Impact Assessments and Accountability: The Co-Construction of Impacts. In *Proceedings of the ACM Conference on Fairness, Accountability, and Transparency*, 735–746.

Microsoft. 2022a. Responsible AI Impact Assessment Guide. https://blogs.microsoft.com/wp-content/uploads/prod/sites/5/2022/06/Microsoft-RAI-Impact-Assessment-Guide.pdf. Accessed: 2024-08-20.

Microsoft. 2022b. Responsible AI Impact Assessment Template. https://blogs.microsoft.com/wp-content/uploads/prod/sites/5/2022/06/Microsoft-RAI-Impact-Assessment-Template.pdf. Accessed: 2024-08-20.

Miles, M.; and Huberman, M. 1994. *Qualitative Data Analysis: A Methods Sourcebook*. Sage.

Mitchell, M.; Wu, S.; Zaldivar, A.; Barnes, P.; Vasserman, L.; Hutchinson, B.; Spitzer, E.; Raji, I. D.; and Gebru, T. 2019. Model Cards for Model Reporting. In *Proceedings of the ACM Conference on Fairness, Accountability, and Transparency*, 220–229.

Moss, E.; Watkins, E.; Singh, R.; Elish, M. C.; and Metcalf, J. 2021. Assembling Accountability: Algorithmic Impact Assessment for the Public Interest. *SSRN Electronic Journal*.

National Institute of Standards and Technology. 2023a. AI Risk Management Framework. https://www.nist.gov/itl/ai-risk-management-framework. Accessed: 2024-08-20.

National Institute of Standards and Technology. 2023b. The EqualAI Algorithmic Impact Assessment Tool. https://www.equalai.org/aia. Accessed: 2024-08-20.

Nielsen, J.; and Landauer, T. K. 1993. A Mathematical Model of the Finding of Usability Problems. In *Proceedings of the INTERACT and Conference on Human Factors in Computing Systems*, CHI '93, 206–213.

of Europe The Committee of Experts on Internet Intermediares (MSI-NET), T. C. 2018. Algorithms and Human Rights. Study on the Human Rights Dimensions of Automated Data Processing Techniques and Possible Regulatory Implications. https://edoc.coe.int/en/internet/7589-algorithms-and-human-rights-study-on-the-human-rights-dimensions-of-automated-data-processing-techniques-and-possible-regulatory-implications.html. Accessed: 2024-08-20.

Raji, I. D.; Smart, A.; White, R. N.; Mitchell, M.; Gebru, T.; Hutchinson, B.; Smith-Loud, J.; Theron, D.; and Barnes, P. 2020. Closing the AI Accountability Gap: Defining an End-to-End Framework for Internal Algorithmic Auditing. In

Proceedings of the ACM Conference on Fairness, Accountability, and Transparency, 33–44.

Sadek, M.; Constantinides, M.; Quercia, D.; and Mougenot, C. 2024. Guidelines for Integrating Value Sensitive Design in Responsible AI Toolkits. In *Proceedings of the ACM Conference on Human Factors in Computing Systems*, 1–20.

Saldaña, J. 2015. *The Coding Manual for Qualitative Researchers*. Sage.

Schiff, D.; Rakova, B.; Ayesh, A.; Fanti, A.; and Lennon, M. 2020. Principles to Practices for Responsible AI: Closing the Gap. arXiv:2006.04707.

Selbst, A. D. 2021. An Institutional View of Algorithmic Impact. *Harvard Journal of Law & Technology*, 35(1).

Shelby, R.; Rismani, S.; Henne, K.; Moon, A.; Rostamzadeh, N.; Nicholas, P.; Yilla-Akbari, N.; Gallegos, J.; Smart, A.; Garcia, E.; and Virk, G. 2023. Sociotechnical Harms of Algorithmic Systems: Scoping a Taxonomy for Harm Reduction. In *Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society*, 723–741.

Sherman, E.; and Eisenberg, I. 2024. AI Risk Profiles: A Standards Proposal for Pre-deployment AI Risk Disclosures. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 38/21, 23047–23052.

Skoric, V. 2023. Critical Criteria for AI Impact Assessment: An Aggregated View. *SSRN Electronic Journal*.

Sokol, K.; and Flach, P. 2020. One Explanation Does Not Fit All: The Promise of Interactive Explanations for Machine Learning Transparency. *KI - Künstliche Intelligenz*, 34(2): 235–250.

Stahl, B. C.; Antoniou, J.; Bhalla, N.; Brooks, L.; Jansen, P.; Lindqvist, B.; Kirichenko, A.; Marchal, S.; Rodrigues, R.; Santiago, N.; Warso, Z.; and Wright, D. 2023. A Systematic Review of Artificial Intelligence Impact Assessments. *Artificial Intelligence Review*, 56(11): 12799–12831.

The Danish Institute for Human Rights. 2023. Introduction to Human Rights Impact Assessment. https://www.humanrights.dk/tools/human-rights-impact-assessment-guidance-toolbox/introduction-human-rights-impact-assessment. Accessed: 2024-08-20.

The Government of Canada. 2023. Algorithmic Impact Assessment Tool. https://www.tbs-sct.canada.ca/pol/doc-eng.aspx?id=32592. Accessed: 2024-08-20.

The High-Level Expert Group on Artificial Intelligence. 2020. The Assessment List for Trustworthy Artificial Intelligence. https://altai.insight-centre.org/. Accessed: 2024-08-20.

United Nations Educational, Scientific and Cultural Organization. 2023. Ethical Impact Assessment. A Tool of the Recommendation on the Ethics of Artificial Intelligence. www.unesco.org/en/articles/ethical-impact-assessment-tool-recommendation-ethics-artificial-intelligence. Accessed: 2024-08-20.

Vakkuri, V.; Kemell, K.-K.; Jantunen, M.; Halme, E.; and Abrahamsson, P. 2021. ECCOLA — A Method for Implementing Ethically Aligned AI Systems. *Journal of Systems and Software*, 182: 111067.

Wang, Q.; Madaio, M.; Kane, S.; Kapania, S.; Terry, M.; and Wilcox, L. 2023. Designing Responsible AI: Adaptations of UX Practice to Meet Responsible AI Challenges. In *Proceedings of the ACM Conference on Human Factors in Computing Systems*, 1–16.

Wang, Z. J.; Kulkarni, C.; Wilcox, L.; Terry, M.; and Madaio, M. 2024. Farsight: Fostering Responsible AI Awareness During AI Application Prototyping. In *Proceedings of the ACM Conference on Human Factors in Computing Systems*, 1–40.

Watkins, E. A.; Moss, E.; Metcalf, J.; Singh, R.; and Elish, M. C. 2021. Governing Algorithmic Systems with Impact Assessments: Six Observations. In *Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society*, 1010–1022.

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.; and Isaac, W. 2023. Sociotechnical Safety Evaluation of Generative AI Systems. arXiv:2310.11986.

World Economic Forum. 2022. A Policy Framework for Responsible Limits on Facial Recognition. Use Case: Law Enforcement Investigations. www.weforum.org/publications/a-policy-framework-for-responsible-limits-on-facial-recognition-use-case-law-enforcement-investigations-revised-2022. Accessed: 2024-08-20.

Wright, D.; and Wadhwa, K. 2012. Introducing a Privacy Impact Assessment Policy in the EU Member States. *International Data Privacy Law*, 3(1): 13–28.

Yeung, L. A. 2021. Guidance for the Development of AI Risk and Impact Assessments. https://cltc.berkeley.edu/wp-content/uploads/2021/08/AI_Risk_Impact_Assessments.pdf. Accessed: 2024-08-20.