Impact Assessment Card: Communicating Risks and Benefits of AI Uses

ANONYMOUS AUTHOR(S)

Communicating the risks and benefits of AI uses is crucial for regulatory compliance and increasing public awareness, but its effectiveness is currently limited to individuals with technical expertise and often presented in highly specialized impact assessment reports. Drawing upon the HCI and CSCW literature on making complex concepts broadly accessible, we propose an impact assessment card for communicating the risks and benefits of AI uses in a way that is accessible to individuals without technical expertise. Through an iterative design process, we conducted three focus groups with a total of 12 participants who identified design requirements for an impact assessment card and designed a set of speculative cards. We then reviewed these speculative cards and iteratively produced the final version of the card. We evaluated this card's effectiveness for conducting a real-world task, that is, to write an email to either recommend the implementation of a hypothetical AI system or advising against it, and compared the task's outcomes (i.e., email quality, efficiency, usability, and preference) against a baseline fully-fledged report in an online study with 235 participants grouped in three cohorts: AI developers; compliance experts; and ordinary individuals who reflect US census in terms of age, sex, and race. After controlling for the type of cohort and task, as well as our participants' expertise in AI and technology more broadly, we found that the most significant difference in the task's outcomes was attributed to the use of card or report. In fact, across all three cohorts, the card was found to be more usable and effective; participants spent less time on executing the task at hand and wrote emails of higher quality. Surprisingly, the card not only helped ordinary individuals but also proved useful to developers and compliance experts—two cohorts that are already attuned to the impact assessment process and frequently use reports as part of it. We reflect on the role of HCI in further refining the card through the use of color, language, and metaphorical representations, aiming to break down barriers to understanding the risks and benefits of AI uses and, ultimately, transforming impact assessment cards into a standardized tool for AI governance.

CCS Concepts: • **Human-centered computing → Interactive systems and tools**; **Collaborative and social computing**.

Additional Key Words and Phrases: impact assessment, regulations, artificial intelligence, visualizations

ACM Reference Format:
Anonymous Author(s). 2018. Impact Assessment Card: Communicating Risks and Benefits of AI Uses. In *Conference on Computer-Supported Cooperative Work & Social Computing (CSCW '25).* ACM, New York, NY, USA, 42 pages. https://doi.org/XXXXXXX.XXXXXXX

1 Introduction

The transformative potential of AI in society requries a thorough understanding of its risks and benefits [43, 86], with policymakers advocating that by providing public with algorithmic advice will improve risk predictions, and, in turn, lead to better and fairer algorithmic decisions [25, 33]. This need has led to the creation of fully-fledged impact assessment reports as a way of identifying and mitigating potential risks associated with AI systems, and communicating AI's potential benefits to

CSCW '25, –, —

ACM ISBN 978-1-4503-XXXX-X/18/06
https://doi.org/XXXXXXX.XXXXXXX

individuals, society, and the environment [56]. Producing such reports requires an in-depth grasp of the AI system, from its initial ideation to its real-world deployment. This includes knowledge of the training data, the underlying algorithms, and the effects these systems might have on society and environment. Moreover, it is essential to effectively share this knowledge with all parties involved, including legal entities and the general public, whose rights are often affected by the AI systems [67]. As AI governance continues to evolve, impact assessment report is set to become a legal requirement. The forthcoming EU AI Act, for example, will require detailed reports on the impact of high-risk AI uses on human rights, the environment, and the public interest [20]. These reports aim to increase transparency regarding AI functionalities, hold corporations accountable for the ethical and societal consequences of their AI systems, and allow ordinary individuals to comprehend the risks and benefits of AI uses to make informed decisions about its adoption.

However, a recent review of more than 300 AI auditing tools found that discovering harms within AI systems and effectively communicating these harms have received far less attention than evaluating the technical performance of those systems [65]. Current reports on AI impact assessments, often filled with technical jargon [50], are mainly aimed at experts and can alienate ordinary individuals impacted by AI's societal integration. This creates a barrier to wider understanding and participation in AI-related discussions. Therefore, it is crucial to explore new methods of communicating the risks and benefits of AI uses that are inclusive and understandable to everyone.

Drawing from the HCI and CSCW literature, as we shall see in §2, we aim to simplify and communicate complex concepts pertaining to AI uses for broader public consumption. For example, the use of simple and clear language, icons, metaphors, and color coding can make complex AI information more accessible to ordinary individuals [31, 34]. With that aim in mind, we made two main contributions:

(1) Through an iterative design process, we conducted three focus groups with 12 participants who identified design requirements for an impact assessment card, and designed a set of speculative cards. The design requirements were grouped into two main categories: those related to the information (i.e., what the card should contain), and those related to the design (i.e., how the card should convey the information). By reviewing these speculative cards and soliciting feedback from the research team, we designed our impact assessment card (Figure 1, §4).

(2) We evaluated our card's effectiveness for conducting a real-world task (e.g., a compliance expert typically writes emails to the ethics committee, recommending implementation of an AI or advising against it), and compared it against a baseline impact assessment report in an online study with 235 participants across three cohorts: AI developers, compliance experts, and ordinary individuals who reflect US census in terms of age, sex, and race (§5). We found a strong preference for the card across the three cohorts, with ordinary individuals expressing the highest favorability. Its user-friendly and accessible format not only allowed for faster reading times but also enabled participants to execute the task more efficiently, resulting in higher-quality emails.

We conclude by discussing how impact assessment cards can help assess AI risks, communicate its benefits, and support AI governance. Additionally, we explore design opportunities and potential applications of the cards across various contexts (§6).

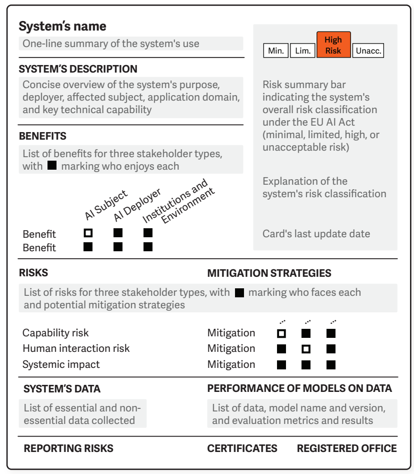


Fig. 1. Template of the impact assessment card. The top section shows the system's name, intended purpose, and overall risk classification under the EU AI Act. The middle section covers benefits, the risk management framework with combined risks and mitigation strategies, and technical details on data and models. The bottom section details reporting mechanisms, registered office, and compliance certifications.

2 Related Work

Next, we surveyed various lines of literature that our work draws upon, and grouped them into two areas: *a)* ways of communicating AI uses' risks and benefits to technical roles (§2.1); and *b)* HCI and CSCW literature concerned with communicating multi-faceted and complex concepts to ordinary individuals (§2.2).

2.1 Communicating Risks and Benefits of AI Uses to Technical Roles

Numerous responsible AI artifacts—defined as processes, tools, documentation templates, and other resources designed to support the ethical creation and use of AI [44]—have been developed for data and model evaluation. However, these artifacts tend to focus more on facilitating risk and benefit communication among technical roles such as developers and engineers [3, 18, 53, 90]. Gebru et al. [29] introduced "datasheets for datasets" for comprehensive dataset descriptions including test key features, test outcomes, and potential biases. Similarly, Bender et al. [4] proposed "data statements" for dataset demographic overviews. For standardized model information, Mitchell et al. [59] suggested "model cards", describing the uses, performance, biases, and limitations of machine learning models. However, a study on completion rates of the cards for HuggingFace' models [50] showed that developers often prioritize completing information on training details, neglecting environmental impacts and evaluations.

Expanding on these artifacts, the prevalent method of communicating the risks and benefits of AI systems to technical roles is through impact assessment reports. Stahl et al. [78] described the impact assessment process as a systematic approach to comprehend the potential positive or negative consequences of an AI system. This process typically entails detailing the AI system's intended use and benefits, evaluating risks, and formulating mitigation strategies. For example, Responsible AI impact assessment template [57] includes system information, identified risks, mitigation measures, and an impact summary. The algorithmic impact assessment [62] further delineates system information into system tasks and operational contexts and categorizes risks as either organizational or those stemming from third-party technologies. These elements are also relevant in specialized fields, evidenced by the algorithmic impact assessment for AI in healthcare [2]. Collectively, these serve as the state-of-the-art example reports for detailing and communicating the risks and benefits of AI systems.

However, the EU AI Act [20] will mandate documenting impacts not at the dataset, model, or AI system level, but for a specific AI system's use, which can be detailed through five components [30]: purpose (the AI's intended goal), AI deployer (the entity managing the AI), AI subject (individuals or groups affected by the AI), capability (the AI's technological feature), and domain (the sector of AI use). To help communicate risks and benefits in this format, Hupont et al. [38] proposed "use cards" that list, among other information, the system's intended use, impacted stakeholders, and Sustainable Development Goals to be supported by the use [85].

2.2 Communicating Multi-Faceted and Complex Concepts to Ordinary Individuals

Communicating AI's risks and benefits to the general public is challenging; however, HCI and CSCW studies provide strategies to simplify these complex concepts for non-experts [26]. Scientific sketchnotes by Fernández-Fontecha et al. [22] combine notes and sketches to introduce complex scientific topics for the layperson. Shen et al. [75] redesigned confusion matrices for binary classification to improve non-experts' understanding of machine learning model performance. They found that by contextualizing terminologies and using flow charts to indicate data reading direction significantly improved comprehension. Similarly, Kehrer and Hauser [45] explored various techniques for visualizing multifaceted scientific data such as abstract representations, data aggregation, and the strategic use of texture and color. The addition of color, particularly red, has been shown to significantly increase perceived risk, a phenomenon observed across multiple cultures despite limited cross-cultural studies [92]. Orange and yellow are the next most commonly used colors for marking risk after red, although people often find it difficult to distinguish which of the two conveys a higher level of risk when used together [92]. Additionally, using prominent typography further enhances the memorability of risk warnings [93]. The length of an artifact (e.g., a card) has also been linked to the comprehension and perceived trustworthiness of an AI. When testing shorter and longer versions of their AI Model Cards among non-experts, Bracamonte et al. [6] found that longer versions of the cards were considered less understandable and interpretable compared to a short version. However, they also found that the short version had a slightly negative effect on the perceived trustworthiness of the AI. Moreover, Kawakami et al. [44] identified additional challenges in ensuring that Responsible AI artifacts such as "datasheets for datasets" [29], effectively serves non-technical stakeholders, including regulators and civil society organizations. These challenges include a misalignment between the technical details provided and the specific decision-making needs of these stakeholders, insufficient clarity in conveying the real-world implications of AI risks, and limited opportunities for stakeholders to evaluate the artifacts. These barriers highlight the need for resources that not only simplify complex AI concepts but also actively engage non-technical actors in the broader governance ecosystem [16, 74]. For example, the AI Failure Cards present real-world AI failures through comic strips that illustrate their impact [81]. They also include structured elicitation questions that help non-technical stakeholders such as frontline workers, service providers, and impacted individuals propose mitigation strategies.

Metaphors are a key tool designers use to shape and influence user expectations effectively or communicate complex information, especially in human-AI collaboration scenarios [46]. One such a metaphor is the use of labels to highlight specific attributes of products or services, aiding consumers in making informed choices. This practice is prevalent in sectors such as agriculture [32], food [41], and energy [77]. For example, "nutrition labels" in the food industry offer a simplified and comprehensible way for consumers to understand a product's nutritional value. Similarly, an impact assessment card for AI systems should distill complex information into a format that helps ordinary individuals to understand the risks and benefits of AI uses such as trade-offs between accuracy and fairness of models [27]. AI Nutrition Facts [84] adopted the metaphor of "nutrition labels" to describe AI services, covering aspects like model type, data use, data retention, privacy practices, and human oversight. Similarly, Open Ethics Label [66] uses the metaphor of "energy label" to disclose details about AI services, including training data provenance, source code, algorithms, and their types of reasoning.

While detailed information is often available on the back of food packaging (similar to how information about AI uses' risks and benefits is presented in full-fledged reports), it can be overly complex for many consumers. This complexity mirrors the challenges end-users encounter with AI documentation. The use of icons [73], charts [27, 54, 69], and straightforward language [28] can render this information more accessible to a diverse audience [31, 34]. For example, using labels with absolute instead of relative rates and conveying probabilities with frequencies (e.g., "3 out of 10") instead of percentages (e.g., "30%") improves understanding of risks in low-numeracy audiences [26]. Deliberate design choices can help not only in conveying the risks and benefits of a product but also in enhancing trust in it [26, 27, 83].

Research Gap. In summary, previous research on communicating the risks and benefits of AI uses has mainly targeted technical audiences, relying primarily on detailed reports. Despite this, the field of HCI and CSCW provides a rich repository of strategies that can be leveraged to create artifacts designed for a wider audience. Our work seeks to bridge this gap by designing and developing an impact assessment card aimed at communicating the risks and benefits of AI uses to both technical and non-technical roles.

3 Author Positionality Statement

Before presenting our impact assessment card, we clarify our positionality to enhance understanding of the methodology, study design, data interpretation, and analysis [15]. We are situated in a Western country in the 21st century, contributing as authors who are predominantly engaged in research within academia and industry at a large technology company.[1] We have contributed to the design, development, and implementation of tools supporting Responsible AI, including guidelines and toolkits. Our team includes four members—two women and two men—from Southern and Eastern Europe, representing diverse ethnic and religious backgrounds. Our combined expertise covers Responsible AI, human-computer interaction (HCI), data visualization, artificial intelligence, and natural language processing. These experiences and backgrounds influenced our data interpretation, the way we incorporated participant feedback into the template's design and development, and the choice of real-world tasks. We recognize the importance of expanding the perspectives presented in this paper and encourage future contributions from individuals with diverse backgrounds, especially those from beyond academia and industry.

4 Design the Impact Assessment Card

To design the impact assessment card, we followed a two-step method that combined insights from existing literature with findings from design activities. First, we reviewed prior studies to identify 14 design patterns commonly used to communicate the risks and benefits of AI applications, which provided a foundation for subsequent speculative design activities conducted in three focus groups (§4.1). Second, we iterated on the outcomes of the focus groups, which included 12 speculative card designs and 8 design requirements. We analyzed the card designs to obtain a preliminary version of our card and then addressed the design requirements to prepare its final version for the user study (§4.2).

[1] REDACTED FOR BLIND REVIEW.

Table 1. Participant demographics of the three focus groups. GID: focus group identifier; PID: participant identifier; Role: AI developer (R_D), compliance expert (R_C), and ordinary individual (R_O).

GID	PID	Age	Gender	Role	Institution	Location
G1	P1	29	F	R_C	Academia	UK
	P2	25	M	R_D	Academia	UK
	P3	34	F	R_O	Industry	Germany
	P4	28	M	R_D	Industry	UK
G2	P5	26	F	R_O	Academia	UK
	P6	59	M	R_O	Industry	Belgium
	P7	27	M	R_C	Academia	UK
	P8	35	F	R_D	Industry	UK
G3	P9	33	M	R_D	Industry	UK
	P10	26	F	R_O	Industry	Portugal
	P11	25	F	R_C	Academia	UK
	P12	27	M	R_D	Academia	UK

4.1 Identify Design Patterns From Literature and Conduct Speculative Design Activities in Focus Groups

4.1.1 Identify Design Patterns From Literature. We started by analyzing three systematic literature reviews that compile tools for communicating AI risks and benefits, as well as labels for trustworthy AI [12, 65, 80]. We extracted an additional set of 4 design patterns from these papers such as nutrition labels for datasets [36], icons for AI legibility [52] or certificates for machine learning methods [60]. Finally, we reviewed studies in agriculture [32], food [41], and energy [77], where labels have been effectively employed to communicate complex information to consumers. This review resulted in one additional design pattern. For the food and energy domains, we did not find any new patterns, as the food metaphor was already used in nutrition labels for datasets [36], and the energy efficiency metaphor was used in the AI ethics label [80]. The only new pattern came from the agricultural domain and was a data hazard label, inspired by chemical hazard labels such as those for flammable substances [94].

We grouped the 14 design patterns derived from the literature into two categories (Appendix A.1, Figure 6): visual representation (i.e., common visual elements for communicating AI uses' risks and benefits), and layout (i.e., how visual elements are combined together). For visual representation, we identified the use of textual descriptions, numeric values, links, tags, icons, charts, data samples, checkboxes, and metaphors (e.g., traffic lights). For the layout, we identified the use of lists, tables, rankings, grids, and groups. The list of the design patterns may not be exhaustive but rather served as a kickstarter in the speculative activities during the focus groups. To facilitate similar activities, we made the the list available at https://anonymous.4open.science/r/AIIA_Card.

4.1.2 Identify Design Requirements and Design Speculative Cards in Three Focus Groups.
Participants. We used snowball sampling and started from identifying initial participants (5) through an internal mailing list at a large tech company. These participants were asked to refer additional participants from their own networks, expanding the sample size through successive referrals. We recruited a total of 12 participants (6 female, 6 male, with a median age of 27.5 years old) representing three different cohorts: AI developers (5), compliance experts (3), and ordinary individuals (4). We then conducted 3 focus groups of 4 participants each, ensuring each group had at least one participant from each cohort (Table 1).

Procedure. The focus group consisted of four phases: *briefing*, a *brainstorming* task, a *speculative design* task, and *debriefing*. During the *briefing*, participants were introduced to the concept of impact assessment cards, along with two examples to familiarize themselves with the topic: AI Nutrition Facts [84] and Open Ethics Label [66], which provide descriptions of AI services in the style of "nutrition labels" and "energy labels, respectively". They then moved to a Figma board environment [23] to engage in the tasks.

During the *brainstorming task*, we aimed to surface the needs of different cohorts for the impact assessment card. It started with an idea generation session where participants used notes to brainstorm about their needs in terms of the card's functionality, specific tasks they think the card will assist with, information content, and format. This was followed by categorizing the ideas into four types of requirements: "must have", "should have", "could have", and "won't have". This categorization is based on the MoSCow method for managing trade-offs during product design [1]. Must-have requirements describe critical features; should-have indicate important but not critical features; could-have describe desirable features (e.g., which could improve user experience); and, won't have indicate features that have been considered but explicitly decided against.

During the *speculative design task*, we aimed to surface visual representations of the impact assessment cards that align with our participants' needs identified in the brainstorming task. Participants were first asked to read a report that documents the risks and benefits of a hypothetical AI system for identifying crime hotspots in public spaces using CCTV footage. Informed by previous studies [51, 70], the use of a hypothetical system with real-world applicability served as a way to help participants contextualize their speculative designs. Participants were then introduced to the 14 design patterns derived from the initial literature review (Appendix A.1, Figure 6), and given five minutes to review all patterns. Finally, they were asked to create a speculative design for an impact assessment card for the hypothetical system. They could either build upon the existing patterns or propose new ones. Participants were instructed to sketch their design using pen and paper, photograph it, and upload it to the Figma board. At the end of this task, each participant explained their design choices.

The focus group ended with a *debriefing* to summarize the main ideas that emerged and provided an opportunity for participants to share any final recommendations for the card. Each group session lasted 1 hour, and was both video and audio recorded upon consenting participants. The audio was automatically transcribed by the video conferencing software. The study was approved by our organization.[2]

Analysis. To derive design requirements, we conducted a qualitative analysis of the recordings and audio transcripts of the focus groups, which include participants' expressed ideas during the brainstorming and speculative design tasks and debriefing sessions. Two authors thematically analyzed these ideas following an inductive approach [7, 55, 58, 71]. The authors used Figma [23] to collaboratively create affinity diagrams based on these participants' inputs. Over the course of six meetings, totaling 16 hours, they discussed and resolved any disagreements that arose during the theme analysis process. From each resulting theme, the authors derived a design requirement and provide example(s) how our participants' phrased the requirement.

Results. Our participants envisioned a wide range of potential uses for the card, including comparing the quality of different AI-based services (2 mentions), understanding the safety of AI-based services before deciding to purchase or subscribe to them, often under time pressure (4 mentions), and contacting relevant authorities or support teams for concerns when problems occur (4 mentions). To support these and similar uses, participants identified eight requirements for the card. We grouped them into two main categories (Table 2): those related to the information (i.e., what

[2]REDACTED FOR BLIND REVIEW.

Table 2. Eight design requirements identified during the focus groups, grouped into those on information and on design, along with implementation decisions for the preliminary (A) and final (B) version of the card.

Theme	Design Requirement	Implementation Decision in A (Preliminary Version)	Implementation Decision in B (Final Version)
Requirements on information			
Data	**R1.** Show information about the system's data, distinguish between its essential and non-essential, and personal identifiable and document later uses of the data	Add a table with icons and tags to distinguish between data types	Replace the table with the heatmap of data types
Model	**R2.** Show information about models and its performance	Add a table detailing model names, versions, and accuracies	Align the model table with the heatmap of data types
Benefits	**R3.** Show information about the system's benefits enjoyed by individuals and the environment influenced by the system	Add a list of benefits for direct stakeholders (AI deployers using the system and AI subjects affected by the system) and indirect (related institutions and environments)	Replace the list with the heatmap of stakeholders enjoying the benefits
Risks and Mitigations	**R4.** Show information about system's risks faced by individuals and the environment influenced by the system and potential mitigation strategies	Add a list of risks and a list of mitigations for direct and indirect stakeholders	Combine the two lists into a table with risks, mitigations and the heatmap of affected stakeholders
Reporting and Governance	**R5.** Show information about reporting mechanisms and who it's responsible for its governance	Add two sections for reporting mechanisms and compliance certifications	Combine sections and include the registered office address
Requirements on design			
Accessible Communication	**R6.** Use accessible textual and visual communication for quick decision-making	Use concise language, avoid technical terms, add summary bar with the system's risk classification	Add concise description of the system including the direct stakeholders, refine the summary bar and provide its explanation
Accessible Medium	**R7.** Use medium that is accessible both physically and digitally even by people with different abilities and those visually impaired	Link the card with a QR code to a longer version of the impact assessment report, ensure print and Braille compatibility, use high-contrast design	Improve the contrast ratios in the summary bar
Cultural Inclusivity	**R8.** Use inclusive textual and visual communication for accommodating diverse cultural perspectives	Avoid the use of culturally sensitive colors and icons	Remove the icons and tags for data types

the card should contain)—*R1-5*, and those related to the design (i.e., how the card should convey the information)—*R6-8*. Regarding the *information*, we identified five requirements about the: *data*, *model*, *benefits*, *risks and mitigation strategies*, and *governance and reporting*. Data is about ensuring that card users are fully informed about the types of data the system collects to enable its use. For example, P2, a developer, suggested that *"the card should include what data an AI system accesses about a certain user, how this data is used by the system (i.e., is it used to train the model or is it stored and for how long)"*. P1, a compliance expert, saw this section of the card as a way to *"help people to choose whether to provide their data for a system, as when signing up to a new service or purchasing tech (e.g., Alexa, Notion AI)"*. Model requirement is about making the inner workings of the system's models transparent to the users. P4, a developer, emphasize that the card *"should specify all data sources that the models have been trained on pass certain assessments, and report the models' accuracy"*. Benefits is about informing users about the broader effects of the system, including its positive impact on people and the planet. P9, a developer, expressed this need by stating *"I want to see the value of the system based on the collected data"*. Risks and mitigation strategies are centered on acknowledging and addressing the potential negative impacts or risks associated with the system's operation. P11, a compliance expert, stated that the card should report *"what is the risk-level of*

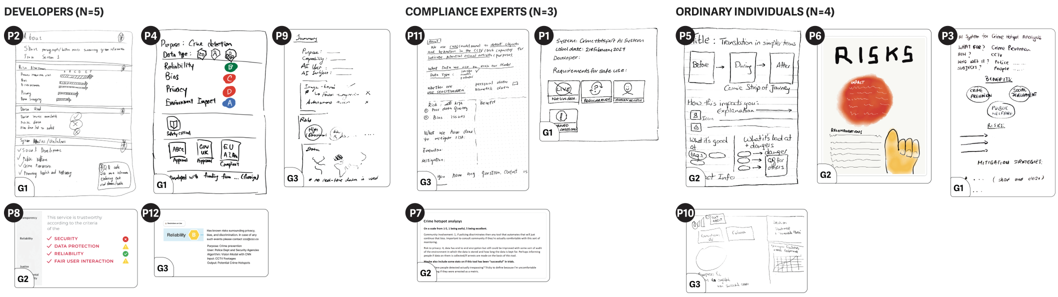


Fig. 2. Speculative impact assessment cards created by 12 participants (P1-12) during three focus groups (G1-G3), sorted by cohort and layout.

the AI system (and how this risk level is decided)". Similarly, P2, a developer, stated that *"[the card] should assist potential users of AI systems to quickly understand how 'safe' they are before deciding to purchase or subscribe to them with a star-based rating"*. Finally, governance and reporting is about the system's regulatory compliance, accountability mechanisms, and the availability of channels for reporting concerns or risks. For example, P3, an ordinary individual, highlighted that *"[the card] should tell me straight away safe the product is and who certified it"*.

Regarding the *design*, we identified three requirements about: *accessible communication*, *accessible medium*, and *cultural inclusivity*. Accessible communication is about ensuring that all system-related information is presented in a manner that is easily understandable and accessible to a wide range of users. For example, P1, a compliance expert, stated that *"the card should use language that is understood by everyone"*, while P8 and P6 stated that it should be *"simple and straightforward"* and *"understandable at a glance"*, respectively. Accessible medium emphasizes the need for the system's information and reports to be accessible across various formats and platforms, catering to diverse user needs. For example, P3, an ordinary individual, pointed towards the idea of providing *"access to more information about the product (e.g., QR Code)"*. Similarly, P2, a developer, stated that the card *"should be available in both physical and digital form, depending on the type of system. An AI-powered smart speaker should have the card printed on the box, but an online subscription-based AI system like ChatGPT should have a digital equivalent shown to the user right before they complete registration"*. Finally, cultural inclusivity involves designing the system in a way that is considerate of and respects diverse cultural backgrounds and perspectives. For example, P5, an ordinary individual, expressed that *"[the card] should be culturally sensitive (e.g. colors used to signify bad vs good)"*.

4.2 Iterate on the Results of the Activities to Design the Impact Assessment Card

4.2.1 Review the Cards From the Speculative Activities to Obtain the Preliminary Version of the Card. The speculative design task during the focus groups resulted in a set of 12 speculative cards (Figure 2) for the crime hotspot analysis system. To demonstrate the generalizability of our card across a variety of AI systems, we chose to implement the card for a similar system that processes personal image data and is more often encountered by ordinary individuals—a biometric supermarket checkout. To do so, we reviewed the set of cards designed by our participants, and for each design requirement (Table 2), we devised a set of implementation decisions that guided our initial card's design for the biometric checkout system.

Implementation decisions for meeting requirements on information. To communicate system's data as per *R1*, we introduced a two-column table inspired by the data nutrition labels [36], with one column for essential data (mandatory for system operation) and another for non-essential data (not critical for operation). Based on the speculative cards from P4, P5 and P8, each collected data is listed as a row with an icon indicating its format (e.g., image icon for image data). Based on P2's and P11's designs, each data is accompanied with tags indicating whether it contains personally identifiable information (as defined by GDPR) and whether it can be potentially re-used in other AI systems. To show model information as per *R2*, we created a section documenting the performance of system models in accordance with the guidelines outlined for the model cards [59] and card from P12. This section is also structured as a table, listing each collected data with corresponding columns for the model's name, version, and accuracy. While we primarily report on accuracy, the table can be extended to include other relevant metrics (e.g., error rates or confidence intervals). To communicate the benefits of the system's use as per *R3*, we included a section listing these benefits (as suggested by P2, P3, P11) across three stakeholders mentioned in the EU AI Act [20]: direct AI deployers (those using the system) and AI subjects (those affected by the system) [30], and indirect related institutions and environment. Using these stakeholders, we structured the subsequent section to list stakeholder-specific risks of system's use alongside potential mitigation strategies, as per *R4* and the cards of P3, P6, P10, P11. To facilitate reporting and governance as per *R5*, we incorporated two sections: one providing information on reporting channels (e.g., dedicated email, phone number) and another showing compliance certifications (as seen on cards by P5, P11, P12) and a QR code linking to the full assessment report.

Implementation decisions for meeting requirements on design. To ensure accessible communication as per *R6*, we refined the language to contain short phrases (maximum 50-65 characters or 8-11 words) and non-technical terms (as seen on cards by P2 and P3). This resulted in a Flesch-Kincaid Grade Level score of 11, indicating suitability for readers aged 16-17. Additionally, we introduced a summary bar similar to those found on food labels and drawn on cards by P4, P10, and P12, denoting the one-letter shortcuts for the system's overall risk classification as per the EU AI Act (with M for Minimal, L for Limited, H for High Risk, and U for Unacceptable risk). To ensure accessible medium as per *R7*, we linked the card with a QR code (as suggested on cards P2 and P5), allowing digital access to the full impact assessment report in print- and Braille-friendly formats. We further improved the card's readability by opting for a high-contrast design, with white background, ample white spaces, and black text in a 14-point sans-serif font with 125% interline spacing to prevent text overcrowding (as in the medical leaflets [17]). To ensure cultural inclusivity as per *R8*, we refrained from employing culturally sensitive or strongly expressive colors and icons such as multiple shades of red for risk levels (visible on cards P4, P6). Instead, we selected a consistent color scheme for our risk summary bar based on established guidelines for the cross-cultural use of color in warnings [92]: red for unacceptable uses, dark orange for high-risk uses, yellow for limited-risk uses, and blue for minimal-risk uses.

Figure 3A presents the first version of the card (nine sections). The top section contains the header with the AI system's name, its intended use, and a risk summary bar. The remaining sections are organized into two columns. The left column consists of four sections addressing various types of impact (benefits, risks, mitigation strategies) and providing information on reporting mechanisms. The right column contains technical details (system's data and model information), compliance certifications, and a QR code for accessing the full impact assessment report.

4.2.2 *Gather Recommendations From the Research Team on the Preliminary Version of the Card.*

During the development of the card, the first author conducted five sessions with the research team, progressively integrating feedback into new versions of the card. By the time version 4 of the card was completed, all necessary feedback was implemented and we ceased further iterations, enhancing the card as follows.

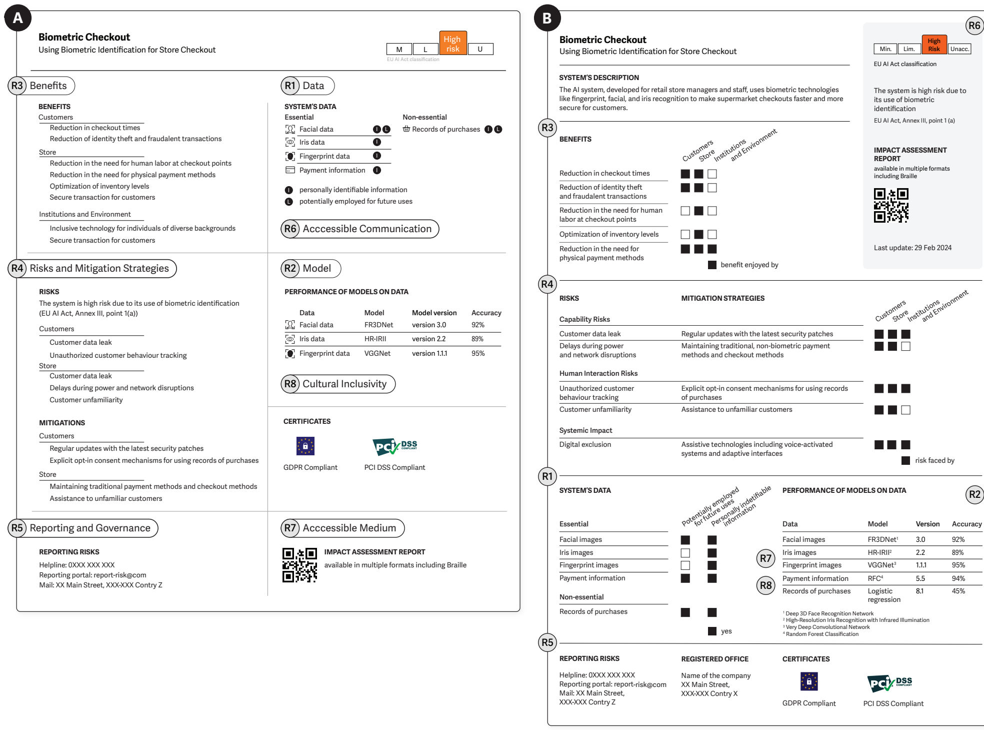


Fig. 3. Impact assessment card: preliminary (A) and final (B) version. Both versions meet the 8 design requirements identified during the focus groups: present information on the system's data (R1), model (R2), benefits (R3), risks and mitigation strategies (R4), and governance (R5), while ensuring accessible communication (R6), medium accessibility (R7), and cultural inclusivity (R8). The final version was the result of four design iterations in the team.

Recommendations for meeting requirements on information. To provide a clearer picture of the system's data (*R1*), we transformed the two-column table into a heatmap. Essential and non-essential data is now displayed in a single column, with adjacent checkboxes replacing the icons and tags. This format enables easier recognition of patterns (e.g., excessive collection) and addition of new criteria (e.g., information about the source of data, licensing, real-time processing), without breaking the card's layout. To gain a better understanding of the model's effectiveness (*R2*), we aligned the model performance section with the data one. Each row of the data's heatmap is linked to a specific model that uses the data and its overall performance. This integration simplifies the evaluation process. To improve the presentation of benefits (*R3*), we explored alternative ways of grouping them. That is because we observed that the benefits were being repetitively listed across the AI deployer and AI subject—direct stakeholders. Similarly to the data, we introduced a heatmap with checkboxes for two key purposes: to clearly indicate the benefits that apply to each stakeholder, and to allow for potential expansion of the stakeholders' list. We also noted that, like benefits, risks were repetitively listed across different stakeholders. To better contextualize them as per *R4*, we made three iterations. First, we categorized them according to capability, human interaction, and systemic risks, aligning with a framework for evaluating sociotechnical harms [91]. Next, for each risk category, we included a set of mitigation strategies. Finally, we used a heatmap to indicate the relevance of each risk to stakeholders, after considering the mitigation strategies.

These iterations resulted in one section presenting a holistic view of risk management, enabling readers to see both the problem and the solution in one place. To improve the presentation of reporting and governance information (*R5*), we restructured the section to combine risk reporting methods and certifications, while also expanding it to include details about the registered office (e.g., the official address of the legal entity responsible for the development and deployment of the system). This helps to build confidence in the system's transparency and adherence to legal standards, reinforcing readers' trust and assurance.

Recommendations for meeting requirements on design. To improve communication accessibility (*R6*), we made two iterations: expanding the header and introducing a corner box. In the expanded header, we added a concise description outlining the system's core aspects using a five-component format [30]: the system's purpose, the overseeing AI deployer, the affected AI subject, the application domain, and technical capability enabling the use. In the new corner box, we placed the risk summary bar, which we refined by replacing vague one-letter shortcuts with clearer abbreviations. Below this bar, we provided explanations for each risk level (e.g., being high risk), linked these to relevant articles from the EU AI Act. We also incuded a QR code for the full report and the date of the card's last edit.

We also refined the language describing the collected data to remove any ambiguities regarding the types of data collected. We iteratively transitioned from general terms in version 1 of the card (e.g., "facial data") to more precise descriptions in version 4 (e.g., "facial images"). To improve medium accessibility (*R7*), we revised the risk classification colors in the summary bar and improved their contrast ratios. Finally, to improve cultural inclusivity (*R8*), we removed the icons representing the types of data collected. Although they work well for systems processing few datasets, their creation becomes problematic as the system expands to multiple datasets or more complex data types. Moreover, the use of numerous icons on a small card could lead to visual clutter, compromising the clarity of the information presented.

4.2.3 Final Version of the Card. Figure 3B presents the final version of the card. The top section of the card contains the expanded header and a corner box. The central section features the system's benefits, the risk management framework with combined risks and mitigation strategies, and the technical details on data and models. The bottom section contains information on reporting mechanisms, registered office and compliance certifications.

5 Evaluate the Impact Assessment Card

Having designed the card, we then evaluated it in a large-scale online study. The study's goal was to explore the effectiveness of the card to communicate the risks and benefits of AI uses in a way that is accessible beyond technical roles. Next, we describe our study's design (i.e., setup (§5.1), execution (§5.3), metrics (§5.2), and results (§5.4).

5.1 Setup

We developed a web-based survey that included a real-world task to be performed either with the card or with the impact assessment report as baseline (Figure 4, Step 2 and Step 5).

Task. We defined a task related to the AI system that participants from each cohort might typically perform as part of their jobs [49] or interactions with AI: writing recommendation and feedback emails. This task was formulated based on insights from three areas: our focus groups about practical actions people take in response to AI systems affecting their lives, including the frequent need to contact relevant authorities or support teams when problems occur; conversations with AI practitioners and compliance experts in our organization about tasks in AI approval processes [65]; and previous user studies on writing AI recommendations by different stakeholders [3, 5].

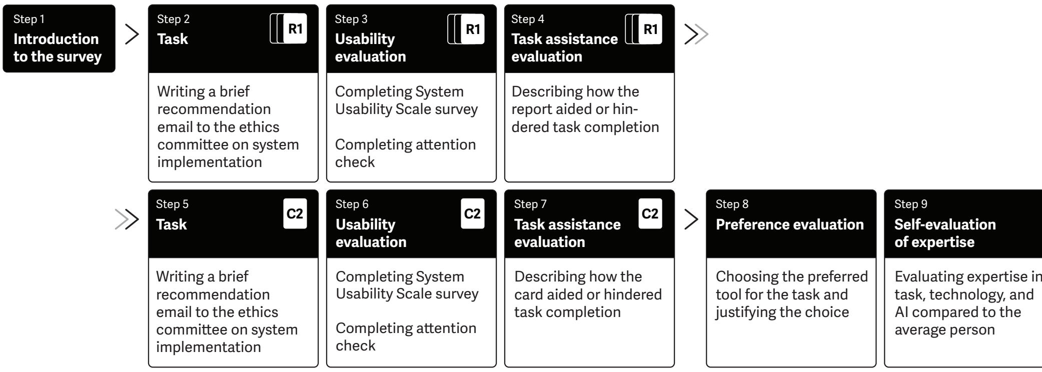


Fig. 4. The online study involved 9 steps. Initially, participants received a brief introduction to the survey and tasks (Step 1). Then, they interacted with the first randomly assigned treatment (e.g., R1 - a report for the biometric checkout), completing a task (Step 2). Subsequently, they assessed the usability (Step 3) and assistance (Step 4) of the treatment. This process was repeated for a second treatment (Steps 5-7) depicting a different AI system (e.g., C2 - a card for the license plate detector). Finally, participants selected their preferred treatment for the task (Step 8), and self-evaluated their knowledge about the task, technology, and AI.

Specifically, for an AI developer, the task was to read the card, and write a brief email to the ethics committee, recommending the implementation of the AI system or advise against it, in either case stating appropriate technical reasons. For a compliance expert, the task was to write an email to the ethics committee, recommending implementation of the system or advising against it. For an ordinary individual, the task was to write an email to the deployers who put in the AI system, asking them to take it out or thanking them, and in either case tell them why. The decision to reject or recommend the system was left entirely up to the participants based on their own judgment.

This task links the information from the card to three advanced decision-making skills typically supported by visualizations [10]: problem-solving (determining appropriate actions), critical thinking (assessing and integrating information on risks, mitigation strategies, and benefits), and reasoning (forming logical arguments to justify actions). It leverages the specific skills and knowledge areas pertinent to each cohort: AI developers use their technical expertise, compliance experts apply their regulatory knowledge, and ordinary individuals draw from their user experience.

We requested that emails from each cohort include between 50 and 250 words, a range that reflects the typical length of descriptions used by AI practitioners in model documentation [50]. This word range ensures conciseness and adequate detail for thematic analysis while preventing survey fatigue among participants.

Treatment. In addition to the card (Appendix A.2, Figures 7-8), we included a baseline condition to compare the card against (Appendix A.3, Figures 9-10). We created an impact assessment report based on current state-of-the-art practices for communicating the risks and benefits of AI systems [2, 62, 78], drawing on examples from published reports [14, 57, 76]. These reports are issued by deployers of high-risk AI systems, as required under the EU AI Act [20], or by organizations seeking certification under AI management standards [39]. The intended audience will primarily include market surveillance authorities, affected stakeholders, independent experts, and civil society organizations to ensure transparency and accountability. Our report mirrored the card's content (e.g., the system's use, data, models, evaluation, risks, mitigations, benefits, contact information, and certificates) but was more descriptive. We alternated between the card and report to eliminate any learning effects. Therefore, a participant was asked to execute the same task with the two conditions. To eliminate any effects from the type of AI systems shown in the card or the report, we selected two hypothetical real-world AI systems that are different in risk levels but are likely familiar to most participants. Next, we provided a brief description of each AI system.

Biometric Checkout. This AI system uses biometric technology such as facial recognition to identify customers during the checkout process in a supermarket. By linking biometric data to payment methods and shopping histories, it enables a seamless and secure checkout experience, eliminating the need for physical cards or cash. This system is categorized as high risk under the EU AI Act [20] due to its extensive use of biometric identification (Appendix A.2, Figure 7; Appendix A.3, Figure 9).

License Plate Detector. This AI system uses cameras and image recognition technology to detect and read license plates of vehicles entering and exiting a supermarket car park. It can be used to monitor parking occupancy, enforce parking time limits, and ensure the security of the parking area. It is categorized as limited risk under the EU AI Act [20] due to its processing of personally identifiable data (Appendix A.2, Figure 8; Appendix A.3, Figure 10).

Both systems, while beneficial to customers and stores, are considered risky under the EU AI Act [35] due to real-time processing of personally identifiable information. Additionally, their excessive information collection and multi-model architecture enable potential future applications beyond their initially stated purpose.

5.2 Metrics

Independent to each cohort, we defined five metrics to capture the effectiveness in conducting the task. The first metric, *task quality*, captured whether the resulting email was considered high quality. The email's quality was scored on a 5-point Likert scale based on how effectively the person used the information from the card or report to justify a recommendation for adopting or rejecting the system. An email scoring 1 was vague, applicable to any AI system, lacked a decisive call to action, and contained no arguments. An email scoring 5 was specific to the system described in the card or report, included a clear recommendation or rejection, and presented diverse arguments covering aspects such as risks, data, benefits, and mitigations. The second metric captured the factors *influencing task quality* (both positively and negatively), with two open-ended questions: *"In what ways did the card (or report) succeed to assist you in completing the task?"* and *"In what ways did the card (or report) fall short to assist you in completing the task?"*. The third metric captured *efficiency* in conducting the task, measured as the average time needed to read the card or report and complete the task. The fourth metric captured the *usability* of the card or report, measured using the System Usability Scale [8]. Finally, the fifth metric captured the overall *preference* for using the card or report for the task.

5.3 Execution

We recruited participants from Prolific [68] and surveyed them across three cohorts: *a)* AI developers; *b)* compliance experts; and *c)* ordinary individuals (Table 3). To recruit a sufficiently large number of participants for each cohort, we controlled for the participants' roles in the organization, the frequency of AI use in their jobs, and their geographic location using Prolific's built-in screeners. Additionally, we controlled for their expertise in the task at hand, technology in general, and AI through a self-reported assessment.

To recruit AI developers, we searched for participants who likely contribute to developing AI systems as part of their software engineering roles, using AI every day. We recruited 65 developers

Table 3. Self-reported knowledge and demographic characteristics of participants.

Control	**Characteristic**	**AI Developers** (n=65)	**Compliance experts** (n=65)	**Ordinary individuals** (n=105)	**US Census** [87, 88]
Expertise	Task	3.38	3.49	3.07	-
	Technology in general	4.20	3.60	3.30	-
	Artificial Intelligence	3.82	3.32	2.96	-
Age	18-29 years	30%	12%	20%	20%
	30-39 years	37%	23%	17%	18%
	40-49 years	18%	22%	17%	16%
	50-59 years	7%	30%	16%	16%
	60 years and above	8%	13%	30%	30%
Sex	Female	11%	48%	50%	50%
	Male	89%	52%	50%	50%
Race	White	54%	57%	60%	62%
	Black	14%	17%	11%	12%
	Asian	25%	15%	6%	6%
	Mixed	5%	9%	10%	10%
	Native American or Alaskan Native	0%	0%	1%	1%
	Other	2%	2%	8%	9%
	Not specified	0%	0%	4%	-

with a median age of 33 years: 7 female and 58 male, mostly White (54%) and Asian (25%). These participants were the most knowledgeable in technology and AI across the three cohorts.

To recruit compliance experts, we searched for participants likely involved in revising AI systems as part of their legal roles, using AI at least 1-6 times a week. We recruited 65 experts with a median age of 42 years: 31 female and 34 male, mostly White (57%) and Black (17%). These participants were the most knowledgeable about the task at hand across the three cohorts, more knowledgeable in technology and AI than ordinary individuals, yet less so than AI developers.

To recruit ordinary individuals, we used stratified random sampling to match US census demographics [87, 88] in terms of age (20% in range 18-29 years, 17% in range 30-39 years, 17% in range 40-49 years, 16% in range 50-59 years, 30% over 60 years), sex (50% female, 50% male), and race (60% White, 11% Black, 10% Mixed, 6% Asian, 1% Native American or Alaskan Native, 8% Other)[3]. Compared to AI developers and compliance experts, as expected, ordinary individuals used AI less frequently in their jobs and had the least knowledge about the task at hand, technology, and AI. We restricted our participant pool to individuals living in the US for one main reason. Involving native English speakers ensured a clear understanding of the study materials, which strengthened the reliability of the findings. All participants were paid on average about $12 (USD) per hour.

Procedure. We administered the survey on Prolific [68]. The survey first provided a brief introduction to the tasks, followed by the first task in which participants had to read either the card or report, and write the email, self-choosing to recommend or reject the system. This was followed by a series of questions to capture the usability of either the card or report, and questions about

[3]Our research does not separately account for ordinary individuals who identify as Hispanic or Latino—the second-largest racial and ethnic group in the U.S.—because our recruitment followed the guidelines of the U.S. Census Bureau [87] and the U.S. Office of Management and Budget [89]. These sources define race using five categories—White, Black or African American, American Indian or Alaska Native, Asian, and Native Hawaiian or Other Pacific Islander, as reported above—while classifying Hispanic or Latino origin as an ethnicity. As a result, ordinary individuals in our sample who identify as Hispanic or Latino are recorded within these five racial categories.

in what ways did they succeeded or fall short in assisting participants in completing the task. Participants repeated the same procedure for the second task. At the end, they were asked to report their overall preference for the card or report in conducting the task.

To ensure response quality, we conducted two attention checks during the survey and implemented two deliberate survey design features. First, after reading task instructions, participants encountered one of the attention-check sentences: *"When asked for your favorite color, you must select 'Blue'"* and *"When asked for your favorite city, you must select 'Rome'"*. Participants had to correctly respond to these checks after completing each task. Second, we disabled pasting from external sources and editing previous responses to ensure original and thoughtful answers.

To control for the extent to which the answers depended on the participants' level of knowledge, we asked them whether they consider themselves more skilled or knowledgeable than most people for the task at hand, as well as for the technology in general and AI. This expertise was assessed using a 5-point Likert scale.

Analysis. We performed both quantitative and qualitative analyses. For the quantitative analysis, we measured for both the task completed with the card and the task completed with the report: the average quality of the task, the average time to complete the task; the average SUS usability scores; and the percentage of participants who preferred the card or report for the task. The evaluation of task quality was conducted by two contributing authors with expertise in Responsible AI, excluding the first author. Their assessment focused on whether the resulting emails were of high quality. Each email was rated by both authors on a 5-point Likert scale, ranging from poor (1) to excellent (5), based on five key criteria: context, recommendation, risks, mitigations, and content clarity. To ensure consistency and accuracy in evaluations, the authors followed a predefined rubric (Appendix A.4). The rating process was blind to the experimental condition—authors did not know whether an email was generated using the card or report. However, they were aware of the cohort (developers, compliance experts, or ordinary individuals) since task formulation differed slightly across these groups. The authors' assessments were largely consistent, with an inter-rater agreement of 85%. In cases where the authors assigned different ratings, they discussed discrepancies in two assessment review meetings with the broader research team to reach a final decision.

We hypothesized five factors that might influence the quality of the task: the type of task (reject or recommend the system), the system (biometric checkout or license plate detector), the participant cohort (AI developers, compliance experts, or ordinary individuals), the participants' level of expertise (low or high), and, crucially, the treatment (card or report). We then conducted linear regression analyses and mean difference testing on these factors.

For the qualitative analysis, we thematically analyzed open-ended responses [7, 55, 58, 71] to understand the factors affecting task quality and preferences for using the card or report.

5.4 Results

We received a total of 235 responses: 65 each from AI developers and compliance experts, and 105 from ordinary individuals. Next, we discuss the quantitative results based on our five metrics (§5.4.1), followed up by qualitative results (§5.4.2).

5.4.1 Quantitative Results. **Regardless of the cohort, participants, on average, spent less time completing their tasks with the card than with the report with even better quality.** Compliance experts achieved the highest average email ratings when using the card (3.59), followed closely by developers (3.53), and then ordinary individuals (2.92) (Figure 5). They also took the longest to complete their tasks with the card (7 min 24 secs) (Appendix A.5, Table 6).

The card was rated higher in usability than the report, especially among ordinary individuals. Developers rated the card with an average SUS score of 67, compared to the report's

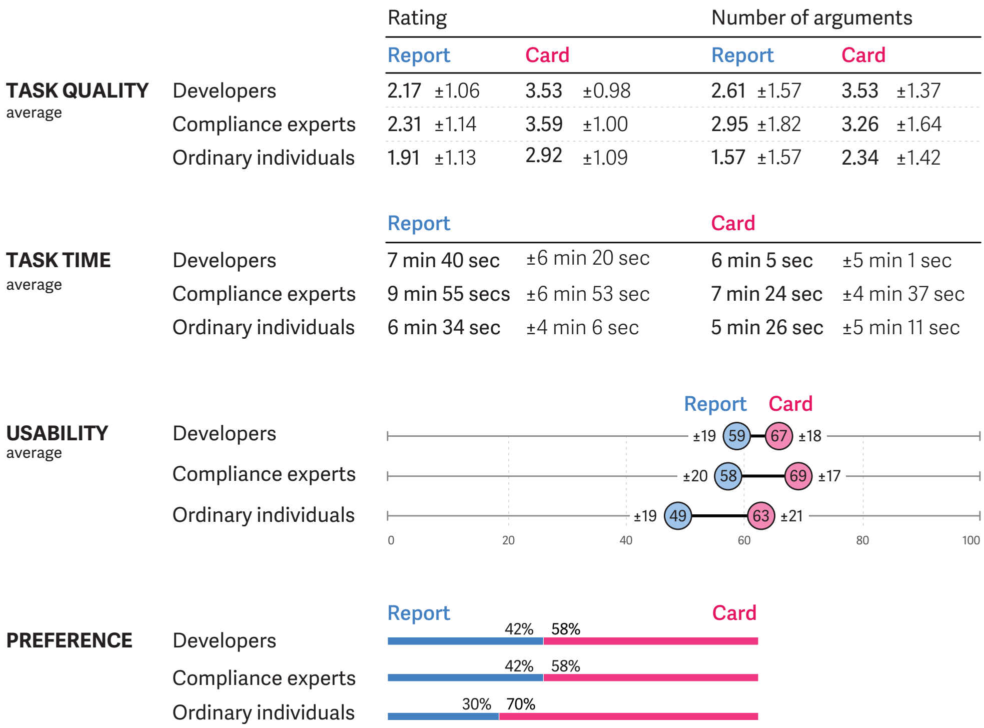


Fig. 5. **Card outperformed report across all quantitative metrics and cohorts**. It helped produce higher quality emails in less time, while being more usable and preferred for the task.

score of 59, indicating a preference for the card's usability (Figure 5) and generally positive user experience [72]. Compliance experts shared this view, scoring the card at 69, with the report at 58. However, this distinction was most pronounced among ordinary individuals, who gave the card a SUS score of 63, compared to a score of 49 for the report (Appendix A.5, Table 7).

All cohorts preferred the card over the report to execute the task at hand, with a higher preference among ordinary individuals compared to developers and compliance experts. Over half of both developers and compliance experts, at 58%, favored the card over the report (Figure 5, Appendix A.5, Table 8). In contrast, 70% of ordinary individuals strongly preferred the card, compared to 30% favoring the report.

The most significant difference in the task quality was attributed to the use of card or report. The most significant difference in task quality was due to treatment (Table 4, Table 5), with the card receiving consistently higher ratings for task quality compared to the report. The type of task (advising for or against either of the two systems) and the participants' expertise levels did not impact the quality.

5.4.2 Qualitative Results. Through thematic analysis of participants' free-form answers, we identified key factors affecting their experience with the card and report, their overall preferences, and suggestions for improving the card. Participant quotes are referenced using CP_N, corresponding to their Prolific ID.

Table 4. **The results of a linear mixed-effects regression analysis with task quality as the dependent variable. The most significant difference in task quality arises from the choice of treatment.** The coefficients represent the effect sizes for each factor relative to its reference category, with statistical significance indicated by: ** for $p < 0.01$, and *** for $p < 0.001$. Non-significant factors (p > 0.05) are also reported for completeness. Random effects were included to account for variability in task quality based on participants' self-selected decisions to reject or recommend the system, ensuring fair comparisons across all fixed factors.

Factor	Values	Coefficient	p-value
Intercept		2.795	0.000
Type of task			
Recommendation	Reject *vs.* Recommend	0.880	0.325
System	Plate Detector *vs.* Checkout	0.150	0.079
Participant's cohort			
Cohort	Developers *vs.* Ordinary individuals	0.131	0.257
Cohort	**Compliance experts *vs.* Ordinary individuals**	**0.287**	**0.007****
Expertise levels			
Task Expertise	Low *vs.* High	-0.013	0.819
Technological Expertise	Low *vs.* High	0.055	0.392
AI Expertise	Low *vs.* High	-0.056	0.382
Treatment			
Treatment type	**Card *vs.* Report**	**-0.987**	**0.000*****

Table 5. **The mean difference testing underscores the strong influence of treatment choice on task quality.** We conducted statistical significance testing on the mean differences between two factor values, presenting Mann-Whitney test p-values with the notations: * for $p < 0.05$, ** for $p < 0.01$, and *** for $p < 0.001$.

Factor	Value Pair	Averages	Difference	p-value
Type of task				
Recommendation	Reject *vs.* Recommend	3.0 *vs.* 3.014	-0.014	0.719
System	Plate Detector *vs.* Checkout	2.801 *vs.* 2.645	-0.156	0.139
Participant's cohort				
Cohort	Developers *vs.* Compliance experts	2.852 *vs.* 2.95	-0.098	0.534
Cohort	**Developers *vs.* Ordinary individuals**	**2.852 *vs.* 2.505**	**0.347**	**0.011***
Cohort	**Compliance experts *vs.* Ordinary individuals**	**2.95 *vs.* 2.505**	**0.445**	**0.002****
Expertise levels				
Task Expertise	Low *vs.* High	2.73 *vs.* 2.714	0.015	0.852
Technological Expertise	Low *vs.* High	2.691 *vs.* 2.851	-0.16	0.25
AI Expertise	Low *vs.* High	2.654 *vs.* 2.803	-0.149	0.204
Treatment				
Treatment type	**Card *vs.* Report**	**3.327 *vs.* 2.12**	**1.207**	**0.0*****

The card was favored for its clear, concise presentation, and quick comprehension of the risks and benefits of AI uses, though some found it overly simplistic for complex decisions. On the positive side, the card was favored for its concise and straightforward presentation of information. Participants found it easier to digest, with visual elements and organized sections that allowed for quick understanding of the main risks and benefits of the presented AI systems. For example, CP9 stated that *"[the card] assisted me by highlighting the risks, accuracy, and benefits,"* while CP23 appreciated *"the card's structured overview of the system's components, facilitating the identification of key technical aspects of the AI system"*. A compliance expert, CP88, mentioned that *"[the card] was easy to use and conveyed the gist of the AI system"*. Additionally, participants commented on the card's format to be *"readily accessible to refer back to"* (CP129, a compliance expert). Participants also echoed the sentiment that despite spending less time with the card, it even helped them produce emails of higher quality. CP190, an ordinary individual, commented that *"the best thing really is just that more thought went into making the card format more digestible and less intimidating, so that it would be easy to get what you need by reading it, without needing time to consult with more technical people to be sure you understand its material correctly"*. On the negative side, some participants noted that the card lacked the depth and detail found in the report. There were also mentions of the card being too simplistic for complex decision-making. CP6, a developer, felt that it *"was a little simple, so I can't help but think there may be something missing in the big picture"*. Despite its concise format, some participants found the card too brief. For example, CP9, a developer, commented that *"the card was brief which I enjoyed, however, it probably could have used a little more substance"*.

The report was valued for its depth and details, though its complexity and dense format challenged quick comprehension and accessibility. On the positive side, participants appreciated the report for its detailed and comprehensive information, which helped them understand the AI system better. They mentioned that the report laid out the pros and cons effectively in a structured way, providing a good foundation of knowledge. CP12, a developer, mentioned that it *"helped explain why the system should be implemented, was organized and listed many different positive aspects"*. Similarly, CP152, an ordinary individual, mentioned that *"the report succeeded in assisting me in completing the task by providing a wide array of information through which I could make a decision"*. On the negative side, a common critique was the report's complexity and length, making it difficult to quickly extract necessary information. Participants found it too detailed at times, with some sections containing excessive technical jargon. CP142, an ordinary individual, stated that *"the report was overly wordy. It had lots of irrelevant information and technical jargon (e.g., the datasets that the models it uses were trained on)."* Similarly, CP157, another ordinary individual, stated that *"the report felt very wordy. Reading it felt like I needed higher education to fully understand some aspects of the technology. I'm not sure if every day people would fully comprehend all the ins and outs of it"*. CP9, a developer, noted that *"the report didn't advise in any direction"*. Additionally, some participants called out the lack of visual elements, which impacted their ease of understanding. CP75, a compliance expert, noted that *"the report was too cluttered"*. Similarly, CP101, an ordinary individual, commented that *"there was lot information to read and some of the reading I didn't understand had to read at least twice"*.

Overall preference. The preference varied among participants. Some preferred the report for its thoroughness and detail, which they found necessary for making informed decisions. Others favored the card for its efficiency and simplicity, allowing for quicker comprehension and easier reference during their tasks. CP9, a developer, found the report more useful, stating that *"the report had much more information that I could use to craft the email"*. Similarly, CP78, a compliance expert, stated that a preference towards *"the report as it provided a more detailed information about the AI system, impacts, risks and mitigation strategies enabling a thorough analysis and recommendation"*. Conversely, CP152, an ordinary individual, mentioned that *"I prefer the card more than the report, because the card was more concise and clear in the information that it presented"* . Similarly, CP101, a compliance expert, preferred the card because *"it seemed easy to read and understand. It showed the entire plan like a minimalistic picture"*.

Card improvements. Participants also suggested ways to further improve the card by providing contextual information, reducing ambiguity, and enhancing its visual elements. Some participants mentioned that the cards were too simplistic and lacked the necessary depth for comprehensive understanding. For example, CP97, a compliance expert, mentioned that *"the card fell short of the optimum aid in completing the task because it did not provide a full explanation of some of the* *material presented"*. Similarly, CP71, another compliance expert, stated that *"the card was overall a helpful tool, but should provide more guidance on how to address complex aspects of the AI system"*. Participants also made recommendations for enhancing the card's visual elements. For example, CP13, a developer, commented that *"the legend at the bottom of each visualization should be moved closer to the top"*. Similarly, CP158, an ordinary individual, mentioned that *"just filling in the box without explaining what a filled out box meant was not useful. It would have been more useful to have a rating with explanation of each rating for each category"*. We addressed the comments about visual elements and provided the link to the latest versions of the card (Appendix A.6, Figures 11-12).

In our user study, we evaluated impact assessment cards for AI systems with a tangible presence in the physical world such as biometric checkout systems and license plate detectors. However, many algorithmic systems operate without a visible manifestation, for example, recommender systems or decision-support algorithms in public services and finance. To illustrate the adaptability of our card beyond physically situated AI, we created examples for two additional digital systems: a music recommender system [42] (Appendix A.7, Figure 13) and a housing benefit allocation assistant [81] (Appendix A.7, Figure 14). These examples demonstrate how our card can incorporate different visual elements and be adapted for AI systems that operate in the background, often without end users being fully aware of their presence.

6 Discussion

We begin by consolidating our findings on the use of impact assessment cards as tools for assessing AI risks, communicating AI benefits, and supporting AI governance (§6.1). Next, we explore opportunities to apply the cards in different contexts (§6.2), and conclude by discussing their limitations (§6.3).

6.1 Cards as Tools for Assessing AI Risks, Benefits, and Governance

The impact assessment card offers a new accessible medium for addressing the ethical and practical aspects of AI systems. Unlike detailed reports aimed at primarily technical audiences, our card can engage diverse stakeholders with a concise and visually appealing format. Next, we discuss three prospective applications of the card.

Assessing AI Risks. HCI and CSCW research has long emphasized the importance of tools that help stakeholders foresee potential failures, risks, and harms in technology design [13, 37, 47]. Our findings demonstrate that impact assessment cards enable stakeholders to identify, contextualize, and reflect on risks more effectively than traditional reports. Our participants engaged deeply with the content, contextualizing risks in relation to AI applications and mitigations. By democratizing access to risk-related discussions, impact assessment cards may also foster informed decision-making and civic engagement in AI governance [9, 67].

Communicating AI Benefits. The public discourse on AI often emphasizes risks, overshadowing potential benefits [63, 65]. Our card aims at addressing this imbalance by presenting benefits prominently alongside risks, drawing inspiration from fields such as medicine and energy communication [17]. Our participants valued this balanced perspective, which encouraged deep reflections on the dual aspects of AI systems. This approach aligns with ethical principles of informed decision-making, ensuring that AI is seen as a tool with both opportunities and challenges [48].

Supporting AI Governance. Existing governance tools (e.g., certification labels and audit frameworks) often target technically skilled users [11, 73]. In contrast, our card synthesizes complex audit information into an accessible format suitable for a broader audience. This design decision is driven by the need for better alignment between technical experts and the broader public in AI governance [24]. Experts benefit from a concise tool for communicating governance decisions, while non-experts gain a practical resource that simplifies regulatory concepts and clarifies their rights. For example, engineers in AI companies could use the card for internal communication, while regulators might adopt it to support compliance with frameworks such as the EU AI Act [20]. Moreover, the cards can empower legal and civil society organizations by providing them with a user-friendly tool to engage in advocacy, oversight, and accountability efforts. By bridging the gap between technical and non-technical audiences, the cards advance inclusivity in AI governance.

6.2 Cards Applied in Different Contexts

We view impact assessment cards as versatile tools that can be adapted to various domains and stakeholder needs. Next, we outline design opportunities and potential applications for the cards in four different contexts.

Participatory Design and Stakeholder Engagement. Participatory design methodologies (e.g., focus groups or co-design workshops [61]) can be used to further refine the cards, ensuring their relevance across diverse use cases. These activities can identify stakeholder-specific needs, ensuring that the resulting card addresses both direct and indirect impacts of AI systems [3]. For example, in healthcare, impact assessment cards could include risk-benefit information tailored to AI-assisted diagnosis tools, highlighting concerns such as data privacy and patient safety,t while showcasing benefits such as early detection of diseases. Similarly, in urban planning, the cards could map AI applications such as predictive traffic management, focusing on stakeholder groups such as residents, city planners, and policy makers.

Regulatory and Compliance Applications. Beyond summarizing risks and benefits, the cards could serve as templates for regulatory reporting, assisting organizations in mapping risks, mitigations, and benefits to regulatory requirements [82]. By integrating data from datasheets and model cards [29, 59], impact assessment cards can help ensure transparency and accountability in AI governance. For example, an AI company developing a recruitment algorithm might customize the card to include categories such as bias mitigation strategies, compliance with anti-discrimination laws, and transparency measures. Including visual markers such as checkboxes or compliance level indicators (e.g., high, medium, low) could enhance their practicality for audits and self-assessments. Moreover, the cards could be integrated into certification processes, serving as an interface between technical audits and public-facing labels. For example, an AI certification body might use cards to communicate whether a system adheres to standards of transparency, fairness, or energy efficiency.

Educational and Advocacy Tools. The cards may also serve as tools for education and public advocacy. In academic settings, they can introduce students to the societal implications of AI through a structured way to explore ethical dilemmas [79]. By presenting complex topics in an accessible format, the cards help bridge the gap between technical knowledge and societal considerations, making them ideal for discussions on AI ethics, governance, and responsible innovation. Advocacy organizations could also employ the cards in public engagement campaigns to facilitate community discussions on AI-related issues such as data privacy, bias, and surveillance. The cards' utility and reach could further expanded by integrating multimedia features such as QR codes linking to additional resources.

Industry-Specific Appropriations. Impact assessment cards can also be tailored to specific industries to address unique risks and benefits. In the financial sector, for example, they could evaluate AI-driven investment tools or fraud detection systems, focusing on transparency about decision-making criteria and biases. Similarly, in the energy and environment domain, the cards might highlight trade-offs in AI applications for renewable energy optimization, helping stakeholders balance gains in efficiency with risks related to system reliability and data accuracy.

Cards are also applicable across systems with varying levels of (physical) visibility. For example, physically situated AI systems (like our exemplary biometric checkout and license plate detector) can be seen as systems with material manifestation as they have a material presence in the physical world (at the point of checkout, or through cameras and CCTVs). On the contrary, AI systems without a material manifestation (e.g., music recommendation platforms [42], benefit allocation assistants [81]) operate entirely in digital environments where their presence is not tied to a physical location but rather integrated into software interfaces or cloud-based services. For systems with material manifestations, cards can provide clear information about data processing and privacy measures. For example, in a biometric checkout system that enables customers to make payments using facial recognition, cards could appear at key moments in the customer journey: during enrollment when users scan their face and link it to a payment method, or on receipts as a QR code to reinforce transparency after a transaction. Conversely, for systems without material manifestations such as recommender systems [42], cards can promote transparency about algorithms and biases. In platforms like Spotify or Netflix, cards could explain how recommendations are generated, including the use of data sources and personalization algorithms, and highlight any associated risks or biases (Appendix A.7 13). These cards could be integrated into digital touchpoints such as during account setup alongside terms and conditions, or embedded in the platform's navigation bar under sections like "About" or "Transparency". By positioning the cards strategically, users can easily access and understand how their data is used, fostering trust and accountability.

6.3 Limitations and Future Directions

Our study and the impact card have four main limitations that suggest directions for future research. First, its brevity may overlook the complexities of AI risks and benefits, requiring more research to adapt it for diverse real-world AI applications. Future designs could involve creating culturally varied card versions [92], or blending physical and digital forms with interactive elements for better risk and benefit understanding [26]. Despite their potential, we believe that cards are not a replacement for detailed reports, particularly in contexts requiring comprehensive evidence to substantiate compliance claims. Participants recommended simplifying language, summarizing key points, and incorporating visual aids to make reports more accessible. Future work could explore how hybrid tools—combining cards and reports—might balance accessibility and depth, further enhancing stakeholder engagement. Second, although the card received higher usability ratings from all cohorts, design improvements could further enhance its usability. The card's score partly reflects the challenge of our endeavor: to create a user-friendly tool that effectively communicates the risks and benefits of AI in a way that is accessible to individuals without technical expertise. In the future, we plan to broaden our engagement to include a more diverse group of stakeholders such as organizational leaders. Third, our study's sample may not completely represent all AI stakeholders like developers, compliance experts, and the ordinary individuals due to limited controls over participants' roles, AI use frequency, and location. While we recruited a sample of Prolific participants matching the US population, the findings and discussions should be interpreted with some limitations. For example, our study does not account for the recently updated standards introducing a combined race and ethnicity question where groups such as Hispanic or Latino are considered a co-equal category alongside the ethnicity categories we used [64]. We encourage future researchers to include additional ethnicity screeners when recruiting participants to improve representation.

Finally, AI students, crucial for future AI development [40], were not part of our cohorts. Inspired by research on the AI Incident Database's educational impact [21], we aim to integrate impact cards with incident reports in future studies to assess AI students' understanding of risks and benefits.

7 Conclusion

Through an iterative design process, we designed and evaluated an impact assessment card for communicating the risks and benefits of AI uses. The card summarizes detailed AI reports, presenting complex information in a clear and accessible way for both experts and laypeople. We evaluated our card's effectiveness in an online study with 235 participants across developers, compliance experts, and ordinary individuals. We found that the card's effectiveness extended beyond ordinary individuals, offering advantages to those who are well-versed in AI impact assessments. Moving forward, our work suggests a promising direction for further refining impact assessment cards, aiming to democratize understanding and participation [19] in AI risk assessment.

References

[1] Philip Achimugu, Ali Selamat, Roliana Ibrahim, and Mohd Naz'ri Mahrin. 2014. A Systematic Literature Review of Software Requirements Prioritization Research. *Information and Software Technology* 56, 6 (2014), 568–585. https://doi.org/10.1016/j.infsof.2014.02.001

[2] Ada Lovelace Institute. 2022. *Algorithmic Impact Assessment: AIA Template.* Retrieved January 22, 2024 from https://www.adalovelaceinstitute.org/resource/aia-template/

[3] Stephanie Ballard, Karen M. Chappell, and Kristen Kennedy. 2019. Judgment Call the Game: Using Value Sensitive Design and Design Fiction to Surface Ethical Concerns Related to Technology. In *Proceedings of the ACM Conference on Designing Interactive Systems (DIS).* 421–433. https://doi.org/10.1145/3322276.3323697

[4] Emily M. Bender and Batya Friedman. 2018. Data Statements for Natural Language Processing: Toward Mitigating System Bias and Enabling Better Science. *Transactions of the Association for Computational Linguistics (ACL)* 6 (2018), 587–604. https://doi.org/10.1162/tacl_a_00041

[5] Glen Berman, Nitesh Goyal, and Michael Madaio. 2024. A Scoping Study of Evaluation Practices for Responsible AI Tools: Steps Towards Effectiveness Evaluations. In *Proceedings of the ACM Conference on Human Factors in Computing Systems (CHI).* Article 294, 24 pages. https://doi.org/10.1145/3613904.3642398

[6] Vanessa Bracamonte, Sebastian Pape, Sascha Löbner, and Frederic Tronnier. 2023. Effectiveness and Information Quality Perception of an AI Model Card: A Study Among Non-Experts. In *Annual International Conference on Privacy, Security and Trust (PST).* IEEE, 1–7.

[7] Virginia Braun and Victoria Clarke. 2006. Using Thematic Analysis in Psychology. *Qualitative Research in Psychology* 3, 2 (2006), 77–101. https://doi.org/10.1191/1478088706qp063oa

[8] John Brooke. 1996. SUS: A Quick and Dirty Usability Scale. *Usability Evaluation in Industry* 189, 3 (1996), 189–194.

[9] Alexander Buhmann and Christian Fieseler. 2021. Towards a Deliberative Framework for Responsible Innovation in Artificial Intelligence. *Technology in Society* 64 (2021), 101475. https://doi.org/10.1016/j.techsoc.2020.101475

[10] A. Burns, C. Xiong, S. Franconeri, A. Cairo, and N. Mahyar. 2020. How To Evaluate Data Visualizations Across Different Levels of Understanding. In *IEEE Workshop on Evaluation and Beyond - Methodological Approaches to Visualization (BELIV).* IEEE Computer Society, 19–28. https://doi.org/10.1109/BELIV51497.2020.00010

[11] Zana Buçinca, Chau Minh Pham, Maurice Jakesch, Marco Tulio Ribeiro, Alexandra Olteanu, and Saleema Amershi. 2023. AHA!: Facilitating AI Impact Assessment by Generating Examples of Harms. arXiv:2306.03280

[12] Peter Cihon, Moritz J. Kleinaltenkamp, Jonas Schuett, and Seth D. Baum. 2021. AI Certification: Advancing Ethical Practice by Reducing Information Asymmetries. *IEEE Transactions on Technology and Society* 2, 4 (2021), 200–209. https://doi.org/10.1109/tts.2021.3077595

[13] Marios Constantinides, Edyta Bogucka, Daniele Quercia, Susanna Kallio, and Mohammad Tahaei. 2024. RAI Guidelines: Method for Generating Responsible AI Guidelines Grounded in Regulations and Usable by (Non-)Technical Roles. *Proceedings of the ACM on Human-Computer Interaction* CSCW (2024), 1–28. https://doi.org/10.1145/3686927

[14] Credo AI. 2024. *AI Vendor Risk Profiles.* Retrieved November 22, 2024 from https://www.credo.ai/ai-vendor-directory

[15] Andrew Gary Darwin Holmes. 2020. Researcher Positionality - A Consideration of Its Influence and Place in Qualitative Research - A New Researcher Guide. *Shanlax International Journal of Education* 8, 4 (2020), 1–10. https://doi.org/10.34293/education.v8i4.3232

[16] Alicia DeVos, Aditi Dhabalia, Hong Shen, Kenneth Holstein, and Motahhare Eslami. 2022. Toward User-Driven Algorithm Auditing: Investigating Users' Strategies for Uncovering Harmful Algorithmic Behavior. In *Proceedings of the ACM Conference on Human Factors in Computing Systems (CHI).* Article 626, 19 pages. https://doi.org/10.1145/3491102.3517441

[17] David Dickinson and Suzy Gallina. 2017. *Information Design: Research and Practice.* Routledge, Chapter Information Design in Medicine Package Leaflets. https://doi.org/10.4324/9781315585680

[18] Salma Elsayed-Ali, Sara E. Berger, Vagner Figueredo De Santana, and Juana Catalina Becerra Sandoval. 2023. Responsible & Inclusive Cards: An Online Card Tool to Promote Critical Reflection in Technology Industry Work Practices. In *Proceedings of the ACM Conference on Human Factors in Computing Systems (CHI)*. Article 5, 14 pages. https://doi.org/10.1145/3544548.3580771

[19] Eva Erman and Markus Furendal. 2024. The Democratization of Global AI Governance and the Role of Tech Companies. *Nature Machine Intelligence* (2024), 1–3. https://doi.org/10.1038/s42256-024-00811-z

[20] 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).* Retrieved June 13, 2024 from https://www.europarl.europa.eu/doceo/document/TA-9-2024-0138-FNL-COR01_EN.pdf

[21] Michael Feffer, Nikolas Martelaro, and Hoda Heidari. 2023. The AI Incident Database as an Educational Tool to Raise Awareness of AI Harms: A Classroom Exploration of Efficacy, Limitations, & Future Improvements. In *Proceedings of the ACM Conference on Equity and Access in Algorithms, Mechanisms, and Optimization (EAAMO)*. 1–11. https://doi.org/10.1145/3617694.3623223

[22] Almudena Fernández-Fontecha, Kay L O'Halloran, Sabine Tan, and Peter Wignell. 2019. A Multimodal Approach to Visual Thinking: The Scientific Sketchnote. *Visual Communication* 18, 1 (2019), 5–29. https://doi.org/10.1177/1470357218759808

[23] Figma. 2016. *Figma: The Collaborative Interface Design Tool.* Retrieved February 22, 2024 from https://www.figma.com

[24] Luciano Floridi, Josh Cowls, Monica Beltrametti, Raja Chatila, Patrice Chazerand, Virginia Dignum, Christoph Luetge, Robert Madelin, Ugo Pagallo, Francesca Rossi, et al. 2018. AI4People—an Ethical Framework for a Good AI Society: Opportunities, Risks, Principles, and Recommendations. *Minds and Machines* 28 (2018), 689–707. https://doi.org/10.1007/s11023-018-9482-5

[25] Riccardo Fogliato, Alexandra Chouldechova, and Zachary Lipton. 2021. THe Impact of Algorithmic Risk Assessments on Human Predictions and Its Analysis via Crowdsourcing Studies. *Proceedings of the ACM on Human-Computer Interaction* 5, CSCW2 (2021), 1–24. https://doi.org/10.1145/3479572

[26] Steven L. Franconeri, Lace M. Padilla, Priti Shah, Jeffrey M. Zacks, and Jessica Hullman. 2021. The Science of Visual Data Communication: What Works. *Psychological Science in the Public Interest* 22, 3 (2021), 110–161. https://doi.org/10.1177/15291006211051956 PMID: 34907835.

[27] A. Gaba, Z. Kaufman, J. Cheung, M. Shvakel, K. m. Hall, Y. Brun, and C. Bearfield. 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 (TVCG)* 30, 01 (2024), 327–337. https://doi.org/10.1109/TVCG.2023.3327192

[28] A. Gaba, V. Setlur, A. Srinivasan, J. Hoffswell, and C. Xiong. 2023. Comparison Conundrum and the Chamber of Visualizations: An Exploration of How Language Influences Visual Design. *IEEE Transactions on Visualization and Computer Graphics (TVCG)* 29, 01 (2023), 1211–1221. https://doi.org/10.1109/TVCG.2022.3209456

[29] Timnit Gebru, Jamie Morgenstern, Briana Vecchione, Jennifer Wortman Vaughan, Hanna Wallach, Hal Daumé III, and Kate Crawford. 2021. Datasheets for Datasets. *Commun. ACM* 64, 12 (2021), 86–92. https://doi.org/10.1145/3458723

[30] Delaram Golpayegani, Harshvardhan J. Pandit, and Dave Lewis. 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 (FAccT)*. 905–915. https://doi.org/10.1145/3593013.3594050

[31] Samantha Goodman, Lana Vanderlee, Rachel Acton, Syed Mahamad, and David Hammond. 2018. The Impact of Front-of-Package Label Design on Consumer Understanding of Nutrient Amounts. *Nutrients* 10, 11 (2018), 1624. https://doi.org/10.3390/nu10111624

[32] Matthew Gorton, Barbara Tocco, Ching-Hua Yeh, and Monika Hartmann. 2021. What Determines Consumers' Use of Eco-Labels? Taking a Close Look at Label Trust. *Ecological Economics* 189 (2021), 107173. https://doi.org/10.1016/j.ecolecon.2021.107173

[33] Ben Green and Yiling Chen. 2021. Algorithmic Risk Assessments Can Alter Human Decision-Making Processes in High-Stakes Government Contexts. *Proceedings of the ACM on Human-Computer Interaction* 5, CSCW2 (2021), 1–33. https://doi.org/10.1145/3479562

[34] Klaus G. Grunert, Sophie Hieke, and Josephine Wills. 2014. Sustainability Labels on Food Products: Consumer Motivation, Understanding and Use. *Food Policy* 44 (2014), 177–189. https://doi.org/10.1016/j.foodpol.2013.12.001

[35] Clara Hainsdorf, Tim Hickman, Sylvia Lorenz, and Jenna Rennie. 2023. *Dawn of the EU's AI Act: Political Agreement Reached on World's First Comprehensive Horizontal AI Regulation.* White & Case. Retrieved December 14, 2023 from https://www.whitecase.com/insight-alert/dawn-eus-ai-act-political-agreement-reached-worlds-first-comprehensive-horizontal-ai

[36] Sarah Holland, Ahmed Hosny, Sarah Newman, Joshua Joseph, and Kasia Chmielinski. 2020. The Dataset Nutrition Label: A Framework to Drive Higher Data Quality Standards. In *Data Protection and Privacy*, Dara Hallinan, Ronald Leenes,

Serge Gutwirth, and Paul De Hert (Eds.). Hart Publishing, Chapter 1, 1–26. https://doi.org/10.5040/9781509932771.ch-001

[37] Matthew K. Hong, Adam Fourney, Derek DeBellis, and Saleema Amershi. 2021. Planning for Natural Language Failures With the AI Playbook. In *Proceedings of the ACM Conference on Human Factors in Computing Systems (CHI)*. 1–11. https://doi.org/10.1145/3411764.3445735

[38] Isabelle Hupont, David Fernández-Llorca, Sandra Baldassarri, and Emilia Gómez. 2024. Use Case Cards: A Use Case Reporting Framework Inspired by the European AI Act. *Ethics and Information Technology* 26, 2 (March 2024). https://doi.org/10.1007/s10676-024-09757-7

[39] ISO/IEC. 2023. *Information Technology – Artificial Intelligence – Management System*. Standard ISO/IEC 42001:2023. International Organization for Standardization. https://www.iso.org/standard/81230.html

[40] Nari Johnson and Hoda Heidari. 2023. Assessing AI Impact Assessments: A Classroom Study. arXiv:2311.11193

[41] Alexandra Jones, Bruce Neal, Belinda Reeve, Cliona Ni Mhurchu, and Anne Marie Thow. 2019. Front-of-Pack Nutrition Labelling To Promote Healthier Diets: Current Practice and Opportunities To Strengthen Regulation Worldwide. *BMJ Global Health* 4, 6 (2019). https://doi.org/10.1136/bmjgh-2019-001882

[42] Marius Kaminskas, Francesco Ricci, and Markus Schedl. 2013. Location-Aware Music Recommendation Using Auto-Tagging and Hybrid Matching. In *Proceedings of the ACM Conference on Recommender Systems (RecSys)* (Hong Kong, China) *(RecSys '13)*. ACM, 17–24. https://doi.org/10.1145/2507157.2507180

[43] Anna Kawakami, Shreya Chowdhary, Shamsi T. Iqbal, Q. Vera Liao, Alexandra Olteanu, Jina Suh, and Koustuv Saha. 2023. Sensing Wellbeing in the Workplace, Why and for Whom? Envisioning Impacts With Organizational Stakeholders. *Proceedings of the ACM on Human-Computer Interaction* 7, CSCW2 (2023), 1–33. https://doi.org/10.1145/3610207

[44] Anna Kawakami, Daricia Wilkinson, and Alexandra Chouldechova. 2024. Do Responsible AI Artifacts Advance Stakeholder Goals? Four Key Barriers Perceived by Legal and Civil Stakeholders. In *Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society*, Vol. 7. 670–682. https://doi.org/10.1609/aies.v7i1.31669

[45] Johannes Kehrer and Helwig Hauser. 2012. Visualization and Visual Analysis of Multifaceted Scientific Data: A Survey. *IEEE Transactions on Visualization and Computer Graphics (TVCG)* 19, 3 (2012), 495–513. https://doi.org/10.1109/TVCG.2012.110

[46] Pranav Khadpe, Ranjay Krishna, Li Fei-Fei, Jeffrey T. Hancock, and Michael S Bernstein. 2020. Conceptual Metaphors Impact Perceptions of Human-AI Collaboration. *Proceedings of the ACM on Human-Computer Interaction* 4, CSCW2 (2020), 1–26. https://doi.org/10.1145/3415234

[47] Sandjar Kozubaev, Chris Elsden, Noura Howell, Marie Louise Juul Søndergaard, Nick Merrill, Britta Schulte, and Richmond Y. Wong. 2020. Expanding Modes of Reflection in Design Futuring. In *Proceedings of the ACM Conference on Human Factors in Computing Systems (CHI)*. 1–15. https://doi.org/10.1145/3313831.3376526

[48] Jobst Landgrebe and Barry Smith. 2022. *Why Machines Will Never Rule the World: Artificial Intelligence Without Fear*. Routledge.

[49] Tianshi Li, Lorrie Faith Cranor, Yuvraj Agarwal, and Jason I. Hong. 2024. Matcha: An IDE Plugin for Creating Accurate Privacy Nutrition Labels. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol. (UbiComp)* 8, 1, Article 33 (2024), 38 pages. https://doi.org/10.1145/3643544

[50] Weixin Liang, Nazneen Rajani, Xinyu Yang, Ezinwanne Ozoani, Eric Wu, Yiqun Chen, Daniel Scott Smith, and James Zou. 2024. What's Documented in AI? Systematic Analysis of 32K AI Model Cards. *arXiv preprint arXiv:2402.05160* (2024).

[51] Q. Vera Liao and S. Shyam Sundar. 2022. Designing for Responsible Trust in AI Systems: A Communication Perspective. In *Proceedings of the ACM Conference on Fairness, Accountability, and Transparency (FAccT)*. 1257–1268. https://doi.org/10.1145/3531146.3533182

[52] Joseph Lindley, Haider Ali Akmal, Franziska Pilling, and Paul Coulton. 2020. Researching AI Legibility Through Design. In *Proceedings of the ACM Conference on Human Factors in Computing Systems (CHI)*. 1–13. https://doi.org/10.1145/3313831.3376792

[53] Ewa Luger, Lachlan Urquhart, Tom Rodden, and Michael Golembewski. 2015. Playing the Legal Card: Using Ideation Cards to Raise Data Protection Issues within the Design Process. In *Proceedings of the ACM Conference on Human Factors in Computing Systems (CHI)*. ACM, 457–466. https://doi.org/10.1145/2702123.2702142

[54] P. Mantri, H. Subramonyam, A. L. Michal, and C. Xiong. 2023. How Do Viewers Synthesize Conflicting Information from Data Visualizations? *IEEE Transactions on Visualization & Computer Graphics (TVCG)* 29, 01 (2023), 1005–1015. https://doi.org/10.1109/TVCG.2022.3209467

[55] Nora McDonald, Sarita Schoenebeck, and Andrea Forte. 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 (2019). https://doi.org/10.1145/3359174

[56] Jacob Metcalf, Emanuel Moss, Elizabeth Anne Watkins, Ranjit Singh, and Madeleine Clare Elish. 2021. Algorithmic Impact Assessments and Accountability: The Co-Construction of Impacts. In *Proceedings of the ACM Conference on*

Fairness, Accountability, and Transparency (FAccT). ACM, 735–746. https://doi.org/10.1145/3442188.3445935
[57] Microsoft. 2022. *Microsoft Responsible AI Impact Assessment Template*. Retrieved January 22, 2024 from https://blogs.microsoft.com/wp-content/uploads/prod/sites/5/2022/06/Microsoft-RAI-Impact-Assessment-Template.pdf
[58] Matthew Miles and Michael Huberman. 1994. *Qualitative Data Analysis: A Methods Sourcebook*. Sage.
[59] Margaret Mitchell, Simone Wu, Andrew Zaldivar, Parker Barnes, Lucy Vasserman, Ben Hutchinson, Elena Spitzer, Inioluwa D. Raji, and Timnit Gebru. 2019. Model Cards for Model Reporting. In *Proceedings of the ACM Conference on Fairness, Accountability, and Transparency (FAccT)*. 220–229. https://doi.org/10.1145/3287560.3287596
[60] Katharina J. Morik, Helena Kotthaus, Raphael Fischer, Sascha Mücke, Matthias Jakobs, Nico Piatkowski, Andreas Pauly, Lukas Heppe, and Danny Heinrich. 2022. Yes We Care!-Certification for Machine Learning Methods Through the Care Label Framework. *Frontiers in Artificial Intelligence* 5 (2022). https://doi.org/10.3389/frai.2022.975029
[61] Michael J. Muller and Sarah Kuhn. 1993. Participatory Design. *Commun. ACM* 36, 6 (1993), 24–28.
[62] National Institute of Standards and Technology. 2023. *The EqualAI Algorithmic Impact Assessment Tool*. Retrieved January 2024 from https://www.equalai.org/aia/
[63] Claudio Novelli, Federico Casolari, Antonino Rotolo, Mariarosaria Taddeo, and Luciano Floridi. 2023. Taking AI Risks Seriously: A New Assessment Model for the AI Act. *AI & Society* (2023), 1–5. https://doi.org/10.1007/s00146-023-01723-z
[64] Office of Management and Budget of the United States Government. 2024. *Revisions to OMB's Statistical Policy Directive No. 15: Standards for Maintaining, Collecting, and Presenting Federal Data on Race and Ethnicity* . Retrieved December 2, 2024 from https://www.federalregister.gov/documents/2024/03/29/2024-06469/revisions-to-ombs-statistical-policy-directive-no-15-standards-for-maintaining-collecting-and
[65] Victor Ojewale, Ryan Steed, Briana Vecchione, Abeba Birhane, and Inioluwa Deborah Raji. 2024. Towards AI Accountability Infrastructure: Gaps and Opportunities in AI Audit Tooling. *arXiv preprint arXiv:2402.17861* (2024).
[66] Open Ethics. 2023. *Open Ethics Label: AI Nutrition Labels*. Retrieved June 30, 2024 from https://openethics.ai/label
[67] Vinodkumar Prabhakaran, Margaret Mitchell, Timnit Gebru, and Iason Gabriel. 2022. A Human Rights-Based Approach to Responsible AI. *arXiv preprint arXiv:2210.02667* (2022).
[68] Prolific. 2014. *Prolific: Quickly find research participants you can trust*. Retrieved March 05, 2024 from https://www.prolific.com
[69] Ghulam Jilani Quadri, Arran Zeyu Wang, Zhehao Wang, Jennifer Adorno, Paul Rosen, and Danielle Albers Szafir. 2024. Do You See What I See? A Qualitative Study Eliciting High-Level Visualization Comprehension. In *Proceedings of the ACM Conference on Human Factors in Computing Systems (CHI)*. ACM, Article 204, 26 pages. https://doi.org/10.1145/3613904.3642813
[70] Malak Sadek, Marios Constantinides, Daniele Quercia, and Celine Mougenot. 2024. Guidelines for Integrating Value Sensitive Design in Responsible AI Toolkits. *Proceedings of the ACM Conference on Human Factors in Computing Systems (CHI)*, Article 472 (2024), 20 pages. https://doi.org/10.1145/3613904.3642810
[71] Johnny Saldaña. 2015. *The Coding Manual for Qualitative Researchers*. Sage.
[72] Jeff Sauro. 2011. *A Practical Guide to the System Usability Scale: Background, Benchmarks & Best Practices*. Measuring Usability LLC.
[73] Nicolas Scharowski, Michaela Benk, Swen J. Kühne, Léane Wettstein, and Florian Brühlmann. 2023. Certification Labels for Trustworthy AI: Insights From an Empirical Mixed-Method Study. In *Proceedings of the ACM Conference on Fairness, Accountability, and Transparency (FAccT)*. 248–260. https://doi.org/10.1145/3593013.3593994
[74] Hong Shen, Alicia DeVos, Motahhare Eslami, and Kenneth Holstein. 2021. Everyday Algorithm Auditing: Understanding the Power of Everyday Users in Surfacing Harmful Algorithmic Behaviors. *Proceedings of the ACM on Human-Computer Interaction* 5, CSCW2, Article 433 (2021), 29 pages. https://doi.org/10.1145/3479577
[75] Hong Shen, Haojian Jin, Ángel Alexander Cabrera, Adam Perer, Haiyi Zhu, and Jason I Hong. 2020. Designing Alternative Representations of Confusion Matrices To Support Non-Expert Public Understanding of Algorithm Performance. *Proceedings of the ACM on Human-Computer Interaction* 4, CSCW2 (2020), 1–22.
[76] Eli Sherman and Ian Eisenberg. 2024. AI Risk Profiles: A Standards Proposal for Pre-deployment AI Risk Disclosures. , 23047–23052 pages. https://doi.org/10.1609/aaai.v38i21.30348
[77] Marcel Stadelmann and Renate Schubert. 2018. How Do Different Designs of Energy Labels Influence Purchases of Household Appliances? A Field Study in Switzerland. *Ecological Economics* 144 (2018), 112–123. https://doi.org/10.1016/j.ecolecon.2017.07.031
[78] Bernd Carsten Stahl, Josephina Antoniou, Nitika Bhalla, Laurence Brooks, Philip Jansen, Blerta Lindqvist, Alexey Kirichenko, Samuel Marchal, Rowena Rodrigues, Nicole Santiago, Zuzanna Warso, and David Wright. 2023. A Systematic Review of Artificial Intelligence Impact Assessments. *Artificial Intelligence Review* 56, 11 (2023), 12799–12831. https://doi.org/10.1007/s10462-023-10420-8
[79] Ioannis Stavrakakis, Damian Gordon, Brendan Tierney, Anna Becevel, Emma Murphy, Gordana Dodig-Crnkovic, Radu Dobrin, Viola Schiaffonati, Cristina Pereira, Svetlana Tikhonenko, et al. 2021. The Teaching of Computer Ethics on

Computer Science and Related Degree Programmes. A European Survey. *International Journal of Ethics Education* (2021), 1–29. https://doi.org/10.1007/s40889-021-00135-1

[80] Kees Stuurman and Eric Lachaud. 2021. Regulating AI. A Label To Complete the Newly Proposed Act on Artificial Intelligence. *SSRN Electronic Journal* (2021). https://doi.org/10.2139/ssrn.3963890

[81] Ningjing Tang, Jiayin Zhi, Tzu-Sheng Kuo, Calla Kainaroi, Jeremy J Northup, Kenneth Holstein, Haiyi Zhu, Hoda Heidari, and Hong Shen. 2024. AI Failure Cards: Understanding and Supporting Grassroots Efforts to Mitigate AI Failures in Homeless Services. In *Proceedings of the ACM Conference on Fairness, Accountability, and Transparency (FAccT)*. 713–732.

[82] The Future of Life Institute. 2024. *EU AI Act Compliance Checker.* The Future of Life Institute. Retrieved June 13, 2024 from https://artificialintelligenceact.eu/assessment/eu-ai-act-compliance-checker

[83] Emma Tonkin, Annabelle M. Wilson, John Coveney, Trevor Webb, and Samantha B. Meyer. 2015. Trust in and Through Labelling–a Systematic Review and Critique. *British Food Journal* 117, 1 (2015), 318–338. https://doi.org/10.1108/BFJ-07-2014-0244

[84] Twilio Inc. 2023. *AI Nutrition Facts.* Retrieved June 30, 2024 from https://nutrition-facts.ai

[85] United Nations. 2023. *The 17 Sustainable Development Goals.* Retrieved November 8, 2023 from https://sdgs.un.org/goals

[86] Scientific United Nations Educational and Cultural Organization. 2023. *Ethical Impact Assessment. A Tool of the Recommendation on the Ethics of Artificial Intelligence.* Recommendation SHS/REI/BIO/REC-AIETHICS-TOOL-EIA/2023. United Nations Educational, Scientific and Cultural Organization, Paris. https://doi.org/10.54678/YTSA7796

[87] U.S. Census Bureau. 2021. *Race and Ethnicity in the United States: 2010 Census and 2020 Census.* https://www.census.gov/library/visualizations/interactive/race-and-ethnicity-in-the-united-state-2010-and-2020-census.html

[88] U.S. Census Bureau. 2022. *Population by Age and Sex. Annual Social and Economic Supplement.* https://www.census.gov/library/visualizations/interactive/race-and-ethnicity-in-the-united-state-2010-and-2020-census.html

[89] U.S. Office of Management and Budget. 1997. *Revisions to the Standards for the Classification of Federal Data on Race and Ethnicity.* Retrieved February 22, 2025 from https://obamawhitehouse.archives.gov/omb/fedreg_1997standards

[90] Zijie J. Wang, Chinmay Kulkarni, Lauren Wilcox, Michael Terry, and Michael Madaio. 2024. Farsight: Fostering Responsible AI Awareness During AI Application Prototyping. In *Proceedings of the ACM Conference on Human Factors in Computing Systems (CHI)*. Article 976, 40 pages. https://doi.org/10.1145/3613904.3642335

[91] Laura Weidinger, Maribeth Rauh, Nahema Marchal, Arianna Manzini, Lisa Anne Hendricks, Juan Mateos-Garcia, Stevie Bergman, Jackie Kay, Conor Griffin, Ben Bariach, Iason Gabriel, Verena Rieser, and William Isaac. 2023. Sociotechnical Safety Evaluation of Generative AI Systems. arXiv:2310.11986

[92] Michael S. Wogalter, Christopher B. Mayhorn, and Olga A. Zielinska. 2015. *Use of Color in Warnings.* Cambridge University Press, 377–400.

[93] Stephen L. Young and Michael S. Wogalter. 1990. Comprehension and Memory of Instruction Manual Warnings: Conspicuous Print and Pictorial Icons. *Human Factors* 32, 6 (1990), 637–649. https://doi.org/10.1177/001872089003200603

[94] Natalie Zelenka, Nina Di Cara, and Huw Day. 2021. *Data Hazard Labels.* Retrieved March 12, 2024 from https://datahazards.com/index.html

A Appendix

A.1 Design Patterns for Communicating Risks and Benefits of AI Uses

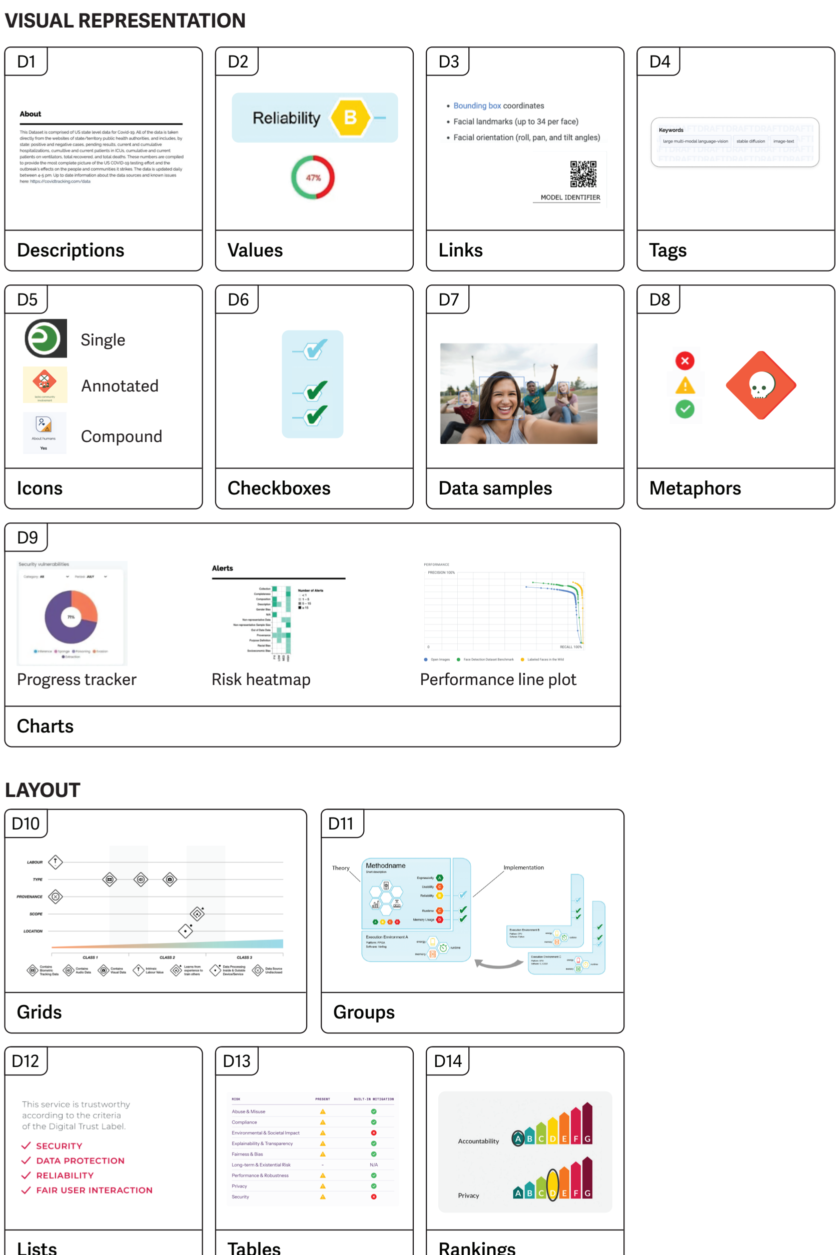


Fig. 6. Fourteen design patterns for **visual representation** (D1-D9) and **layout** (D10-D14) to communicate the risks and benefits of AI technologies, derived from the literature review [12, 32, 36, 41, 52, 60, 65, 77, 80, 94]. Available also at https://anonymous.4open.science/r/AIIA_Card/README.md#design-patterns.

A.2 Impact Assessment Cards

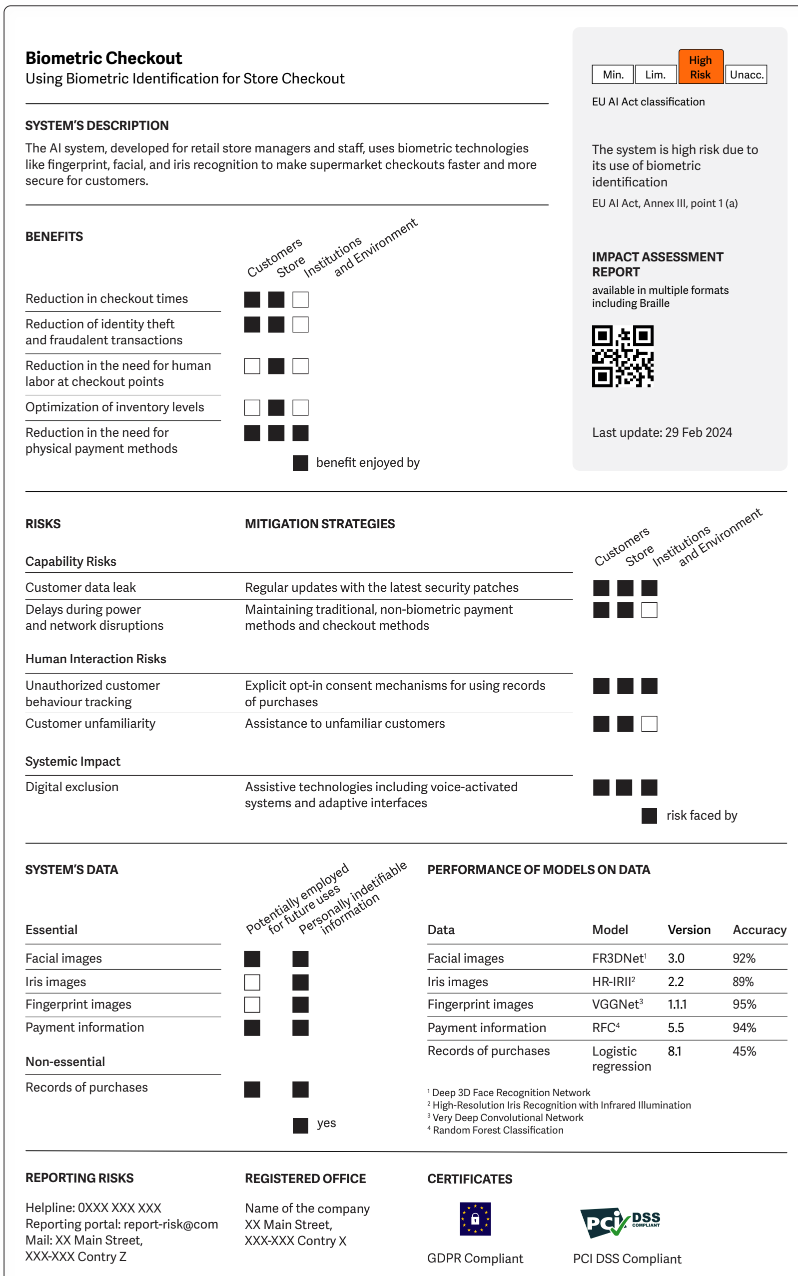


Fig. 7. Impact assessment card for a store checkout system using biometric identification, used during the large-scale online study. Available also at https://anonymous.4open.science/r/AIIA_Card/impact-assessment-card-biometric-checkout.pdf.

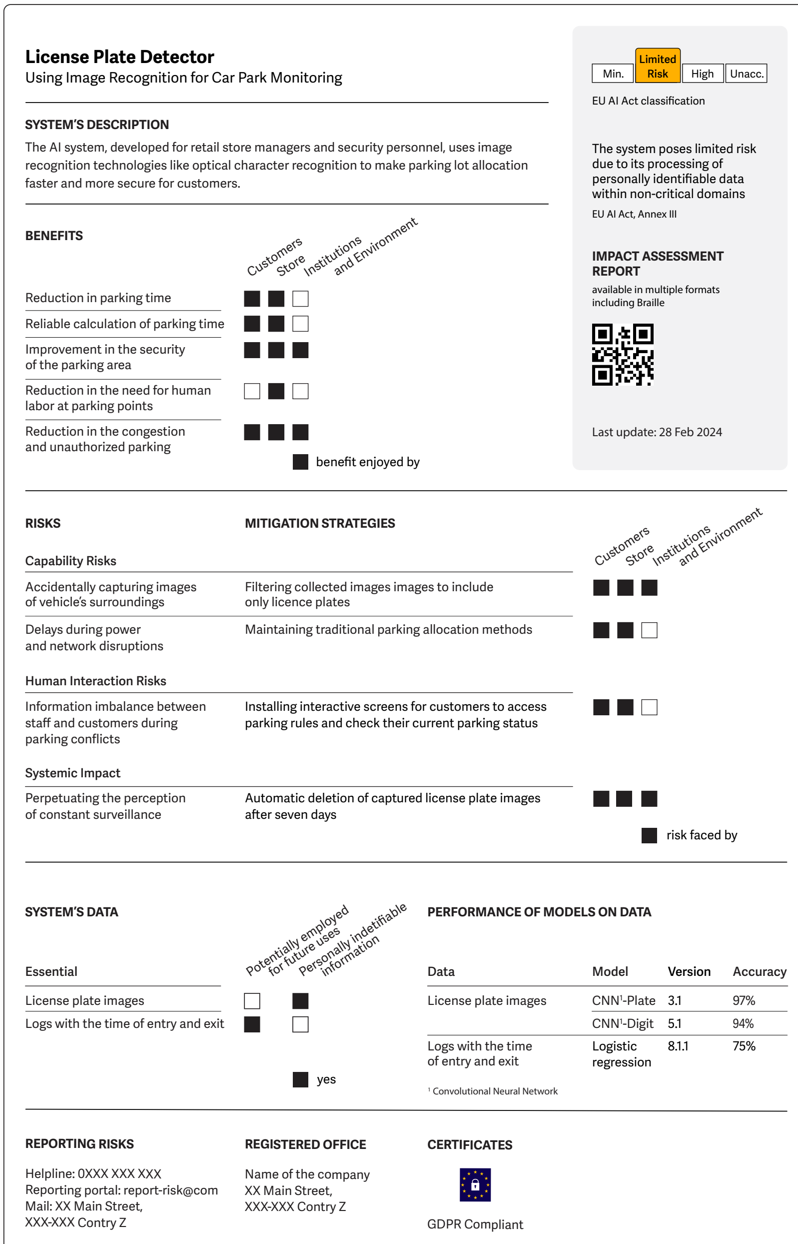
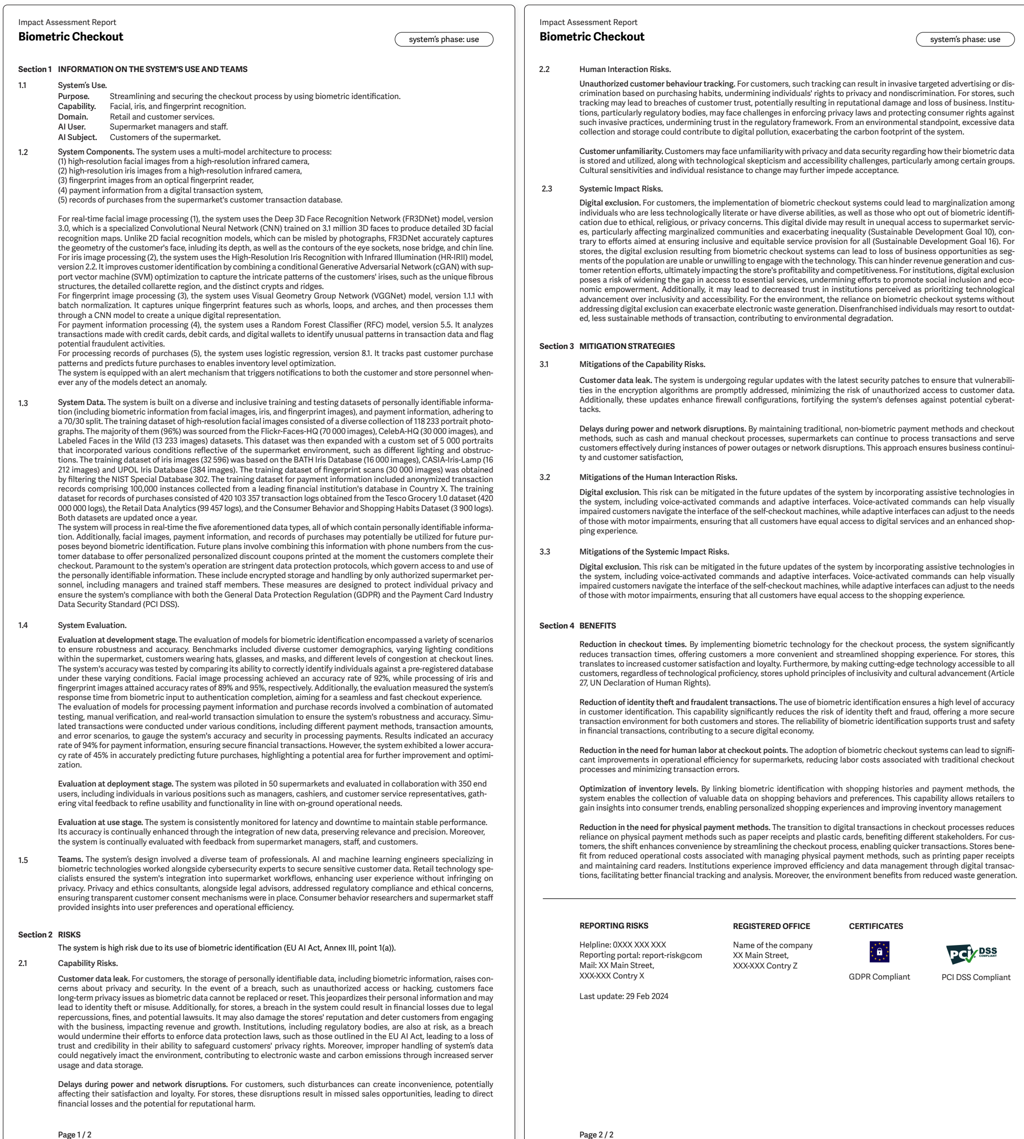


Fig. 8. Impact assessment card for a car park monitoring system using image recognition, used during the large-scale online study. Available also at https://anonymous.4open.science/r/AIIA_Card/impact-assessment-card-license-plate-detector.pdf.

A.3 Impact Assessment Reports



Impact Assessment Report

Biometric Checkout

system's phase: use

Section 1 INFORMATION ON THE SYSTEM'S USE AND TEAMS

1.1 System's Use.

Purpose. Streamlining and securing the checkout process by using biometric identification.
Capability. Facial, iris, and fingerprint recognition.
Domain. Retail and customer services.
AI User. Supermarket managers and staff.
AI Subject. Customers of the supermarket.

1.2 **System Components.** The system uses a multi-model architecture to process:
(1) high-resolution facial images from a high-resolution infrared camera,
(2) high-resolution iris images from a high-resolution infrared camera,
(3) fingerprint images from an optical fingerprint reader,
(4) payment information from a digital transaction system,
(5) records of purchases from the supermarket's customer transaction database.

For real-time facial image processing (1), the system uses the Deep 3D Face Recognition Network (FR3DNet) model, version 3.0, which is a specialized Convolutional Neural Network (CNN) trained on 3.1 million 3D faces to produce detailed 3D facial recognition maps. Unlike 2D facial recognition models, which can be misled by photographs, FR3DNet accurately captures the geometry of the customer's face, inluding its depth, as well as the contours of the eye sockets, nose bridge, and chin line.
For iris image processing (2), the system uses the High-Resolution Iris Recognition with Infrared Illumination (HR-IRII) model, version 2.2. It improves customer identification by combining a conditional Generative Adversarial Network (cGAN) with support vector machine (SVM) optimization to capture the intricate patterns of the customers' irises, such as the unique fibrous structures, the detailed collarette region, and the distinct crypts and ridges.
For fingerprint image processing (3), the system uses Visual Geometry Group Network (VGGNet) model, version 1.1.1 with batch normalization. It captures unique fingerprint features such as whorls, loops, and arches, and then processes them through a CNN model to create a unique digital representation.
For payment information processing (4), the system uses a Random Forest Classifier (RFC) model, version 5.5. It analyzes transactions made with credit cards, debit cards, and digital wallets to identify unusual patterns in transaction data and flag potential fraudulent activities.
For processing records of purchases (5), the system uses logistic regression, version 8.1. It tracks past customer purchase patterns and predicts future purchases to enables inventory level optimization.
The system is equipped with an alert mechanism that triggers notifications to both the customer and store personnel whenever any of the models detect an anomaly.

1.3 **System Data.** The system is built on a diverse and inclusive training and testing datasets of personally identifiable information (including biometric information from facial images, iris, and fingerprint images), and payment information, adhering to a 70/30 split. The training dataset of high-resolution facial images consisted of a diverse collection of 118 233 portrait photographs. The majority of them (96%) was sourced from the Flickr-Faces-HQ (70 000 images), CelebA-HQ (30 000 images), and Labeled Faces in the Wild (13 233 images) datasets. This dataset was then expanded with a custom set of 5 000 portraits that incorporated various conditions reflective of the supermarket environment, such as different lighting and obstructions. The training dataset of iris images (32 596) was based on the BATH Iris Database (16 000 images), CASIA-Iris-Lamp (16 212 images) and UPOL Iris Database (384 images). The training dataset of fingerprint scans (30 000 images) was obtained by filtering the NIST Special Database 302. The training dataset for payment information included anonymized transaction records comprising 100,000 instances collected from a leading financial institution's database in Country X. The training dataset for records of purchases consisted of 420 103 357 transaction logs obtained from the Tesco Grocery 1.0 dataset (420 000 000 logs), the Retail Data Analytics (99 457 logs), and the Consumer Behavior and Shopping Habits Dataset (3 900 logs). Both datasets are updated once a year.
The system will process in real-time the five aforementioned data types, all of which contain personally identifiable information. Additionally, facial images, payment information, and records of purchases may potentially be utilized for future purposes beyond biometric identification. Future plans involve combining this information with phone numbers from the customer database to offer personalized personalized discount coupons printed at the moment the customers complete their checkout. Paramount to the system's operation are stringent data protection protocols, which govern access to and use of the personally identifiable information. These include encrypted storage and handling by only authorized supermarket personnel, including managers and trained staff members. These measures are designed to protect individual privacy and ensure the system's compliance with both the General Data Protection Regulation (GDPR) and the Payment Card Industry Data Security Standard (PCI DSS).

1.4 System Evaluation.

Evaluation at development stage. The evaluation of models for biometric identification encompassed a variety of scenarios to ensure robustness and accuracy. Benchmarks included diverse customer demographics, varying lighting conditions within the supermarket, customers wearing hats, glasses, and masks, and different levels of congestion at checkout lines. The system's accuracy was tested by comparing its ability to correctly identify individuals against a pre-registered database under these varying conditions. Facial image processing achieved an accuracy rate of 92%, while processing of iris and fingerprint images attained accuracy rates of 89% and 95%, respectively. Additionally, the evaluation measured the system's response time from biometric input to authentication completion, aiming for a seamless and fast checkout experience.
The evaluation of models for processing payment information and purchase records involved a combination of automated testing, manual verification, and real-world transaction simulation to ensure the system's robustness and accuracy. Simulated transactions were conducted under various conditions, including different payment methods, transaction amounts, and error scenarios, to gauge the system's accuracy and security in processing payments. Results indicated an accuracy rate of 94% for payment information, ensuring secure financial transactions. However, the system exhibited a lower accuracy rate of 45% in accurately predicting future purchases, highlighting a potential area for further improvement and optimization.

Evaluation at deployment stage. The system was piloted in 50 supermarkets and evaluated in collaboration with 350 end users, including individuals in various positions such as managers, cashiers, and customer service representatives, gathering vital feedback to refine usability and functionality in line with on-ground operational needs.

Evaluation at use stage. The system is consistently monitored for latency and downtime to maintain stable performance. Its accuracy is continually enhanced through the integration of new data, preserving relevance and precision. Moreover, the system is continually evaluated with feedback from supermarket managers, staff, and customers.

1.5 **Teams.** The system's design involved a diverse team of professionals. AI and machine learning engineers specializing in biometric technologies worked alongside cybersecurity experts to secure sensitive customer data. Retail technology specialists ensured the system's integration into supermarket workflows, enhancing user experience without infringing on privacy. Privacy and ethics consultants, alongside legal advisors, addressed regulatory compliance and ethical concerns, ensuring transparent customer consent mechanisms were in place. Consumer behavior researchers and supermarket staff provided insights into user preferences and operational efficiency.

Section 2 RISKS

The system is high risk due to its use of biometric identification (EU AI Act, Annex III, point 1(a)).

2.1 Capability Risks.

Customer data leak. For customers, the storage of personally identifiable data, including biometric information, raises concerns about privacy and security. In the event of a breach, such as unauthorized access or hacking, customers face long-term privacy issues as biometric data cannot be replaced or reset. This jeopardizes their personal information and may lead to identity theft or misuse. Additionally, for stores, a breach in the system could result in financial losses due to legal repercussions, fines, and potential lawsuits. It may also damage the stores' reputation and deter customers from engaging with the business, impacting revenue and growth. Institutions, including regulatory bodies, are also at risk, as a breach would undermine their efforts to enforce data protection laws, such as those outlined in the EU AI Act, leading to a loss of trust and credibility in their ability to safeguard customers' privacy rights. Moreover, improper handling of system's data could negatively imact the environment, contributing to electronic waste and carbon emissions through increased server usage and data storage.

Delays during power and network disruptions. For customers, such disturbances can create inconvenience, potentially affecting their satisfaction and loyalty. For stores, these disruptions result in missed sales opportunities, leading to direct financial losses and the potential for reputational harm.

Page 1 / 2

Impact Assessment Report

Biometric Checkout

system's phase: use

2.2 Human Interaction Risks.

Unauthorized customer behaviour tracking. For customers, such tracking can result in invasive targeted advertising or discrimination based on purchasing habits, undermining individuals' rights to privacy and nondiscrimination. For stores, such tracking may lead to breaches of customer trust, potentially resulting in reputational damage and loss of business. Institutions, particularly regulatory bodies, may face challenges in enforcing privacy laws and protecting consumer rights against such invasive practices, undermining trust in the regulatory framework. From an environmental standpoint, excessive data collection and storage could contribute to digital pollution, exacerbating the carbon footprint of the system.

Customer unfamiliarity. Customers may face unfamiliarity with privacy and data security regarding how their biometric data is stored and utilized, along with technological skepticism and accessibility challenges, particularly among certain groups. Cultural sensitivities and individual resistance to change may further impede acceptance.

2.3 Systemic Impact Risks.

Digital exclusion. For customers, the implementation of biometric checkout systems could lead to marginalization among individuals who are less technologically literate or have diverse abilities, as well as those who opt out of biometric identification due to ethical, religious, or privacy concerns. This digital divide may result in unequal access to supermarket services, particularly affecting marginalized communities and exacerbating inequality (Sustainable Development Goal 10), contrary to efforts aimed at ensuring inclusive and equitable service provision for all (Sustainable Development Goal 16). For stores, the digital exclusion resulting from biometric checkout systems can lead to loss of business opportunities as segments of the population are unable or unwilling to engage with the technology. This can hinder revenue generation and customer retention efforts, ultimately impacting the store's profitability and competitiveness. For institutions, digital exclusion poses a risk of widening the gap in access to essential services, undermining efforts to promote social inclusion and economic empowerment. Additionally, it may lead to decreased trust in institutions perceived as prioritizing technological advancement over inclusivity and accessibility. For the environment, the reliance on biometric checkout systems without addressing digital exclusion can exacerbate electronic waste generation. Disenfranchised individuals may resort to outdated, less sustainable methods of transaction, contributing to environmental degradation.

Section 3 MITIGATION STRATEGIES

3.1 Mitigations of the Capability Risks.

Customer data leak. The system is undergoing regular updates with the latest security patches to ensure that vulnerabilities in the encryption algorithms are promptly addressed, minimizing the risk of unauthorized access to customer data. Additionally, these updates enhance firewall configurations, fortifying the system's defenses against potential cyberattacks.

Delays during power and network disruptions. By maintaining traditional, non-biometric payment methods and checkout methods, such as cash and manual checkout processes, supermarkets can continue to process transactions and serve customers effectively during instances of power outages or network disruptions. This approach ensures business continuity and customer satisfaction,

3.2 Mitigations of the Human Interaction Risks.

Digital exclusion. This risk can be mitigated in the future updates of the system by incorporating assistive technologies in the system, including voice-activated commands and adaptive interfaces. Voice-activated commands can help visually impaired customers navigate the interface of the self-checkout machines, while adaptive interfaces can adjust to the needs of those with motor impairments, ensuring that all customers have equal access to digital services and an enhanced shopping experience.

3.3 Mitigations of the Systemic Impact Risks.

Digital exclusion. This risk can be mitigated in the future updates of the system by incorporating assistive technologies in the system, including voice-activated commands and adaptive interfaces. Voice-activated commands can help visually impaired customers navigate the interface of the self-checkout machines, while adaptive interfaces can adjust to the needs of those with motor impairments, ensuring that all customers have equal access to the shopping experience.

Section 4 BENEFITS

Reduction in checkout times. By implementing biometric technology for the checkout process, the system significantly reduces transaction times, offering customers a more convenient and streamlined shopping experience. For stores, this translates to increased customer satisfaction and loyalty. Furthermore, by making cutting-edge technology accessible to all customers, regardless of technological proficiency, stores uphold principles of inclusivity and cultural advancement (Article 27, UN Declaration of Human Rights).

Reduction of identity theft and fraudalent transactions. The use of biometric identification ensures a high level of accuracy in customer identification. This capability significantly reduces the risk of identity theft and fraud, offering a more secure transaction environment for both customers and stores. The reliability of biometric identification supports trust and safety in financial transactions, contributing to a secure digital economy.

Reduction in the need for human labor at checkout points. The adoption of biometric checkout systems can lead to significant improvements in operational efficiency for supermarkets, reducing labor costs associated with traditional checkout processes and minimizing transaction errors.

Optimization of inventory levels. By linking biometric identification with shopping histories and payment methods, the system enables the collection of valuable data on shopping behaviors and preferences. This capability allows retailers to gain insights into consumer trends, enabling personalized shopping experiences and improving inventory management

Reduction in the need for physical payment methods. The transition to digital transactions in checkout processes reduces reliance on physical payment methods such as paper receipts and plastic cards, benefiting different stakeholders. For customers, the shift enhances convenience by streamlining the checkout process, enabling quicker transactions. Stores benefit from reduced operational costs associated with managing physical payment methods, such as printing paper receipts and maintaining card readers. Institutions experience improved efficiency and data management through digital transactions, facilitating better financial tracking and analysis. Moreover, the environment benefits from reduced waste generation.

REPORTING RISKS

Helpline: 0XXX XXX XXX
Reporting portal: report-risk@com
Mail: XX Main Street,
XXX-XXX Contry X

Last update: 29 Feb 2024

REGISTERED OFFICE

Name of the company
XX Main Street,
XXX-XXX Contry Z

CERTIFICATES

GDPR Compliant

PCI DSS Compliant

Page 2 / 2

Fig. 9. Impact assessment report for a store checkout system using biometric identification, used during the large-scale online study. Available also at https://anonymous.4open.science/r/AIIA_Card/report-biometric-checkout.pdf.

Impact Assessment Report

License Plate Detector

system's phase: use

Section 1 INFORMATION ON THE SYSTEM'S USE AND TEAMS

1.1 System's Use.

Purpose. Streamlining and securing the parking allocation process by using image recognition.
Capability. Optical character recognition.
Domain. Retail and customer services.
AI User. Supermarket security managers and staff.
AI Subject. Customers of the supermarket.

1.2 **System Components.** The system uses a multi-model architecture to process:
(1) license plate images from a Pan-Tilt-Zoom CCTV camera,
(2) logs with the time of entry and exit of the vehicle from a loop sensor.

For real-time license plate image processing (1), the system uses two specialized Convolutional Neural Networks. CNN-Plate (version 3.1) is used to detect and isolate the license plate from the rest of the vehicle image. This involves recognizing the plate's shape and size. CNN-Digit (version 5.1) is used to segment the characters from the plate, including registration numbers, letters, and symbols.
For time log processing (2), the system uses logistic regression, version 8.1.1. It identifies outliers and anomalies in the time log data, such as a vehicle taking an unusually long time to exit or enter, which could indicate a problem or an exception that needs attention. The system is equipped with an alert mechanism that triggers notifications to both the customer and store personnel whenever any of the models detect an anomaly.

1.3 **System Data.** The system utilizes a comprehensive training and testing dataset of 612 437 images, adhering to a 70/30 split. Captured under various conditions such as daylight and nighttime, and during different weather scenarios including rain, snow, and fog, these images are also taken from multiple angles, encompassing both the front and back of the vehicle. They feature varying levels of blur due to vehicle movement. Moreover, to ensure accurate character recognition, the dataset includes a wide range of license plates from different regions, showcasing variations in plate designs, fonts, and colors.
The dataset was obtained by combining the Chinese City Parking Dataset (300 000 images), the Vehicle Make and Model Recognition dataset (291 752 images), the Stanford Cars dataset (16 185 images), and the UFPR-ALPR Dataset (4 500 images). The dataset is updated twice a year to reflect new license plate formats and adapt to changes in vehicle registration designs, ensuring the system remains effective over time.
The system processes license plate images that contain personally identifiable information. It also records logs of vehicles' entry and exit times, which, while not classified as personally identifiable information, could potentially be used for purposes beyond their intended use. Future plans involve combining these logs with phone numbers from the customer database to offer personalized shopping discounts that are synchronized with the customer's typical shopping hours, providing them with special offers precisely when they are most likely to visit the store. To safeguard system data, stringent protocols are in place governing data access, storage, and processing. Access to this information is specifically limited to designated supermarket staff, including security managers and IT support personnel. These measures are designed to protect individual privacy and ensure the system's compliance with the General Data Protection Regulation (GDPR).

1.4 System Evaluation.

Evaluation at development stage. The evaluation of models for optical character recognition encompassed a variety of scenarios to ensure robustness and accuracy. Benchmarks included various conditions, including different vehicle types, license plate designs, and environmental settings like lighting and weather. The system's accuracy was tested by comparing its ability to correctly identify license plates against a pre-registered database under these varying conditions. License plate recognition achieved an accuracy rate of 97%, while digit recognition reached 94%. Additionally, the evaluation measured the system's response time from capturing the vehicle's entry and exit times via loop sensors to the moment the parking lot barrier goes up, aiming for a seamless and fast parking experience.
The evaluation of the model to identify outliers and anomalies within the time log data involved a combination of automated testing, manual verification, and real-world transaction simulation to ensure the system's robustness and accuracy. Simulated parking scenarios were conducted under various conditions, including peak hours, weekends, and adverse weather conditions, to gauge the system's accuracy and security in processing parking time information. Results indicated an accuracy rate of 75%, while instances where vehicles took unusually long durations to enter or exit were closely scrutinized to derive critical indicators of potential exceptions requiring personnel's attention.

Evaluation at deployment stage. The system was piloted in 50 supermarkets and evaluated in collaboration with 350 end users, including individuals in various positions such as security managers and customer service representatives, gathering vital feedback to refine usability and functionality in line with on-ground operational needs.

Evaluation at use stage. The system is consistently monitored for latency and downtime to maintain stable performance. Its accuracy is continually enhanced through the integration of new data, preserving relevance and precision. Moreover, the system is continually evaluated with feedback from supermarket security managers, staff, and customers.

1.5 **Teams.** The system's design involved a diverse team of professionals. Computer vision specialists focused on license plate recognition, and experts in privacy and data security to ensure the protection of customer information. Collaboration with parking management professionals ensured the system met operational needs for efficient parking management and security. Legal and ethical advisors provided compliance guidance, while input from supermarket staff and customers and was integrated to tailor the system towards user-centric functionality.

Section 2 RISKS

The system poses limited risk due to its processing of personally identifiable data within non-critical domains (EU AI Act, Annex III).

2.1 Capability Risks.

Accidentally capturing images of vehicle's surroundings. For customers, if the system captures images of drivers and passengers, this can lead to privacy concerns, potentially resulting in distrust and reluctance to use the supermarket's services. For stores, this could negatively impact their reputation and brand image. If customers perceive that their privacy is not adequately protected, they may opt to shop elsewhere, resulting in a loss of revenue. Additionally, non-compliance with data protection regulations could lead to legal penalties and fines, further damaging the store's reputation. Institutions, including regulatory bodies, are also at risk, as a privacy breaches would undermine their efforts to enforce data protection laws, leading to a loss of trust and credibility in their ability to safeguard customers' privacy rights. Moreover, improper handling of system's data could negatively imact the environment, contributing to electronic waste and carbon emissions through increased server usage and data storage.

Delays during power and network disruptions. For customers, such disturbances can create inconvenience, potentially affecting their satisfaction and loyalty. For stores, these disruptions result in missed sales opportunities, leading to direct financial losses and the potential for reputational harm.

Page 1 / 2

Impact Assessment Report

License Plate Detector

system's phase: use

2.2 Human Interaction Risks.

Information imbalance between staff and customers during parking conflicts. In situations where the system's data contradicts a customer's account of their parking activities, staff may face challenges in mediating disputes, impacting customer trust and satisfaction.

2.3 Systemic Impact.

Perpetuating the perception of constant surveillance. Although the system primarily targets license plates, its existence may engender a perception of perpetual surveillance, affecting the psychological well-being of customers, staff, and local residents. This constant surveillance impression could prompt heightened scrutiny from regulatory authorities, potentially triggering audits and investigations that disrupt institutional operations. Furthermore, the increased energy consumption needed to maintain the system's continuous functionality may worsen environmental degradation, undermining the long-term sustainability goals outlined in the ESG strategies of these stores.

Section 3 MITIGATION STRATEGIES

3.1 Mitigations of the Capability Risks

Accidentally capturing images of vehicle's surroundings. The system's operators can strategically position cameras to focus specifically on areas where license plates are expected. Additionally, installing physical shields around cameras can limit their field of view and prevent them from capturing extraneous surroundings. These measures help ensure that only relevant images containing license plates are collected, reducing the risk of privacy infringement.

Delays during power and network disruptions. Implementing manual backup systems like ticket booths or manual attendants can ensure parking operations continue smoothly. Additionally, installing backup power sources such as generators or UPS systems can keep the automated parking allocation system running during power outages, minimizing disruptions for customers. These measures provide resilience against unforeseen technical issues, ensuring efficient parking management even in adverse conditions.

3.2 Mitigations of the Human Interaction Risks

Information imbalance between staff and customers during parking conflicts. Installling interactive screens at prominent locations within the store, such as near parking entrances or payment kiosks, allows customers to easily access parking rules and check their current parking status. These screens can also display real-time information about available parking spots, time limits, and any special regulations, empowering customers to make informed decisions and reducing the likelihood of conflicts with staff.

3.3 Mitigations of the Systemic Impact Risks

Perpetuating the perception of constant surveillance. Implementing automatic deletion of captured license plate images after seven days can enhance privacy protection for individuals. This policy can be enforced through automated scripts or scheduled tasks integrated into the system's backend infrastructure. Exceptions can be made for law enforcement and court orders, with secure access controls ensuring that only authorized personnel can access retained images for legitimate legal purposes.

Section 4 BENEFITS

Reduction in parking time. For customers, shorter parking durations enhance convenience and efficiency, reducing the time spent searching for parking spots and waiting to exit the facility. This improved experience increases customer satisfaction and loyalty, encouraging repeat visits. For stores, the streamlined parking process increases turnover rates, allowing more customers to access the facility within a shorter timeframe. This leads to higher foot traffic, improved sales opportunities, and ultimately boosts revenue for the store.

Reliable calculation of parking time is essential. For customers, accurate parking time calculations ensure fair and transparent billing, preventing overcharges and disputes. For stores, reliable parking time calculations optimize parking space utilization and help ensure a smooth flow of traffic.

Improvement in the security of the parking area. For customers, enhanced security provides peace of mind, reducing the risk of theft, vandalism, or other criminal activities affecting their vehicles. For stores and institutions, heightened security measures deter criminal behavior, safeguarding assets and reducing potential liabilities. Additionally, a secure parking area contributes to community safety, enhancing the overall well-being of the neighborhood.

Reduction in the need for human labor at parking points. The model can monitor the performance of the parking system over time, helping the stores to identify periods of inefficiency or increased demand. The model's predictions can be used to optimize staff scheduling for peak times or plan maintenance work during predicted low usage periods

Reduction in the congestion and unauthorized parking. By understanding patterns of enter end exit logs, the system can optimize the allocation of parking spaces and potentially reduce wait times for the customers.

REPORTING RISKS
Helpline: 0XXX XXX XXX
Reporting portal: report-risk@com
Mail: XX Main Street,
XXX-XXX Contry Z

Last update: 29 Feb 2024

REGISTERED OFFICE
Name of the company
XX Main Street,
XXX-XXX Contry Z

CERTIFICATES
GDPR Compliant
PCI DSS Compliant

Page 2 / 2

Fig. 10. Impact assessment report for a car park monitoring system using image recognition, used during the large-scale online study. Available also at https://anonymous.4open.science/r/AIIA_Card/report-license-plate-detector.pdf.

A.4 Rubric for Evaluating Email Quality

The quality of emails was assessed based on five key criteria. Each email was evaluated for its ability to: address the real-world use of the system (*context*), provide a clear call to action for or against implementation (*recommendation*), identify and discuss risks associated with the system (*risks*), mention actionable strategies to mitigate these risks (*mitigations*), and present information in a clear and coherent manner (*content clarity*). Lower-rated emails often failed to directly engage with the system described and were vague. Higher-rated emails demonstrated a nuanced understanding of the system, offered balanced arguments covering risks and benefits, and included solutions for addressing identified risks.

Detailed Criteria for Email Quality Ratings

Rating 1: Poor quality

Context: No mention of the system's real-world use.
Recommendation: Lacks a decisive recommendation.
Risks: Fails to mention any risks.
Mitigations: Does not include any mitigation strategies or references to actions for reducing risks.
Content clarity: Content is highly vague or incomprehensible. Structure or grammar issues significantly hinder readability.

Rating 2: Fair quality

Context: Briefly mentions the system's real-world use but lacks elaboration.
Recommendation: The recommendation is unclear or weak.
Risks: Includes at least one risk. Risks are mentioned but lack relevance to the system's real-world use.
Mitigations: Includes at least one mitigation strategy. Mitigation strategies are mentioned but are not clearly tied to the specific risks of the system's real-world use.
Content clarity: Content is somewhat vague or challenging to follow. Structure lacks focus and clarity.

Rating 3: Good quality

Context: Explains the system's real-world use in at least one sentence, demonstrating a basic understanding of the system.
Recommendation: The recommendation is clear (recommend or reject) but could be more compelling.
Risks: Includes at least two risks. Risks are moderately connected to the system.
Mitigations: Includes at least two mitigation strategies tied to specific risks of the system's real-world use.
Content clarity: Content is well-written with minimal vagueness. Logical structure supports the argument, though it may lack sophistication.

Rating 4: Very good quality

Context: Explains the system's real-world use in detail, demonstrating a clear grasp of its operational implications and relevance to its users and subjects.
Recommendation: The recommendation is clear (recommend or reject) and decisive.
Risks: Identifies and discusses multiple risks (>2). Risks are clearly connected to the system. Attempts to prioritize key risks.
Mitigations: Includes more than two mitigation strategies tied to specific risks of the system's real-world use. The included mitigations are actionable.
Content clarity: Content is very well-written and logically structured, making the information easy to follow. Arguments are cohesive and well-supported.

Rating 5: Excellent quality

Context: Demonstrates a nuanced understanding of the system's real-world use, with concrete examples and scenarios related to its users and subjects.
Recommendation: The recommendation is clear (recommend or reject) and decisive. It balances pros and cons with depth and foresight.
Risks: Identifies and thoroughly discusses all key risks, including subtle or rare risks, or identifies new risks expanding the scope of the treatment. Effectively prioritizes risks with clear justification.
Mitigations: Includes more than two mitigation strategies tied to specific risks of the system's real-world use. The included mitigations are actionable, specific, and technically detailed .
Content clarity: Exceptionally clear, precise, and insightful writing style. Engages the reader and delivers a compelling, logical argument.

Examples of Emails and Their Quality Ratings

Rating 1: Poor quality

Email text: I hope this email finds you and your team doing well. I recently stumbled upon and read your document implementing the new AI system in the store. This will definitely help us improve our understanding and familiarity with artificial intelligence.
Justification for rating: The email lacks specific references to the AI system's use, does not state a recommendation or rejection, and fails to provide arguments related to the system's use, risks, or mitigations.

Rating 2: Fair quality

Email text: Dear team, it has come to my attention that there are significant risk associated with the use of the biometric checkout system. Considering these risks and the potential negative impact on both customers and the environment, I kindly request that we consider the re-implementation of the system.
Justification for rating: The email references the specific AI system's use, but the recommendation is unclear. It mentions general risks and includes re-development as one mitigation strategy.

Rating 3: Good quality

Email text: Hello, I'm writing this email to you to recommend implementing the system of biometric checkout. Although, a good and safe option would be to make an opt-in system and inform the customers. That way people can't complain. I understand some people will say this is too much surveillance and will invade our privacy. Overall, it will reduce identity theft and fraudulent transactions. Let's say, you lose your card, someone else grabs it and tries to buy something at a store, with biometric checkout, this person will be caught.
Justification for rating: The email references the specific use of the AI system, and its recommendation is clear. It identifies two primary risks associated with the system: the potential for surveillance and privacy invasion. To address these risks, the email outlines two specific mitigation strategies: implementing an opt-in system to ensure user consent and proactively informing customers about how their data will be used. Additionally, the email provides a balanced perspective by including arguments that highlight both the potential benefits and drawbacks of the system.

Rating 4: Very good quality

Email text: Dear Ethics Committee, I am writing this email to advise against implementing this license plate detector for use in the parking lot. I think the potential risks with this AI system outweighs the potential benefits. I believe that accidentally capturing images of vehicle's surroundings will lead to conflict between staff and customers, as well as raise privacy concerns among customers. The mitigation strategies included, such as installing physical shields and shortening data storage time, will increase cost for a solution that doesn't help much. The benefits of this system are more for the store's benefit rather than the customers', which could potentially lead to losing customers.
Justification for rating: The email provides a detailed explanation of the AI system's real-world use, showcasing a strong understanding of its operational implications and relevance to both customers and the store. The recommendation is clear and decisive, addressing four key risks: conflicts between staff and customers, privacy concerns among customers, potential loss of customers, and increased costs for the store. It offers more than two targeted mitigation strategies tied to these risks, presented in a well-written and logically structured manner.

Rating 5: Excellent quality

Email text: Dear Ethics Committee, I have reviewed the information presented and I *advise against* adopting this system. The following risks are described: (1) Risk of customer data leak - it is unacceptable to collect sensitive data such as biometrics and put it at risk of being stolen or leaked. "Latest security patches" is not sufficient as a strategy to safeguard high-value data, especially if your organization is specifically targeted for data theft. (2) Unauthorized customer tracking - there is no guarantee that a hacker, rogue employee, or rogue vendor cannot access the data and use it for their own purposes, even when safeguards exist. (3) Customer unfamiliarity - What if customers become familiar through the materials ans and don't want to use it? They will stop using the store, or they will complain on social media or to government regulators. The benefit does not outweigh the risk. In short, you must judge the system by what happens when it fails, not when everything goes right. In this case the legal and ethical risk is too high when the system fails.
Justification for rating: The email provides a clear and well-justified recommendation. It explains the AI system's real-world use in detail, demonstrating concrete scenarios beyond those explicitly mentioned in the treatment, related to affected individuals. The recommendation is clear and decisive, prioritizing three key risks and including three targeted mitigation strategies directly tied to these risks.

A.5 Results of Regression Analyses for Predicting Task Completion Time, Usability Ratings, and Preference for Cards vs. Reports.

A.5.1 Predicting Task Completion Time.

Table 6. **Factors influencing completion time include treatment and participant's cohort.** Using the report significantly increases completion time compared to the card, while legal experts take longer to complete the task than ordinary individuals. We conducted an ordinary least squares regression analysis with completion time as the dependent variable. The coefficients represent the changes in completion time (in seconds) relative to the reference category, with statistical significance indicated by: * for $p < 0.05$, and ** for $p < 0.01$. Non-significant factors (p > 0.05) are also reported for completeness.

Factor	Comparison (*vs.* Reference Category)	Coefficient	p-value
Intercept		302.870	0.000***
Type of task			
System	Plate Detector *vs.* Checkout	-3.409	0.908
Participant's cohort			
Cohort	Developers *vs.* Ordinary individuals	-50.423	0.200
Cohort	**Legal Experts *vs.* Ordinary individuals**	**105.228**	**0.013***
Expertise levels			
Task Expertise	High *vs.* Low	13.841	0.474
Technological Expertise	High *vs.* Low	-26.685	0.227
AI Expertise	High *vs.* Low	34.298	0.119
Treatment			
Treatment type	**Report *vs.* Card**	**102.617**	**0.001****

A.5.2 Predicting Usability Ratings.

Table 7. **Factors influencing usability ratings include treatment and participant's cohort.** Participants across all cohorts find the report less usable than the card, and ordinary individuals give lower usability ratings compared to developers and legal experts. We conducted an ordinary least squares regression analysis with usability as the dependent variable. The coefficients represent the changes in usability scores relative to the reference category, with statistical significance indicated by: * for $p < 0.05$, ** for $p < 0.01$, and *** for $p < 0.001$. Non-significant factors (p > 0.05) are also reported for completeness.

Factor	Comparison (*vs.* Reference Category)	Coefficient	p-value
Intercept		50.124	0.000
Type of task			
System	Plate Detector *vs.* Checkout	3.076	0.083
Participant's cohort			
Cohort	**Developers *vs.* Ordinary individuals**	**4.769**	**0.043***
Cohort	**Legal Experts *vs.* Ordinary individuals**	**6.199**	**0.004****
Expertise levels			
Task Expertise	High *vs.* Low	1.425	0.218
Technological Expertise	High *vs.* Low	2.492	0.060
AI Expertise	High *vs.* Low	-0.776	0.555
Treatment			
Treatment type	**Report *vs.* Card**	**-11.750**	**0.000*****

A.5.3 Predicting Preference for Cards vs. Reports.

Table 8. **Factors influencing preference for cards include recommendation type, participant's cohort, and AI expertise.** "Reject" and "Unclear" recommendations, belonging to the legal experts cohort, and greater AI expertise all reduce the likelihood of preferring cards. We conducted a binomial logistic regression analysis with preference for cards (vs. reports) as the dependent variable. The coefficients represent the log-odds of preferring a card relative to a report for each factor, compared to its reference category. Statistical significance is indicated by: * for $p < 0.05$, and ** for $p < 0.01$. Non-significant factors (p > 0.05) are also reported for completeness.

Factor	Comparison (*vs.* Reference Category)	Coefficient	p-value
Intercept		1.635	0.001
Type of task			
Recommendation	**Reject *vs.* Recommend**	**-0.542**	**0.018****
Recommendation	**Unclear *vs.* Recommend**	**-0.936**	**0.002****
System	Plate Detector *vs.* Checkout	-0.012	0.953
Participant's cohort			
Cohort	Developers *vs.* Ordinary individuals	-0.440	0.105
Cohort	**Legal Experts *vs.* Ordinary individuals**	**-0.707**	**0.005****
Expertise levels			
Task Expertise	High *vs.* Low	0.151	0.258
Technological Expertise	High *vs.* Low	0.178	0.240
AI Expertise	**Low *vs.* High**	**-0.477**	**0.002****
Treatment			
Treatment type	Report *vs.* Card	0.132	0.523

A.6 Updated Impact Assessment Cards

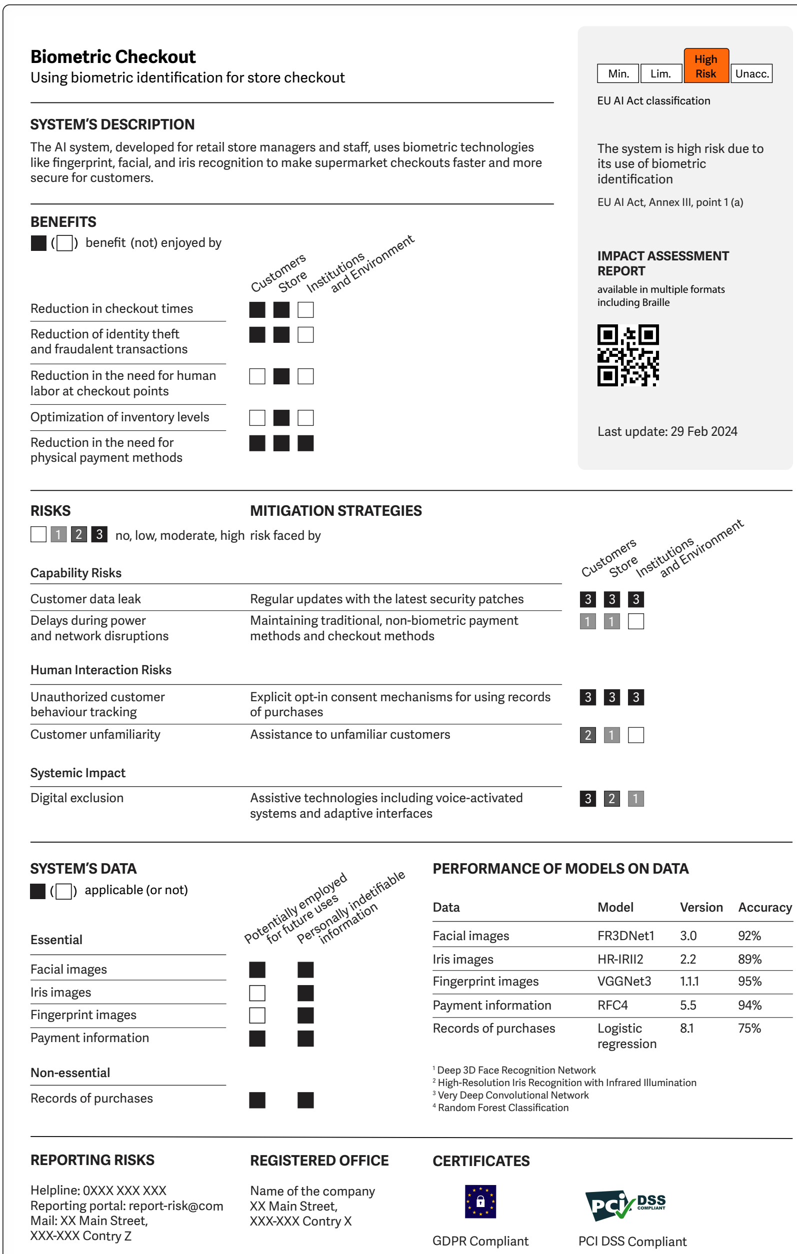


Fig. 11. Updated impact assessment card for a store checkout system using biometric identification, including the risk severity ratings. Available also at https://anonymous.4open.science/r/AIIA_Card/impact-assessment-card-biometric-checkout-version5.pdf.

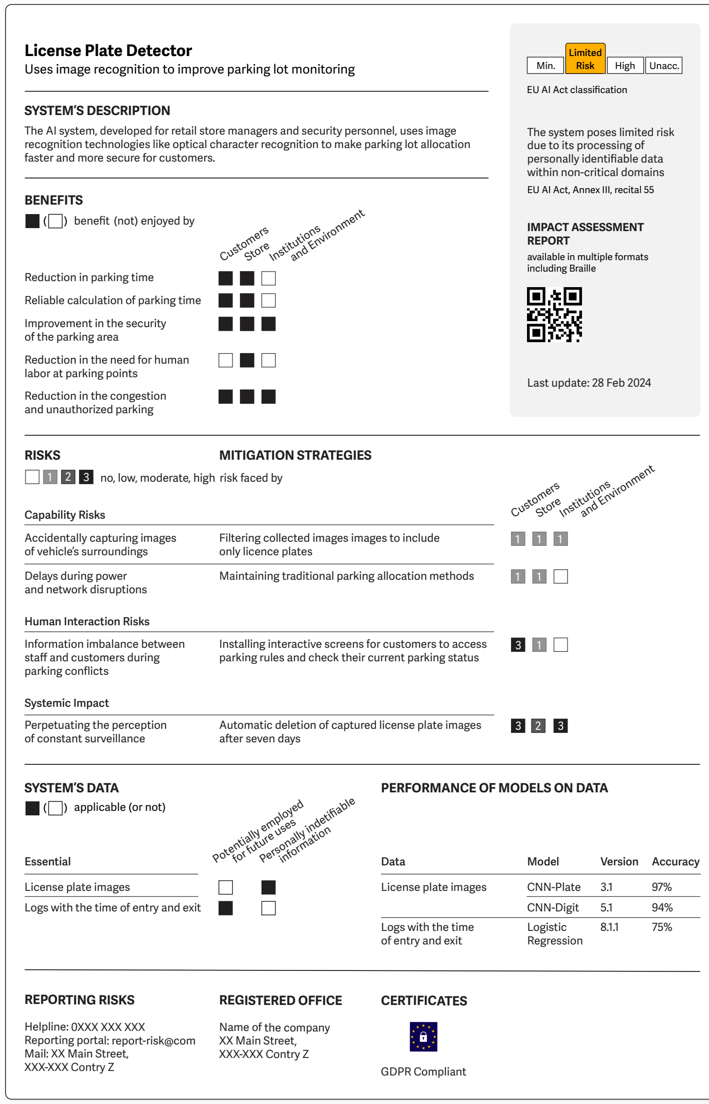


Fig. 12. Updated impact assessment card for a car park monitoring system using image recognition, including the risk severity ratings. Available also at https://anonymous.4open.science/r/AIIA_Card/impact-assessment-card-license-plate-detector-version5.pdf.

A.7 Impact Assessment Cards for Digital AI Systems

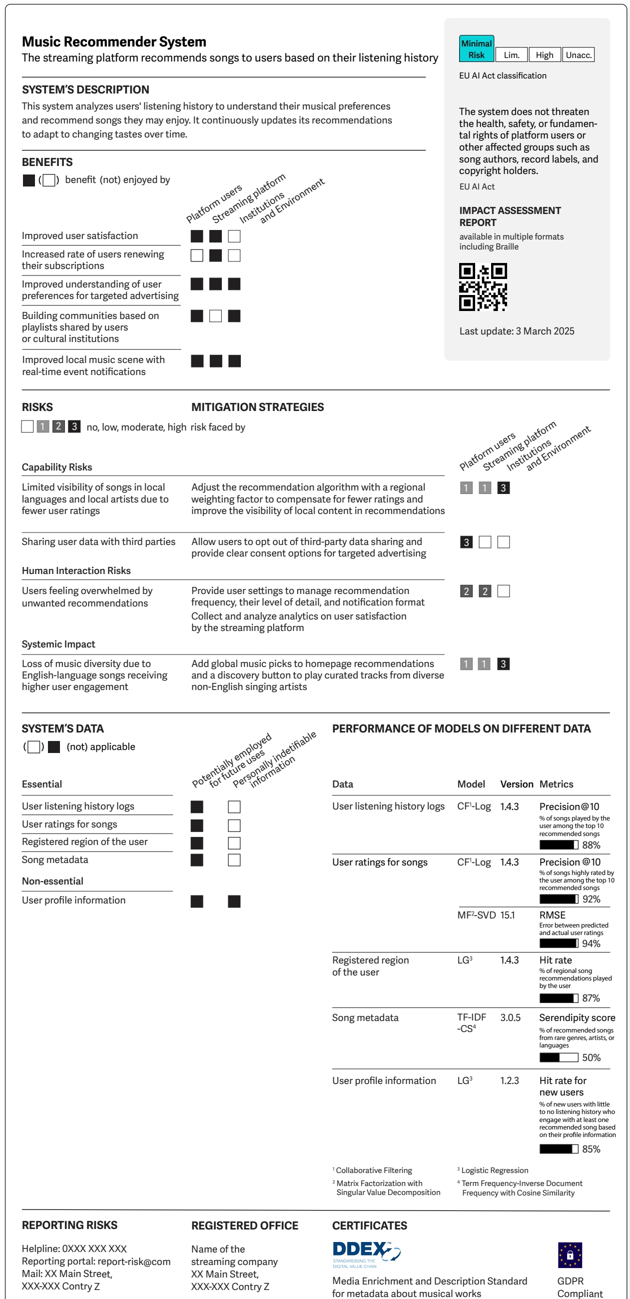


Fig. 13. Impact assessment card for a music recommender system that suggests songs to platform users based on their listening history. Available also at https://anonymous.4open.science/r/AIIA_Card/impact-assessment-card-music-recommender.pdf.

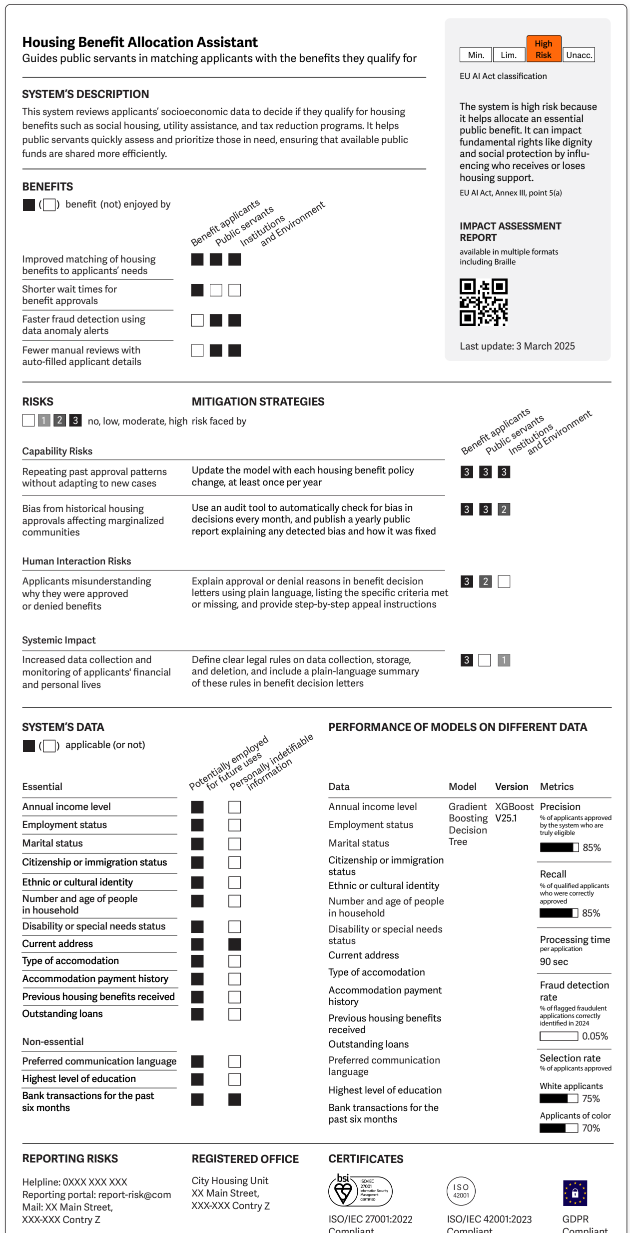


Fig. 14. The impact assessment card for a housing benefit allocation assistant for public servants reviews applicants' socioeconomic data to match them with eligible benefits such as utility assistance and tax reduction programs. In addition to the risk determination from the EU AI Act [20], the summary box includes the system's life cycle stage and the required approvals for operation, as mandated by ISO 42001. [39]. Available also at https://anonymous.4open.science/r/AIIA_Card/impact-assessment-card-benefit-assistant.pdf.

Received 20 February 2007; revised 12 March 2009; accepted 5 June 2009