Mapping the Design Space of Responsible AI Cards: Evaluating Existing Decks and Providing Design Recommendations

Anonymous Author(s)

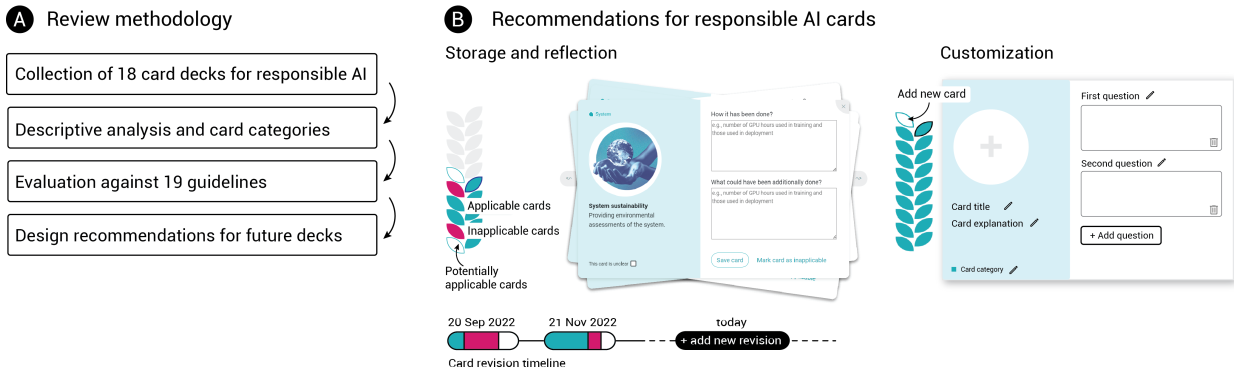

Fig. 1: We systematically mapped out the design space of responsible AI card decks, and proposed two design recommendations to enrich the space, that is, storage and reflection on previous card interactions and card customization.

Abstract— Developing Artificial Intelligence (AI) is a highly intricate process involving socio-technical and practical factors. To simplify and facilitate this process, tools rooted in card designs have been proposed. In particular, cards have been suggested as means to identify possible effects of AI on individuals, society, and the environment, and to help in predicting negative consequences of AI development – the emerging area of responsible AI. However, the space of existing responsible AI cards lack systematic mapping against design and information visualization guidelines. To partly map out the design space, we collected, reviewed, and evaluated 18 responsible AI card decks published between 2017 and 2022. We did so by using 19 design guidelines grouped into two main categories: *content quality* and *visual properties*. On the upside, we found that existing card decks are straightforward to begin with, are understandable by their intended users, and are consistent in design choices without distracting visual elements. On the downside, we found that due to their physical form existing cards are not suitable for the use in distributed teams, and that they fall short in customization. In light of these findings, we provide two design recommendations for responsible AI cards. Future cards should: (1) allow for storage and reflection on previous card interactions; and (2) allow for customization through editable decks and the ability to add new cards.

Index Terms—card-based visualizations, design cards, responsible AI, design ethics, AI ethics, literature review, guidelines evaluation

1 Introduction

Developing Artificial Intelligence (AI) is not an easy task, let alone building responsible AI (RAI) — an approach to designing, developing, and deploying AI systems in a safe, trustworthy, and ethical way [41, 66, 68]. In fact, building RAI systems is a highly challenging and intricate process involving socio-technical and practical factors [24]. If we view AIs as socio-technical systems, this entails a combination of technical components (i.e., the data and code) and social elements (i.e., the society in which an AI is deployed) that we interact with each other [23, 25]. As with all socio-technical systems, AIs are subject to the so-called "socio-technical gap," which, in Ackerman's words, is *"the divide between what we know we must support socially and what we can support technically"* [6]. That is why calls for addressing these challenges in the development of responsible AI are mounting [64].

To facilitate the development of responsible AI, and, in turn, address some of these challenges, technology companies and scholars have proposed and developed frameworks, toolkits, and guidelines. Prominent examples include IBM's 360 toolkits on identifying biases [3] and providing explainability in models [2], Meta's Fairness Flow [42], or Google's What-If-Tool [34]. More recently, tools rooted in card designs have been proposed. Cards are typically used to assist in technology development, foster creativity, and facilitate user participation in the early design stages [54], not least because they are simple, tangible, interactive, and easy to manipulate [76]. For example, the Feminist Tech Card Deck aims at initiating an ethical debate on 12 principles of fair policy-making and technology, such as ensuring accessibility, equitable participation, and representation [67]; the IDEO AI Ethics Cards aim at engaging all stakeholders in considering the practical implications of an AI system, with reusable cards at different stages of the development lifecycle (e.g., onset, early prototyping, decision-making, and planning) [38]; and the Microsoft Judgement Call is a speculative card-based game that invites players to empathize with various AI systems and to write fictional systems' reviews [11]. The game successfully empowered product teams to think broadly about ethical considerations that are directly applicable to their work.

Despite the recent upsurge and efforts in developing cards for responsible AI, these efforts are scattered and fragmented. At the time of writing, the cards' design space lacks systematic mapping against design and information visualization guidelines. To map out the design space, we formulated two Research Questions (RQs):

RQ_1: Which of the state-of-the-art card decks support responsible design and development of AI, and how?

RQ_2: To what extent do the state-of-the-art card decks fulfill design guidelines pertaining to content quality and visual properties?

Manuscript received xx xxx. 201x; accepted xx xxx. 201x. Date of Publication xx xxx. 201x; date of current version xx xxx. 201x. For information on obtaining reprints of this article, please send e-mail to: reprints@ieee.org. Digital Object Identifier: xx.xxxx/TVCG.201x.xxxxxxx

In answering our RQs, we made three sets of contributions:

- We collected and thematically reviewed 18 card decks specifically designed for responsible AI, which were developed between 2017 and 2022 (RQ_1, §3).
- We evaluated these card decks using 19 design guidelines pertaining to content qualities and visual properties. On the upside, we found that existing cards are understandable by their intended users and are consistent in design choices without distracting visual elements. On the downside, we found that these cards are not suitable for the use in distributed teams and fall short in customization (RQ_2, §4).
- In light of these findings, we provide two design recommendations for developing future responsible AI cards. Future cards should: (1) allow for storage and reflection on previous card interactions, and (2) allow for customization through editable decks and the ability to add new cards (§5).

2 RELATED WORK

We surveyed various lines of research that our work draws upon, and grouped them into two main areas: *i)* visualization tools for assisting AI development, and *ii)* cards for human-centered and responsible AI.

2.1 Visualization Tools for Assisting AI Development

The massive adoption of AI surfaced an array of questions relating to, for example, whether it is possible to develop and deploy AI models that are fair, interpretable, secure and privacy-preserving, and sustainable [69, 71]. To partly answer these questions, the data visualization and interface design communities have proposed many tools to assist AI development. These tools draw inspiration from a mixed set of domains, such as visual analytics [9, 21], value sensitive design [11, 73], and new genres of AI-related visualizations, such as visual explorative stories [32]. Visualization tools assisting (responsible) AI development are abundant, and can be generally grouped into three areas: *i)* data understanding; *ii)* documentation of AI; and *iii)* ideation for AI development.

2.1.1 Data Understanding

The first group of tools uses visualization to understand the characteristic of the system data, such as clusters, abnormal distributions, errors, or missing data [65]. These practices use well-established static chart types, such as heatmaps, line charts or bar charts, and incorporate them into typical data processing environments such as computational notebooks. Examples of such visualizations include the IBM's AIF360 [3], Microsoft's Fairlearn [14, 26], and Aequitas [61]. More interactive approaches to data understanding are offered in explorative dashboards (e.g., Know Your Data by Google to visualize image datasets [31], Responsible AI dashboard by Microsoft [48]).

2.1.2 AI Documentation

The second group of tools uses visualization to aid system documentation and dissemination. For example, interactive model and data cards aim at ensuring model transparency and accountability (e.g., Model Card++ by NVIDIA [15]). A model card documents a model's: details (e.g., model type, parameters), its intended uses, its performance across a variety of relevant factors, including groups, instrumentation, and environments, evaluation metrics, as well as documentation of the training and evaluation data. Datasheets' documentation includes data provenance and key characteristics, relevant regulations, test results, and significant yet more subjective information such as potential bias, strengths, and weaknesses of a dataset [30]. Finally, visual stories offer high-level explanations that narrow the gap between the expert developers and potential non-expert users of the system. For instance, Google's "AI explorables" [32] are engaging, formula-free visualizations that use interactive techniques to having a broad audience familiarized with important AI-related concepts such as detecting bias in AI models or assessing fairness in input datasets.

2.1.3 AI System Ideation

The third group of tools uses visualizations as part of the AI ideation process. This includes tools for assisting in the creative process of generating ideas, for example, using games. For instance, Microsoft's *Judgment Call* is a card game that leverages speculative thinking and motivates participants to empathize with diverse AI systems while writing imaginary reviews for them. This game has successfully enabled product teams to think broadly about ethical considerations that are directly applicable to their work. Similarly, the *Ethics Inc.* card game invites players to form a think tank to develop a reliable AI system that complies with the European guidelines put forward by the AI High-Level Expert Group [69]. In general, card games can be played by multiple users (e.g., AI researchers and engineers), enabling collaboration and teamwork. In turn, by discussing and building on each other's ideas, users may come up with more robust and creative AI systems.

2.2 Cards for Human-Centered Design and Responsible AI

Cards have emerged as the dominant tool for collaborative ideation, with a significant surge in their usage over the past five years [54]. For example, the meta-review of 155 design decks showed that one out of five decks (21.3%) aims at helping the technical or human-centered design of digital products and systems, such as user interfaces, interactive devices, and computer games. However, there is a gap in knowledge of card ideation tools for newer technologies [54], and very little is known about their design and functionalities to aid responsible AI development.

Regarding responsible AI, previous literature highlights three reasons why cards are good candidates for addressing issues concerning AI's ethical and responsible development and use. First, cards offer a way to present complex information related to responsible AI in a concise and easily understandable manner [8]. They enable users to quickly scan and comprehend their contents without the need to navigate through complex internal documents [19]. This makes them a valuable reference source for teams, helping team members to harmonize knowledge. Second, analog tools such as cards aid cognitive processes [22] and enable technically-oriented teams to generate ideas that are less focused on technology [46]. By simulating various AI scenarios and prompting developers to consider the broader implications of AI, cards can facilitate decision-making and allow players to observe the consequences of their choices. This helps to build empathy and understanding of the impact of AI on different stakeholders during the system development, deployment, and use. Third, cards offer a secure and non-threatening environment for experimenting with AI through fictional scenarios, without fear of guilt or judgment [11]. This can lead to a bottom-up development of more responsible AI practices and help to identify potential risks and challenges early on.

In summary, previous works have proposed various data visualization techniques to assist in the responsible development of AI, such as visual stories and dashboards. These techniques are mainly used for improving data understanding and system documentation, whereas cards are still an unexplored area. Despite the recent upsurge and efforts in developing cards for responsible AI, there is a lack of comparative analyses of their benefits to guide the design of future card decks.

3 METHODS

To answer our RQs, we first collected and described **18** card decks designed for responsible AI, which were released between 2017 and 2022. We then evaluated these cards using 19 design guidelines grouped into two main categories: *content quality* and *visual properties*.

3.1 Positionality Statement

Understanding researcher positionality is essential to demystifying our lens on data collection and analysis [28, 37]. We situate this paper in the United Kingdom in the 21st century, writing as authors primarily working as academic and industry researchers. We identify as one female and three males from Poland, Cyprus, Iran, and Italy with diverse ethnicity and religious backgrounds. Our shared research backgrounds

Table 1: Descriptive analysis of 18 cards decks for responsible AI (in the rows) based on 9 descriptive dimensions (in the columns). Figure 2 presents one exemplary card from each deck. Cats: cards' categories derived from K-modes clustering; Date: the year in which the deck was released; Medium: whether the card deck is physical, digital, or hybrid; Type: methods being used to support ideation; Origin: the decks' designers origin and affiliation; Purpose: the intended use of the card deck; Use Phase: the phase of an AI's lifecycle in which the card deck can be applied; Methodology: instructions related to the use of the deck card; Customization: whether the deck and the cards are editable; Formal qualities: what visual elements are present on the cards.

Cats	#N	Card Deck	Date	Medium	Type	Origin	Purpose	Phase	Methodology	Customization	Formal Qualities
Design triggers	1	AI Awareness Cards [56]	2018	Physical	Prompts	Design	Participatory design	Anytime	Instructions	Not provided	Text, Image, Categories
	2	The Intelligence Augmentation Design Toolkit [29]	2017	Physical	Prompts	Design	Agenda-driven	Beginning	Instructions	Not provided	Text, Image, Categories
	3	IDEO AI Ethics Cards [38]	2019	Physical	Prompts	Design	Repository	As needed	Suggestion for use	Not provided	Text, Image, Categories
	4	AI Meets Design Ideation Cards [7]	2019	Physical	Prompts	Design	Repository	Beginning	Suggestion for use	Not provided	Text, Image, Categories
	5	The Tarot Cards of Tech [10]	2018	Physical	Prompts	Design	Participatory design	Beginning	Suggestion for use	Not provided	Text, Image
	6	Microsoft Guidelines for Human-AI Interaction [8]	2019	Physical	Concepts	Commercial	Repository	As needed	Suggestion for use	Not provided	Text, Image, Categories
	7	The Feminist Tech Card Deck [67]	2022	Physical	Prompts	NGO	Participatory design	As needed	Suggestion for use	Optional	Text, Image, Categories
	8	UnBias Fairness Toolkit [57]	2020	Physical	Components	Design	Participatory design	Specific point	Suggestion for use	Not provided	Only text, Categories
Research nudges	9	Moral-IT [73]	2021	Physical	Components	Academic	Repository	Specific point	Instructions	Optional	Text, Image, Categories
	10	I Love Algorithms [17]	2019	Physical	Concepts	Academic	Agenda-driven	Specific point	Instructions	Not provided	Text, Image
	11	Biotech Ethics Cards [50]	2021	Physical	Prompts	Academic	Participatory design	Specific point	Suggestion for use	Not provided	Text, Image, Categories
	12	ML Ethical Dilemmas Cards [13]	2020	Physical	Prompts	Academic	Participatory design	Specific point	Instructions	Not provided	Text, Image, Categories
	13	ECCOLA [74]	2021	Physical	Components	Academic	Repository	Specific point	Suggestion for use	Not provided	Only text, Categories
	14	Microsoft Judgment Call [11]	2019	Physical	Components	Commercial	Repository	Beginning	Instructions	Required	Text, Image, Categories
Hybrid checklists	15	AI Blindspots Card Set V1 [43]	2019	Hybrid	Concepts	Academic	Participatory design	Beginning	Suggestion for use	Optional	Only text
	16	AI Blindspots Card Set V2 [44]	2019	Hybrid	Concepts	Academic	Participatory design	As needed	Instructions	Optional	Only text, Categories
	17	AI Blindspots Healthcare [45]	2019	Hybrid	Concepts	Academic	Participatory design	As needed	Instructions	Optional	Only text, Categories
	18	AI Blindspots (MIT) [16]	2019	Hybrid	Concepts	Academic	Participatory design	As needed	Suggestion for use	Optional	Text, Image, Categories

include human-centered computing, privacy, security, software engineering, AI, social computing, information visualization, urbanism, and conducting systematic literature reviews.

3.2 Criteria for Cards Collection

First, we defined a set of four **inclusion** criteria:

- **Responsible AI**: We used the definition of AI provided by the National Institute of Standards and Technology (NIST): "AI system is an engineered or machine-based system that can, for a given set of objectives, generate outputs such as predictions, recommendations, or decisions influencing real or virtual environments. AI systems are designed to operate with varying levels of autonomy." [51]. As such, responsible AI system is an AI system that is *safe*, *fair*, *reliable*, *privacy-preserving*, *accountable*, *explainable*, and *sustainable*. These terms originate from the industry, academic, and government guidelines [12, 33, 49, 51, 52, 59, 70] and we used them to show the breadth of research in this area. Other researchers may use other terms to discuss similar concepts.
- **Responsible AI Card Decks**: We considered card deck a participatory tool that visualizes responsible AI practices and facilitates the process of developing ethical AI systems. Single card withih a deck resembles physical card appearance and allow for game-like interactions such as flipping or reshuffling.
- **Availability:** Analyzed cards should be available to the public, formulated in English, and accompanied by author descriptions of their context, purpose, and method of use.

Next, we defined two **exclusion** criteria:

- **Data Documentation Tools:** We excluded tools or applications that considered cards merely as a naming convention or as an interface design pattern—a content element visually resembling a playing card whose purpose was to document related data or model information without facilitating any game-like interactions. As such, we did not include in our analysis Model Cards [35], Data Cards [58], or AWS AI Service Cards [55].
- **Bespoke Cards:** We focused on finding cards that contain generalizable practices and form bodies of knowledge on responsible AI. Therefore, we excluded bespoke card decks designed for a particular AI, as well as decks that were only applicable in a limited domain and did not provide cues on how to adapt them for broader use. For example, we excluded the Value Cards [63] because they were tailored to recidivism prediction system, limiting their reproducibility in other AI uses.

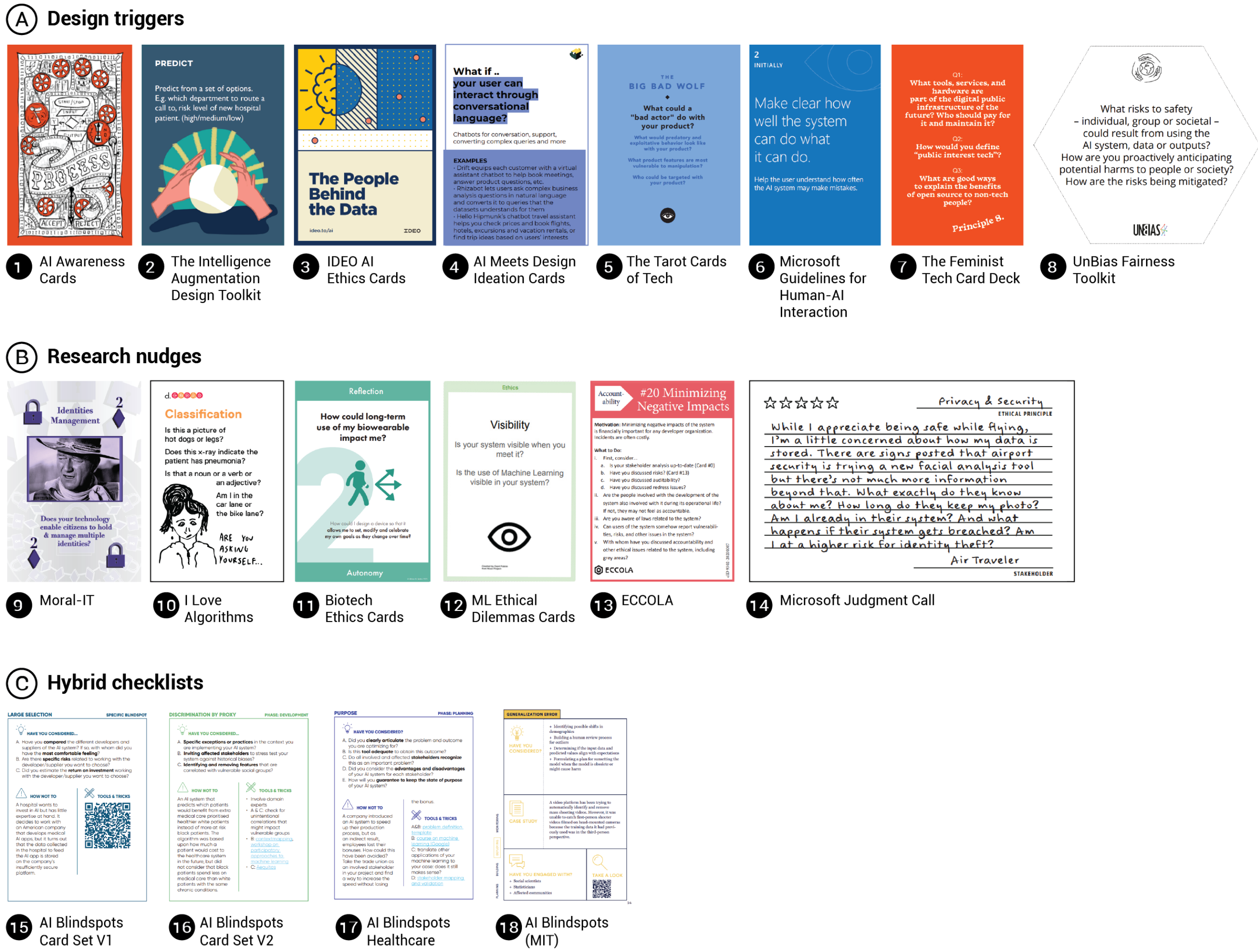

Fig. 2: **Responsible AI cards categories.** Design triggers (1-8), research nudges (9-14), and hybrid checklists (15-18). Numbers refer to Table 1.

3.3 Cards Collection

To collect card decks that are relevant to our inclusion criteria (§3.2), we mainly focused on four resources:

- **Deckaholic** is one of the largest online collection of card decks [1]. However, out of 97 decks featured on the website, none referred to responsible AI practices.
- **Knowledge Center Data & Society** is a large collection of practical tools for legal, ethical, and societal aspects of AI and data-driven applications [4]. Out of 35 tools featured on the website, 4 were related to our search scope.
- **Meta-reviews** provide a survey of current research practices. We read three meta-reviews relevant to card decks [53, 60, 76] and checked which of their references (going back in time) and citations (going forward in time) were related to responsible AI. Out of 230 decks found in these papers, 3 of them were included in our analysis.
- **Industry-driven Cards** are listed on major technology companies' websites (e.g., Google Research, Microsoft AI, Amazon AWS, Meta AI) and a subset of them is related to responsible AI. Out of 6 tools found in these websites, 2 of them were included in our analysis.

After collecting all the card decks (N=368), through discussions and meetings among three authors, a total of **18** card decks met our inclusion and exclusion criteria (§3.2). The final set of 18 cards includes three card decks [43–45] that originate from the initial deck quantifying blind spots of AI systems [16]. Because these three decks differ in the dimensions upon which we thematically reviewed them (as we shall see next, they differed in the dimensions of industry of origin, use phase, methodology of use and formal qualities), we decided to include all of them in the analysis.

3.4 Methods for Analyzing Cards

Our analysis includes three steps: *a)* descriptive analysis of 18 responsible AI card decks; *b)* categorization of these decks with K-modes clustering algorithm; and *c)* evaluation of the decks using 19 design guidelines for assessing content quality and visual properties.

3.4.1 Descriptive Analysis

To describe the 18 card decks, three authors collaboratively collected 9 descriptive dimensions (e.g., type, origin, purpose in Table 1) from two meta-analysis publications [60, 76]. Then, the three authors inspected these cards and used the 9 descriptive dimensions to characterize them. Any conflicts were resolved through discussions.

3.4.2 Categorization with K-modes

The purpose of categorizing the 18 card decks was to understand current trends and patterns in the design and development of responsible AI card decks. Given that the 9 descriptive dimensions are categorical, we used the K-modes clustering algorithm (similar to K-means used for numerical data). The algorithm is an unsupervised machine learning

Table 2: Design guidelines used in the evaluation were grouped in two main categories: content quality [36] and visual properties [27]. Distributions show the extent to which the 18 cards decks score in each guideline, ranging from “strongly disagree” to “strongly agree”.

Category	Guideline	Guideline description	Distribution
Content Quality	Understandable	Information can be easily understood by expert and non-expert producers.	
	Relevant	Prompts, guidelines, and examples are relevant and can be implemented effectively.	
	Effective	Users can produce appropriate answers to most questions with little confusion.	
	Multi-party adaptable	The card deck is helpful to all stakeholders.	
	Multi-stage adaptable	Questions are coherent, complete, and appropriately describe all stages of an AI’s lifecycle.	
	Deep	Questions and prompts clearly state the fidelity of information or specification.	
	Recognizable	Sections are organized logically so that users can recognize context and where to provide information.	
	Standard-compliant	The card deck uses industry-standard terms, or offers an additional context where relevant.	
	Clear	The same term means the same concept, every time it’s used.	
	Graspable	The relative meaning and importance of keywords and visual summaries are easy to grasp.	
	Effortless	The goals of questions and expectations of responses are clear at a glance.	
	Contextualizable	The card deck solicits background knowledge and context for enriching user’s understanding.	
Visual Properties	Visually meaningful	Visual elements (e.g., typography, colors, and symbols) of the cards are meaningfully presented.	
	Actionable	Users’ workload is minimized with respect to the number actions to be accomplished on each card.	
	Customizable	Users can customize the card deck so that it fits their working strategies, habits, and tasks.	
	Legible	Users can easily orient due to the balanced distribution of cards’ layout, legibility, and space efficiency.	
	Consistent	Design choices (e.g., uniform look of cards) are coherent and maintained throughout the card deck.	
	Memorable	The user does not have to memorize a lot of information to carry out tasks.	
	Minimalist	The deck does not include any extra information that can distract the eye from seeing the card content.	

technique used to group data points together with multiple categorical attributes [20]. In our case, a data point represents a card deck (i.e., rows in Table 1), and the algorithm uses the Hamming distance as a dissimilarity measure to find the number of categorical attributes (i.e., columns in Table 1) that differ between cards. The resulting cluster centroids then are represented by the mode, that is, the most frequently occurring value for each categorical property within the cluster.

3.4.3 Criteria for Card Scoring

To score the card decks, we used two sets of design guidelines: *content quality* [36] and *visual properties* [27] (Table 2).

Content Quality. The first set of guidelines was developed by Google researchers to evaluate model cards’ transparency, design, content, and usability [36]. By developing the Data Cards [58], the researchers also provided a set of guidelines to evaluate against, and we use these guidelines in our study. In total, this set consists of 12 guidelines considering whether the content quality is: understandable, relevant, effective, multi-party and multi-stage adaptable, deep, recognizable, standard-compliant, clear, graspable, efortless, and contextualizable.

Visual Properties. The second set of guidelines comes from information visualization theory, and is typically used to identify common and crucial usability issues in visualizations [27, 75] and to foster visualization analysis skills [62]. The set consists of 10 guidelines, but we removed three of them because they were only relevant to cards with data-driven interaction mechanisms. The excluded guidelines refer to dataset reduction (as the cards did not have any visual elements derived from datasets), orientation and help (since the decks did not include interface functions like redoing or undoing of actions), and prompting (no interface recommendations for alternative actions or inputs). This left us with 7 guidelines considering whether visual properties are: visually meaningful, actionable, customizable, legible, consistent, memorable, and minimalist. Additionally, we slightly modified the phrasing of the guidelines to match the context of the card deck being analyzed—a standard practice when conducting visualization evaluation [47].

3.4.4 Card Scoring

To assess each card deck against the guidelines, three authors collaborated and assigned a joint score between 1 and 5 for each criterion, with 1 indicating strong disagreement and 5 indicating strong agreement. The Likert-type scale with 5 response points for each item is prevalently used for guideline compliance rating. The authors also noted areas where the content and design of the card deck could be improved.

In Section 4, we present the results of the descriptive analysis and evaluation against guidelines. Based on our assessment, we analyzed which guidelines were most frequently followed and violated. Subsequently, we distilled 2 design recommendations to ensure that the new card decks for responsible AI overcome the identified drawbacks.

3.5 Limitations of Methods

To maintain coherence with prior research on card decks, we began our selection process with established decks highlighted in meta-reviews [53, 60, 76] and then systematically augmented our collection with card decks documented in publicly available online sources until February 2023. It is possible that practitioners may use design cards that are not documented in research papers, relying instead on internal tools employed by industry experts. To address this limitation, we examined technologies listed on two landscape overview websites: the Machine learning, Artificial intelligence and Data Landscape website [72] and UK Artificial Intelligence Industry Landscape website [39]. However, this strategy did not yield any new card decks to be analyzed.

4 Results

4.1 Descriptive Analysis

Next, we discuss the results of our descriptive analysis of responsible AI cards along 9 dimensions, that is, *release date*, *medium*, *type*, *origin*, *purpose*, *phase*, *methodology*, *customization*, and *formal qualities*. Table 1 summarizes these results.

Date. Half of the card decks were published in 2019 (9 decks), followed by 2 decks in 2020, 3 decks in 2021, and 1 deck in 2022. One deck was released in 2017, and two decks in 2018.

Table 3: Three authors scored each card deck against the 19 design guidelines, ranging from 1 ("strongly disagree") to 5 ("strongly agree"). Cats: Cards are grouped together based on their category derived from K-modes clustering.

Cats	#N	Card Deck	Mean Score
Design triggers	1	AI Awareness Cards [56]	3.95
	2	The Intelligence Augmentation Design Toolkit [29]	2.53
	3	IDEO AI Ethics Cards [38]	**4.05**
	4	AI Meets Design Ideation Cards [7]	3.89
	5	The Tarot Cards of Tech [10]	3.79
	6	Microsoft Guidelines for Human-AI Interaction [8]	4.00
	7	The Feminist Tech Card Deck [67]	3.11
	8	UnBias Fairness Toolkit [57]	3.26
Research nudges	9	Moral-IT [73]	3.74
	10	I Love Algorithms [17]	3.26
	11	Biotech Ethics Cards [50]	3.58
	12	ML Ethical Dilemmas Cards [13]	3.42
	13	ECCOLA [74]	3.95
	14	Microsoft Judgment Call [11]	**4.42**
Hybrid Checklists	15	AI Blindspots Card Set V1 [43]	3.84
	16	AI Blindspots Card Set V2 [44]	4.05
	17	AI Blindspots Healthcare [45]	3.84
	18	AI Blindspots (MIT) [16]	**4.11**

Medium. All cards are intended for physical use during in-person brainstorming sessions. Potential card users can download the decks at no cost by accessing the PDF files and print them. Nevertheless, having card decks freely available to download is a notably positive practice in contrast to the reported accessibility of design card decks, which stands at 19% [5]. Four decks provide additional hybrid extensions to their cards, such as interactive online versions of these cards or QR codes that link to additional recommended resources. The length of the decks varied. Half of the decks contained 20 or less than 20 cards (9 decks), while seven of them had between 21 to 52 cards. The largest deck included 77 cards.

Type. We did not identify a singular predominant approach for how the cards' content was intended to promote responsible AI ideation. Instead, we observed two almost equally popular strategies: *i)* broadening the scope of AI development with thought-provoking prompts (8 decks) and *ii)* clarifying the concepts and key terms related to the AI system (6 decks). The examples of prompts include what-if questions, such as *"What if you can sense and respond to your users' emotions in real-time using facial expression detection?"* (AI meets design Ideation Cards [7]), or *"What's the worst headline about your product you can imagine?"* (The Tarot Cards of Tech [10]). Only four decks focused on introducing the components of the AI system.

We also did not come across any decks that predominantly utilized activities such as embodiment, construction, or storytelling for ideation support [60]. Although none of the decks solely contained stories, a few included short statements about system components and concepts. For example, the IDEO AI Ethics cards [38] give examples how the use of the AI guidelines led to a beneficial outcome for a stakeholder who was auditing the AI system. Contrarily, the Feminist Tech Card Deck [67] augments their guidelines with short fictional stories envisioning the future use of AI.

Origin. The card decks come from various industries, with academic research groups (9 decks) and design communities (6 decks) being the primary contributors. Less common were decks released by commercial companies (2 decks) or non-governmental organizations (1 deck). These shares are likely explained by the fact that seven decks are supported by research papers introducing the rationale behind the card design process and report card evaluation results.

Purpose. Cards were used for two important purposes. One purpose of using cards was to inspire responsible development and challenge traditional perspectives (6 decks), while the other was to involve all stakeholders in the process and foster mutual sensitivity and empathy (10 decks). Two decks were agenda-driven as they did not directly expose users to responsible AI practices but instead focused on specific algorithms and machine learning components that could be used for designing services.

Phase. Cards have demonstrated their usefulness in various stages of ensuring responsible AI practices. The first category comprises 5 decks that can be employed at the beginning of the process to shape system development. The second category included 7 decks that are beneficial for both the initial phase of idea generation and when a team, for example, needs to review the design, troubleshoot issues, and explore alternatives. The third category consisted of 6 decks that are utilized at specific points of the process, such as during a dedicated workshop with all stakeholders.

Methodology. When examining the usage methodology, 10 decks provide recommendations on how to utilize the cards, whereas 8 decks come with highly specific card instructions, such as rules for card games or predefined activities to perform. The recommended duration for using the cards was seldom specified (16 decks). However, when provided, the suggested time frame ranged from 5 minutes to 180 minutes.

Customization. The decks largely constitute a self-contained body of knowledge on responsible AI practices, with 11 decks lacking significant user customization opportunities. When customization is available (6 decks), it is primarily accomplished by providing blank cards whereby users can record their comments and then add the custom card to the stack. Only one deck includes empty templates that necessitate customization by the user.

Formal Qualities. Out of the total number of decks, the majority (14 decks) offer a thematic structure for the cards and provide guidance on how they are interrelated, often by categorizing them. Most cards (13 decks) include imagery and text, while only 5 decks exclusively contain imagery or text. We also observed visual and style differences across the decks, with some featuring photographs, digital collages, metaphorical drawings, comics, or abstract forms and gradients.

4.2 Categorization with K-modes

The K-mode clustering revealed three categories (clusters) of responsible AI cards: *design triggers*, *research nudges*, and *hybrid checklists (Figure 2)*.

Design triggers (8 decks) are cards that rely on thought-provoking prompts or abstract visuals to stimulate divergent thinking during AI ideation (Figure 2A). A typical "design trigger" deck is segmented into specific categories (e.g., human-AI interactions, AI system errors, AI use cases [29]), without allowing for any card modifications. Trigger cards can be employed throughout the entire process of AI ideation, including problem analysis, development, implementation, evaluation and iteration. Due to their open-ended framing, this type of cards provides only general recommendations for usage. Design triggers typically originate from design companies and creative agencies. Yet, the suggested audience of the triggers is very broad, including youth (UnBias Fairness Toolkit [56]), policy makers (AI For Decision Makers [57]), non-profit organisations (The Feminist Tech Card Deck [67]), designers, managers and innovators (AI meets design Ideation Cards [7], The Tarot Cards of Tech [10]).

Research nudges (6 decks) are cards that provide insights into the various approaches and methods used in the development and implementation of AI systems. (Figure 2B). They encompass a broad range of AI techniques, including components, concepts, and prompts. These cards are intended to be introduced at designated points of use, such as workshops, and therefore come with detailed instructions. These cards are designed by either academic or industry researchers and are backed by scientific evidence presented in research papers. Moreover, the proposed use of nudges is mostly education-centered—academic researchers design nudges for youth (Biotech Ethics Cards [50]), high school students (ML Cards [13], I Love Algorithms [17]), and university students and researchers (ECCOLA [74], Moral-IT [73]).

Hybrid checklists (4 decks) are cards that feature a predefined list

Table 4: Rotated Component Matrix for the 6 principal components derived from scoring the 18 responsible AI cards against the 19 design guidelines: *Informative*; *Curated Customization*; *Adaptable*; *Comprehensible*; *Visually Appealing*; and *Efortless*.

Guideline	Informative	Curated Customization	Adaptable	Comprehensible	Visually Appealing	Effortless
1. Consistent	.942					
2. Deep	.863					
3. Effective	.624					
4. Relevant	.578					
5. Recognizable		.889				
6. Customizable		.538				
7. Standard-compliant		.526				
8. Multi-stage adaptable			.890			
9. Multi-party adaptable			.705			
10. Minimalist				.808		
11. Graspable				.774		
12. Legible				.664		
13. Visually meaningful					.969	
14. Effortless						.529

component loadings below .5 are suppressed.

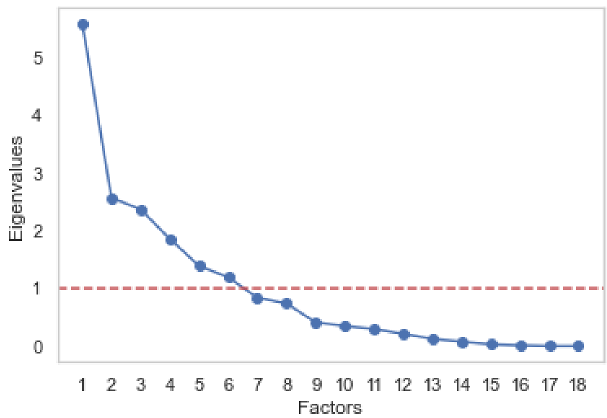

Fig. 3: **Scree plot.** It shows the principal components (the x-axis) and their respective eigenvalues (the y-axis) in a descending order.

of items (e.g., blindspots or potential harms) and external links that a team needs to verify, check, or inspect for an AI system (Figure 2C). While the decks are intended to be used as physical cards, their specific elements such as embedded links are only accessible when the cards are viewed digitally. The purpose of these elements is to provide additional context and information for the checklist items, utilizing external sources of knowledge such as research papers, blog posts, and industry websites. Hybrid checklists are typically the result of collaborations between universities and companies. This category is aimed at teams, companies, and organizations involved in developing or implementing AI systems in their operations.

4.3 Card Scoring

Tables 1 and 3 show the results of our evaluation of the 18 card decks against the guidelines for content quality and visual properties.

From a content quality standpoint, all card decks performed very well in offering **clarity** of meanings every time a concept was used (μ = 5), for example, by using **standard** terms for these concepts (μ = 4.44). The decks were also characterized by the right amount of **depth** of questions and prompts (μ = 4.78). From the visual properties perspective, the decks achieved highest scores in ensuring the **consistency** of the design (μ = 4.78), minimizing the initial information needed to start using cards, that is, **memorable** (μ = 4.78), and reducing the number of visual distractions, that is **minimalist**, (μ = 4.44).

Yet, the evaluation against the 19 guidelines uncovered multiple weaknesses in the design space. In terms of content quality, the cards received a low score (μ = 2.17) for their effectiveness in eliciting background knowledge and context from all users (**contextual**). Moreover, the decks were not **complete** as they lacked coherent activities to appropriately describe all stages of an AI's lifecycle (μ = 2.77). Deck users were also not equipped in any additional tools to customize the cards so they can fit their working strategies, habits and tasks, that is, **customization** (μ = 1.94). Finally, while the visual design was consistent, the **visual meaning** of typography, colors, and symbols could be enriched (μ = 2.39).

We also observed that certain guidelines had similar wording. By testing the correlation matrix between the 19 guidelines (rows were the 18 card decks, and columns the 19 guidelines), it showed significant correlations of above .3 at least one other item (guideline), suggesting reasonable factorability. To this end, we performed a factor analysis to identify the main orthogonal dimensions that explain most of the variance in the cards' scores.

First, we performed an adequacy check to test whether factor analysis is feasible, that is, the Bartlett's Test of Sphericity. The Bartlett's Test of Sphericity was found to be significant ($\chi^2(17)$ = 905.62, p < .01). Then, to determine the number of components that explain most of the variance in our card decks scores, we inspected the eigenvalues of the principal components. The eigenvalue of a principal component reflects the variance in all the variables explained by this component, and a higher value indicates higher variance in the variables loaded together in this component. The Kaiser's rule [40] suggests eigenvalues above 1.0 to be retained (resulting in 6 principal components), while the Cattell [18] proposed a visual inspection of the Scree Plot (Figure 3) to determine the 'elbow' point in which a leveling effect is observed. This resulted in 6 principal components that explained 75% of the total variance in the data. We considered the loadings of 14 items (guidelines) of the 6 principal components, while loadings of 5 items that were below .5 were suppressed.

Table 4 shows the resulting six principal components and their items (guidelines). The first component, **informative**, refers to cards' consistent look (consistent), sufficient depth of information (deep), effectiveness in decision making (effective), and relevance to the problem (relevant) at hand. The second component, **curated customization**, indicates that cards' content is easy to recognize (recognizable), compliant with industry-wide standards (standard-compliant), and can be customized to meet stakeholders' requirements (customizable). The third component, **adaptable**, refers to cards' adaptability to multiple stakeholders (multi-party adaptable) and multiple AI development stages (multi-stage adaptable). The fourth component, **comprehensible**, refers to whether cards' content is clear and readable, for example, through balancing card composition (legible), using relevant keywords and categories (graspable), and avoiding the use of distracting visual elements (minimalist). The fifth component, **visually appealing**, considers whether cards' content is presented using well-thought visual elements (visually meaningful). The sixth component, **effortless**, speaks to cards' ease of use.

To make use of these six components in future evaluations studies of responsible AI cards (or broadly card-based visualization systems), we paraphrased them into six questions that can be answered in a Likert scale from 1 to 5. In the case of **adaptable**, for example, 1 would indicate "not adaptable at all" and 5 would indicate "very adaptable".

$Q_{Adaptable}$: How adaptable the card deck is to multiple stakeholders or multiple development stages?

$Q_{Effortless}$: How effortless is the use of the card deck?

$Q_{Informative}$: How informative the card deck is based on its look and content?

$Q_{Comprehensible}$: How comprehensible the card deck is?

$Q_{Visually\_appealing}$: How visually appealing the card deck is?

$Q_{Curated\_Customization}$: How standard-compliant and customizable the card deck is?

5 Design Recommendations

As discussed in the previous section, the evaluation of the decks against the guidelines revealed two major shortcomings.

First, existing decks do not support **storage of and reflection upon knowledge** from involved stakeholders, regardless of their location, levels of expertise and time of contribution. The current design of physical decks presents challenges in maintaining a comprehensive

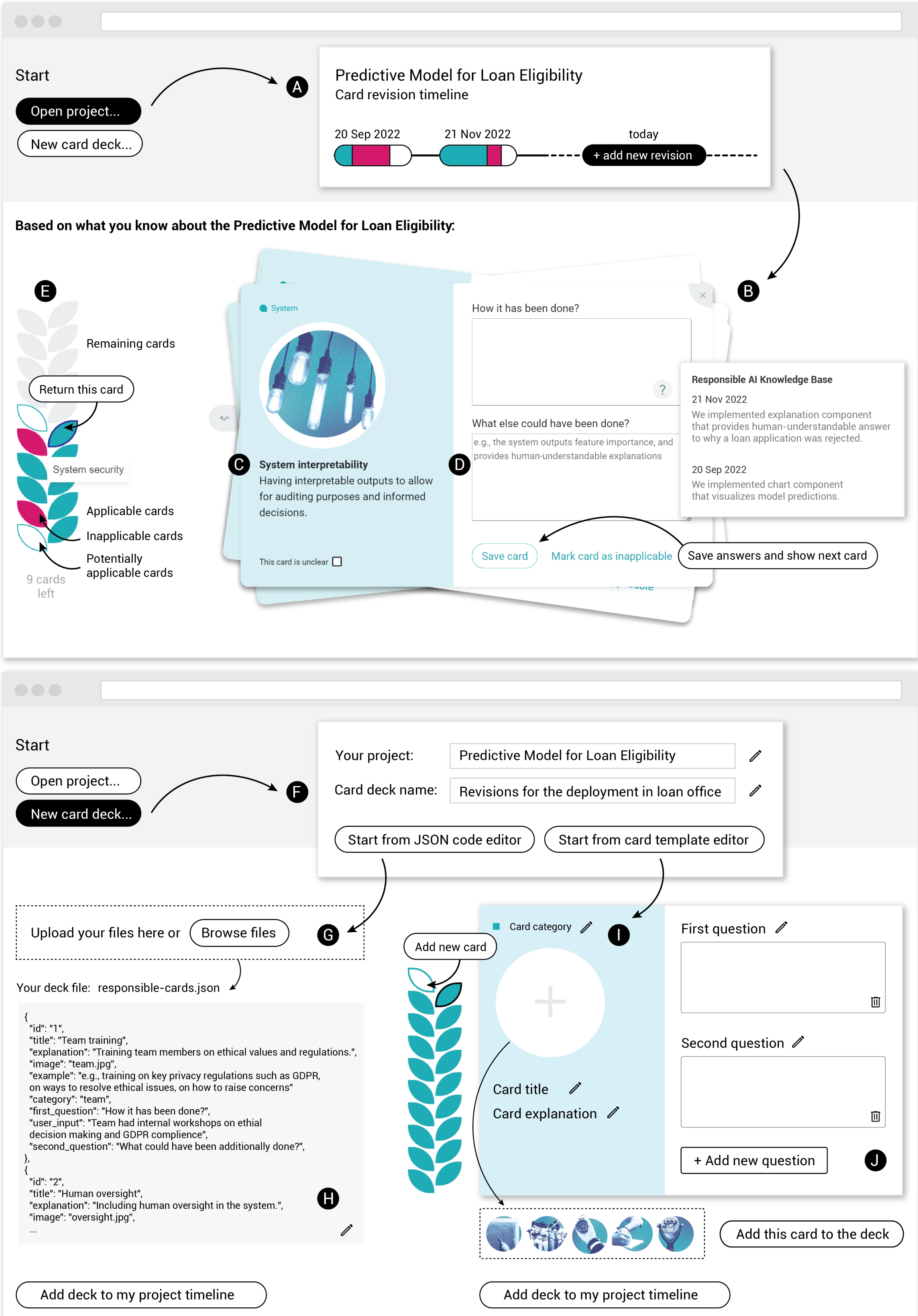

Fig. 4: **Design recommendations: soliciting and storing background knowledge (top) and card customization (bottom).** The new digital interface allow for: tracking the progress of an AI system based on previous cards' answers (A), initiating card revisions (B), displaying general recommendations to ensure responsible AI development (C), contextualizing each recommendation in the AI system at hand (D), obtaining real-time feedback on whether a card is applicable, inapplicable, or potentially applicable in a future iteration of the AI system (E), customization card content (F) via source code (G, H) or template edits (I, J).

record of previous card usage and reflecting on the ongoing iterations of the AI. While the hybrid checklist card archetypes try to tap to external sources of knowledge such as research papers or blog posts, new cards can elevate the internal knowledge of the stakeholders. As such, new functionalities could explore the opportunities of storing, arranging and clustering cards with key concepts, events, people, or decisions that are related to the developed AI system. However, these possibilities are only accessible when the cards are viewed digitally.

Second, users are not able to **customize** existing cards for their specific AI systems and their typical workflows for developing them. Cards typically focused on a single stage of the dataset or model's lifecycle and did not account for different stakeholders, processes, phases, and social implications of the system. As a result, existing card decks have limited applicability, making it challenging to use or adapt one single deck for an entire project (e.g., an AI project). For instance, at the beginning of the AI project, there might be a need to use flashcards to test and agree on key terms, concepts, or definitions that the stakeholders needs to memorize. Later in the project, the interest can be moved to model cards that document AI's intended use, performance characteristics, and limitations.

To solve these issues, we recommend to extend the possibilities of physical card decks by implementing new visual tools that allow for: (1) storage and reflection on previous card interactions; and (2) customization through editable decks and the ability to add new cards. To do so, we start by describing a new card deck system, and then its two functionalities.

5.1 Card Deck

To demonstrate these recommendations, we make use of a card deck that is comprised of 20 cards, with each card describing a technique that a developer can implement to ensure responsible AI development. Let us examine a scenario of an AI that is deciding who is eligible for a loan. In that case, a card in our fictional card deck would prompt a developer to report implementation details in relation to, for example, system interpretability.

Creating digital interface for responsible AI techniques requires creating new visual representations for each card and providing functionalities that go beyond the existing card design space. For example, to conform to the size limits of the digital screens, technique cards will are stacked on each other. Each card has two sides (Figure 4). The left side presents the ***general recommendation*** by featuring a title for the recommendation, a concise textual explanation, and a symbolic graphic illustration (Figure 4C). The right side of the card is a placeholder for the ***user-contextualized recommendation***. Its initial state contains two questions and two input fields with examples of how a recommendation can be implemented in an AI. Users can then replace the examples in the input fields with their context-dependent implementation of the recommendation (Figure 4D).

Once the user submits the card, the internal scoring system categorizes the user's input and provides real-time feedback on whether a card is applicable, inapplicable, or potentially applicable in future development iterations of the system. As the user interacts with the cards, the counter on the right side indicates the ongoing assignment of the cards to the three different groups and displays the number of remaining cards (Figure 4E). Applicable cards are indicated by green leaves, potentially applicable are indicated by pink leaves, and inapplicable are indicated by empty leaves. Each card can be returned to the stack by clicking on the associated leaf. Once the card process is completed, the user receives the process summary and can download the cards as a printable PDF or an editable JSON file.

5.2 Storage and Reflection

To facilitate learning and help to reinforce ethical considerations throughout the AI lifecycle, our proposed interface allows for storing and revising cards (i.e., keeping a history). This can be done in two ways. The interface allows for tracking changes (Figure 4A), and initiating new card revisions (Figure 4B). When tracking changes, the color coding shows the proportions of cards that are applicable or not at a given point in time. When revising a card's content, the input fields display recommendations based on users' previous answers that are stored in a so-called "responsible AI knowledge base". In that way, distributed teams can benefit by sharing these recommendations to build a common understanding about the project at hand.

5.3 Customization

Card decks should be easily customized to suit different audiences and topics related to responsible AI practices. Moreover, digital cards offer the possibility of more frequent updates that ensure users have access to the most up-to-date information. Two different functionalities can accommodate these needs (Figure 4F). The first functionality, the card code editor, allows loading the JSON card file with the initial 20 recommendations. Users can then edit the card content by changing the source code of the deck (Figure 4G, H). The second functionality aims to provide a simple card template that would allow adding a new card to an existing deck or building a new card deck (Figure 4I, J). Users can edit the left side of the card by giving the name of the recommendation, a concise textual explanation, and a symbolic graphic illustration. The right side of the card can be edited by adding and removing new questions. The deck can be then added to the project as a new point on the project timeline.

6 Conclusion

By reviewing 18 responsible AI cards that were developed and published between 2017-2022, we provided a systematic mapping of the design space at the intersection of card designs and responsible AI. First, we identified three card categories, that is, *design triggers*, *research nudges*, and *hybrid checklists*. Design triggers typically come from design companies and creative agencies, and are used to stimulate divergent thinking about AI using prompts; nudges are typically created by academic or industry researchers, and are used for education purposes; hybrid checklists are typically the result of collaboration between universities and companies, are used to verify, check, or inspect an AI system. Second, by evaluating these cards against 19 design guidelines, we found that the state-of-the-art cards are straightforward to use, understandable by their intended users, and consistent in design choices without distracting visual elements. However, we found that the physical form of existing cards makes them unsuitable for use in distributed teams in addition to lack of customization. To this end, we proposed two design recommendations for allowing storage and reflection on previous card interactions, and customization through editable decks and the ability to add new cards.

References

[1] Deckaholic. http://www.deckaholic.com. Accessed: March 2023. 4

[2] IBM AI Explainability 360. https://aix360.mybluemix.net. Accessed: March 2023. 1

[3] IBM AI Fairness 360. https://aif360.mybluemix.net. Accessed: March 2023. 1, 2

[4] Knowledge Centre Data & Society. https://data-en-maatschappij.ai/en/tools. Accessed: March 2023. 4

[5] T. Aarts, L. K. Gabrielaitis, L. C. de Jong, R. Noortman, E. M. van Zoelen, S. Kotea, S. Cazacu, L. L. Lock, and P. Markopoulos. Design card sets: Systematic literature survey and card sorting study. In *Proceedings of the 2020 ACM Designing Interactive Systems Conference*, DIS '20, p. 419–428. Association for Computing Machinery, New York, NY, USA, 2020. doi: 10.1145/3357236.3395516 6

[6] M. S. Ackerman. The intellectual challenge of CSCW: The gap between social requirements and technical feasibility. *Human–Computer Interaction*, 15(2-3):179–203, 2000. doi: 10.1207/S15327051HCI1523_5 1

[7] AIxDesign. AI Meets Design Ideation Cards. https://www.aixdesign.co/toolkit, 2019. Accessed: February 2023. 3, 6

[8] S. Amershi, D. Weld, M. Vorvoreanu, A. Fourney, B. Nushi, P. Collisson, J. Suh, S. Iqbal, P. Bennett, K. Inkpen, J. Teevan, R. Kikin-Gil, and E. Horvitz. Guidelines for Human-AI interaction. In *CHI 2019*. ACM, May 2019. CHI 2019 Honorable Mention Award. doi: 10.1145/3290605.3300233 2, 3, 6

[9] N. Andrienko, G. Andrienko, L. Adilova, and S. Wrobel. Visual analytics for human-centered machine learning. *IEEE Computer Graphics and Applications*, 42(1):123–133, 2022. doi: 10.1109/MCG.2021.3130314 2

[10] Artefact Group. The Tarot Cards of Tech. https://www.artefactgroup.com/case-studies/the-tarot-cards-of-tech/, 2018. Accessed: February 2023. 3, 6

[11] S. Ballard, K. M. Chappell, and K. Kennedy. Judgment Call the Game: Using Value Sensitive Design and Design Fiction to Surface Ethical Concerns Related to Technology. In *Proceedings of the 2019 on Designing Interactive Systems Conference*, DIS '19, p. 421–433. ACM, 2019. doi: 10.1145/3322276.3323697 1, 2, 3, 6

[12] A. Barredo Arrieta, N. Díaz-Rodríguez, J. Del Ser, A. Bennetot, S. Tabik, A. Barbado, S. Garcia, S. Gil-Lopez, D. Molina, R. Benjamins, R. Chatila, and F. Herrera. Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. *Information Fusion*, 58:82–115, 2020. doi: 10.1016/j.inffus.2019.12.012 3

[13] K.-E. K. Bilstrup, M. H. Kaspersen, and M. G. Petersen. Staging reflections on ethical dilemmas in machine learning: A card-based design workshop for high school students. In *Proceedings of the 2020 ACM Designing Interactive Systems Conference*, DIS '20, p. 1211–1222. Association for Computing Machinery, New York, NY, USA, 2020. doi: 10.1145/3357236.3395558 3, 6

[14] S. Bird, M. Dudík, R. Edgar, B. Horn, R. Lutz, V. Milan, M. Sameki, H. Wallach, and K. Walker. Fairlearn: A toolkit for assessing and improving fairness in AI. Technical report, Microsoft, May 2020. 2

[15] M. Boone, N. Pope, C. Xiao, and A. Anandkumar. Enhancing AI Transparency and Ethical Considerations with Model Card++, 2022. Accessed: March 2023. 2

[16] A. Calderon, D. Taber, H. Qu, and J. Wen. AI Blindspots (MIT). https://aiblindspot.media.mit.edu/, 2019. Accessed: February 2023. 3, 4, 6

[17] C. Carter, M. Stariha, K. Schmutte, T. Forshaw, S. Harte, M. Grundberg, and D. Frumhoff. I love Algorithms. https://dschool.stanford.edu/resources/i-love-algorithms, 2019. Accessed: February 2023. 3, 6

[18] R. B. Cattell. The scree test for the number of factors. *Multivariate Behavioral Research*, 1(2):245–276, 1966. doi: 10.1207/s15327906mbr0102_10 7

[19] C. Ceccarini, E. P. Bogucka, I. Sen, M. Constantinides, C. Prandi, and D. Quercia. Visualizing internal sustainability efforts in big companies. *IEEE Computer Graphics and Applications*, 42(3):87–98, 2022. doi: 10.1109/MCG.2022.3163063 2

[20] A. Chaturvedi, P. E. Green, and J. D. Caroll. K-modes clustering. *Journal of classification*, 18:35–55, 2001. doi: 10.1007/s00357-001-0004-3 5

[21] J. Choo and S. Liu. Visual analytics for explainable deep learning. *IEEE Computer Graphics and Applications*, 38(4):84–92, 2018. doi: 10.1109/MCG.2018.042731661 2

[22] S. Clatworthy. Service innovation through touch-points: Development of an innovation toolkit for the first stages of new service development. *International Journal of Design*, 5:15–28, 08 2011. 2

[23] V. Dignum. *Responsible artificial intelligence: how to develop and use AI in a responsible way*. Springer, 2019. 1

[24] U. Ehsan and M. O. Riedl. Human-centered explainable ai: Towards a reflective sociotechnical approach. In *HCI International 2020-Late Breaking Papers: Multimodality and Intelligence: 22nd HCI International Conference, HCII 2020, Copenhagen, Denmark, July 19–24, 2020, Proceedings 22*, pp. 449–466. Springer, 2020. doi: 10.1007/978-3-030-60117-1_33 1

[25] U. Ehsan, K. Saha, M. De Choudhury, and M. O. Riedl. Charting the sociotechnical gap in explainable ai: A framework to address the gap in xai. *arXiv preprint arXiv:2302.00799*, 2023. 1

[26] Fairlearn. Improve fairness of AI systems. https://fairlearn.org, 2021. Accessed: November 2022. 2

[27] C. Forsell and J. Johansson. An heuristic set for evaluation in information visualization. In *Proceedings of the International Conference on Advanced Visual Interfaces*, AVI '10, p. 199–206. Association for Computing Machinery, New York, NY, USA, 2010. doi: 10.1145/1842993.1843029 5

[28] H. Frluckaj, L. Dabbish, D. G. Widder, H. S. Qiu, and J. D. Herbsleb. Gender and Participation in Open Source Software Development. *Proc. ACM Hum.-Comput. Interact.*, 6(CSCW2), Nov 2022. doi: 10.1145/3555190 2

[29] Futurice. The Intelligence Augmentation Design Toolkit. https://futurice.com/ia-design-kit, 2017. Accessed: February 2023. 3, 6

[30] T. Gebru, J. Morgenstern, B. Vecchione, J. W. Vaughan, H. Wallach, H. D. III, and K. Crawford. Datasheets for Datasets. *Commun. ACM*, 64(12):86—-92, nov 2021. doi: 10.1145/3458723 2

[31] Google. Know Your Data. https://knowyourdata.withgoogle.com, 2021. Accessed: March 2023. 2

[32] Google. AI Explorables. https://pair.withgoogle.com/explorables, 2022. Accessed: November 2022. 2

[33] Google. Responsible AI practices. https://ai.google/responsibilities/responsible-ai-practices, 2022. Accessed: February 2023. 3

[34] Google. Using the What-If Tool. https://cloud.google.com/ai-platform/prediction/docs/using-what-if-tool, 2022. Accessed: November 2022. 1

[35] Google Research. Google Cloud Model Cards. https://modelcards.withgoogle.com/model-reports, 2023. Accessed: March 2023. 3

[36] Google Research. Heuristics for Transparency Artifacts, 2023. Accessed: March 2023. 5

[37] L. Havens, M. Terras, B. Bach, and B. Alex. Situated Data, Situated Systems: A Methodology to Engage with Power Relations in Natural Language Processing Research. In *Proceedings of the Second Workshop on Gender Bias in Natural Language Processing*, pp. 107–124. Association for Computational Linguistics, Dec. 2020. 2

[38] IDEO. AI needs ethical compass. This tool can help. https://www.ideo.com/blog/ai-needs-an-ethical-compass-this-tool-can-help, 2019. Accessed: November 2022. 1, 3, 6

[39] Innovation Eye. UK Artificial Intelligence Industry Landscape. https://mindmaps.innovationeye.com/reports/ai-in-uk, 2021. Accessed: March 2023. 5

[40] H. F. Kaiser. The application of electronic computers to factor analysis. *Educational and Psychological Measurement*, 20(1):141–151, 1960. 7

[41] H. Kissinger, E. Schmidt, and D. P. Huttenlocher. *The age of AI: And our human future*. John Murray London, 2021. 1

[42] I. Kloumann and J. Tannen. How we're using Fairness Flow to help build AI that works better for everyone. https://ai.facebook.com/blog/how-were-using-fairness-flow-to-help-build-ai-that-works-better-for-everyone, 2021. Accessed: November 2022. 1

[43] Knowledge Centre Data and Society. AI Blindspots Card Set V1. https://data-en-maatschappij.ai/en/tools/ai-blindspot, 2019. Accessed: February 2023. 3, 4, 6

[44] Knowledge Centre Data and Society. AI Blindspots Card Set V2. https://data-en-maatschappij.ai/en/tools/ai-blindspots-2.0, 2019. Accessed: February 2023. 3, 4, 6

[45] Knowledge Centre Data and Society. AI Blindspots Healthcare. https://data-en-maatschappij.ai/en/tools/ai-blindspots-healthcare, 2019. Accessed: February 2023. 3, 4, 6

[46] J. Kwiatkowska, W. Bartkowski, O. Górnicka, and D. Lamas. Applying generative techniques to avoid technology push effect in ideas and prototypes created by technology-oriented people. In *Proceedings of the Mulitimedia, Interaction, Design and Innnovation*, MIDI '15. Association for Computing Machinery, New York, NY, USA, 2015. doi: 10.1145/2814464.2814480 2

[47] R. Langevin, R. J. Lordon, T. Avrahami, B. R. Cowan, T. Hirsch, and G. Hsieh. Heuristic evaluation of conversational agents. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, CHI '21. Association for Computing Machinery, New York, NY, USA, 2021. doi: 10.1145/3411764.3445312 5

[48] Microsoft. AI Lab project: Responsible AI dashboard. https://www.microsoft.com/en-us/ai/ai-lab-responsible-ai-dashboard, 2021. Accessed: March 2023. 2

[49] Microsoft. Responsible AI. https://www.microsoft.com/en-us/ai/responsible-ai, 2022. Accessed: February 2023. 3

[50] Z. Minh-Tam Dao-Kroeker, A. Kitson, A. N. Antle, Y. Murai, and A. Adibi. Designing biotech ethics cards: Promoting critical making during an online workshop with youth. In *Interaction Design and Children*, IDC '21, p. 450–455. Association for Computing Machinery, New York, NY, USA, 2021. doi: 10.1145/3459990.3465182 3, 6

[51] National Institute of Standards and Technology. AI Risk Management Framework. https://www.nist.gov/itl/ai-risk-management-framework. Accessed: February 2023. 3

[52] Nokia Bell Labs. Responsible AI. https://www.bell-labs.com/research-innovation/responsible-ai, 2022. Accessed: February 2023. 3

[53] D. Peters, L. Loke, and N. Ahmadpour. Toolkits, cards and games – a review of analogue tools for collaborative ideation. *CoDesign*, 17:410–434, 02 2020. doi: 10.1080/15710882.2020.1715444 4, 5

[54] D. Peters, L. Loke, and N. Ahmadpour. Toolkits, cards and games – a review of analogue tools for collaborative ideation. *CoDesign*, 17(4):410–434, 2021. doi: 10.1080/15710882.2020.1715444 1, 2

[55] V. Philomin and P. Hallinan. Introducing AWS AI Service Cards: A new resource to enhance transparency and advance responsible AI, 2022. 3

[56] Proboscis. AI Awareness Cards. https://unbias.wp.horizon.ac.uk/2018/09/06/unbias-fairness-toolkit/, 2018. Accessed: February 2023. 3, 6

[57] Proboscis. UnBias Fairness Toolkit. http://proboscis.org.uk/6346/ai-for-decision-makers/, 2020. Accessed: February 2023. 3, 6

[58] M. Pushkarna, A. Zaldivar, and O. Kjartansson. Data cards: Purposeful and transparent dataset documentation for responsible ai. In *2022 ACM Conference on Fairness, Accountability, and Transparency*, FAccT '22, p. 1776–1826. Association for Computing Machinery, New York, NY, USA, 2022. doi: 10.1145/3531146.3533231 3, 5

[59] PwC. PwC's Responsible AI. https://www.pwc.com/gx/en/issues/data-and-analytics/artificial-intelligence/what-is-responsible-ai/pwc-responsible-ai.pdf, 2022. Accessed: February 2023. 3

[60] R. Roy and J. P. Warren. Card-based design tools: a review and analysis of 155 card decks for designers and designing. *Design Studies*, 63:125–154, 2019. doi: 10.1016/j.destud.2019.04.002 4, 5, 6

[61] P. Saleiro, B. Kuester, L. Hinkson, J. London, A. Stevens, A. Anisfeld, K. T. Rodolfa, and R. Ghani. Aequitas: A bias and fairness audit toolkit. *arXiv preprint arXiv:1811.05577*, 2018. 2

[62] B. S. Santos, B. Q. Ferreira, and P. Dias. Using heuristic evaluation to foster visualization analysis and design skills. *IEEE Computer Graphics and Applications*, 36(1):86–90, 2016. doi: 10.1109/MCG.2016.7 5

[63] H. Shen, W. H. Deng, A. Chattopadhyay, Z. S. Wu, X. Wang, and H. Zhu. Value cards: An educational toolkit for teaching social impacts of machine learning through deliberation. In *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*, FAccT '21, p. 850–861. Association for Computing Machinery, New York, NY, USA, 2021. doi: 10.1145/3442188.3445971 3

[64] B. Shneiderman. Bridging the gap between ethics and practice: Guidelines for reliable, safe, and trustworthy human-centered ai systems. *ACM Trans. Interact. Intell. Syst.*, 10(4), oct 2020. doi: 10.1145/3419764 1

[65] B. Shneiderman. Bridging the gap between ethics and practice: Guidelines for reliable, safe, and trustworthy human-centered ai systems. *ACM Trans. Interact. Intell. Syst.*, 10(4), oct 2020. doi: 10.1145/3419764 2

[66] B. Shneiderman. *Human-centered AI*. Oxford University Press, 2022. 1

[67] SUPERRR Lab. The Feminist Tech Card Deck. https://superrr.net/feministtech/deck, 2022. Accessed: November 2022. 1, 3, 6

[68] M. Tahaei, M. Constantinides, D. Quercia, S. Kennedy, M. Muller, S. Stumpf, Q. V. Liao, R. Baeza-Yates, L. Aroyo, J. Holbrook, et al. Human-centered responsible artificial intelligence: Current & future trends. *arXiv preprint arXiv:2302.08157*, 2023. 1

[69] The High-Level Expert Group on Artificial Intelligence. The Assessment List for Trustworthy Artificial Intelligence. https://altai.insight-centre.org, 2020. Accessed: November 2022. 2

[70] The Organisation for Economic Co-operation and Development (OECD). Recommendation of the Council on Artificial Intelligence, 2019. Accessed: February 2023. 3

[71] N. C. Thompson, K. Greenewald, K. Lee, and G. F. Manso. Deep learning's diminishing returns: the cost of improvement is becoming unsustainable. *IEEE Spectrum*, 58(10):50–55, 2021. doi: 10.1109/MSPEC.2021.9563954 2

[72] M. Turck and J. Wu. The 2023 MAD (Machine Learning, Artificial Intelligence and Data) Landscape. https://mad.firstmark.com/, 2023. Accessed: March 2023. 5

[73] L. D. Urquhart and P. J. Craigon. The moral-it deck: a tool for ethics by design. *Journal of Responsible Innovation*, 8(1):94–126, 2021. doi: 10.1080/23299460.2021.1880112 2, 3, 6

[74] V. Vakkuri, K.-K. Kemell, M. Jantunen, E. Halme, and P. Abrahamsson. Eccola — a method for implementing ethically aligned ai systems. *Journal of Systems and Software*, 182:111067, 2021. doi: 10.1016/j.jss.2021.111067 3, 6

[75] E. Wall, M. Agnihotri, L. Matzen, K. Divis, M. Haass, A. Endert, and J. Stasko. A heuristic approach to value-driven evaluation of visualizations. *IEEE Transactions on Visualization and Computer Graphics*, 25(1):491–500, 2019. doi: 10.1109/TVCG.2018.2865146 5

[76] C. Wölfel and T. Merritt. Method card design dimensions: A survey of card-based design tools. In P. Kotzé, G. Marsden, G. Lindgaard, J. Wesson, and M. Winckler, eds., *Human-Computer Interaction – INTERACT 2013*, pp. 479–486. Springer Berlin Heidelberg, Berlin, Heidelberg, 2013. doi: 10.1007/978-3-642-40483-2_34 1, 4, 5