The AI-DEC: A Card-based Design Method for User-centered AI Explanations

Christine P Lee
Department of Computer Sciences
University of Wisconsin–Madison
Madison, Wisconsin, USA
cplee5@cs.wisc.edu

Min Kyung Lee*
School of Information
The University of Texas at Austin
Austin, Texas, USA
minkyung.lee@austin.utexas.edu

Bilge Mutlu*
Department of Computer Sciences
University of Wisconsin–Madison
Madison, Wisconsin, USA
bilge@cs.wisc.edu

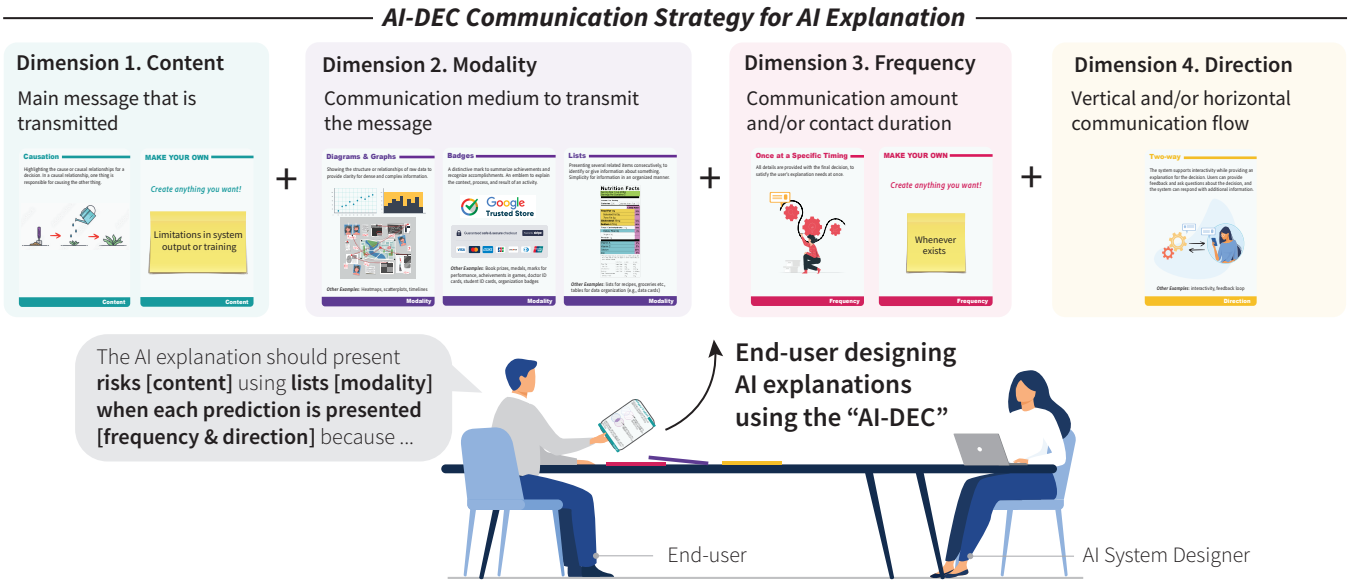


Figure 1: We introduce *AI-DEC*, a card-based design method for user-centered AI explanations. AI-DEC presents design examples for key communication dimensions of AI explanations: content, modality, frequency, and direction. The figure provides an example of our co-design session where a worker (*i.e.,* end-user) utilizes the AI-DEC to design personalized AI explanations.

ABSTRACT

Increasing evidence suggests that many deployed AI systems do not sufficiently support end-user interaction and information needs. Engaging end-users in the design of these systems can reveal user needs and expectations, yet effective ways of engaging end-users in the AI explanation design remain under-explored. To address this gap, we developed a design method, called *AI-DEC*, that defines four dimensions of AI explanations that are critical for the integration of AI systems—communication content, modality, frequency, and direction—and offers design examples for end-users to design AI explanations that meet their needs. We evaluated this method through co-design sessions with workers in healthcare, finance, and management industries who regularly use AI systems in their daily work. Findings indicate that the AI-DEC effectively supported workers in designing explanations that accommodated diverse levels of performance and autonomy needs, which varied depending on the AI system's workplace role and worker values. We discuss the implications of using the AI-DEC for the user-centered design of AI explanations in real-world systems.

CCS CONCEPTS

• **Human-centered computing** → **User studies**; **Empirical studies in HCI**; **User centered design**; **Participatory design**; **Interface design prototyping**.

KEYWORDS

Design cards, user-centered design, human-AI interaction

ACM Reference Format:
Christine P Lee, Min Kyung Lee, and Bilge Mutlu. 2024. The AI-DEC: A Card-based Design Method for User-centered AI Explanations. In *Designing Interactive Systems Conference (DIS '24), July 01–05, 2024, IT University of Copenhagen, Denmark.* ACM, New York, NY, USA, 19 pages. https://doi.org/10.1145/3643834.3661576

*Equal senior contribution.

This work is licensed under a Creative Commons Attribution International 4.0 License.

DIS '24, July 01–05, 2024, IT University of Copenhagen, Denmark

© 2024 Copyright held by the owner/author(s).

ACM ISBN 979-8-4007-0583-0/24/07
https://doi.org/10.1145/3643834.3661576

1 INTRODUCTION

Artificial Intelligence (AI) systems are increasingly deployed into real-world applications across many domains, including healthcare [56, 60, 144], finance [30, 35, 116], management [25, 98], education [4, 51, 52], and government [3, 63]. Such an increase is largely due to the rapid advancement in recent machine learning (ML) and AI fields, enabling AI systems to provide decision support or final decisions on workers' behalf. However, such deployment is often accompanied by adaptation and integration challenges [137], due to a mismatch of worker expectations, organizational issues, and technical constraints [111, 138]. A majority of these constraints and limited success of integration stem from a lack of consideration for end-users—users who actually engage with the AI system. These technologies are often developed without sufficient consideration of the deployed environment, end-users' interaction and information needs, and how users will use and adopt the AI system.

There has been increasing interest in designing AI systems to be human-centric, with a focus on making them comprehensible to end users and meeting their needs [23, 79]. Several design methods have been used to facilitate user-centered AI, including participatory design [75], co-design [10, 47, 74, 81, 145], field studies [46, 61, 62], and design workshops [6, 71, 103, 120, 123, 143]. However, there are still limited methods for effectively engaging end-users in explainable AI (XAI) design. Prior work has emphasized the importance of understanding user needs, situational factors, and deployed environment to design accessible and comprehensible AI explanations[36, 77, 140]; yet little research has engaged end-users in AI explanation design process. This difficulty is partially due to a limit in intuitive and accessible design methods that render AI explanations as accessible design material for end-users. Design methods that facilitate prototyping or collaboration with end-users have the potential to design AI explanations that better support various workplaces.

In this work, we developed a card-based design method called the ***AI-DEC***, (*i.e.,* **AI** Explanation **De**sign **C**ards) to specifically meet end-users informational and organizational needs for designing effective AI explanations. The AI-DEC defines four dimensions of AI explanations that are critical for the integration of AI systems in deployment environments: communication content, modality, frequency, and direction. The cards in the AI-DEC provide design examples for end-users to design AI explanations that meet their needs. The aim of the AI-DEC is to directly involve end-users in the design process of AI explanations, supporting them in prototyping and expressing their visions. We then used the AI-DEC for co-design sessions with 16 workers who regularly engage with AI systems in their workplace to explore whether the AI-DEC helped them design AI explanations aligned with their information and interaction needs. We recruited workers from three domains: healthcare, finance, and management to assess the applicability of the AI-DEC in various contexts. Our findings show that the AI-DEC assisted workers in designing AI explanations tailored to their performance and autonomy needs. These needs vary depending on the specific role of the AI system in the workplace and the workers' individual values. Finally, we present implications on how to use the AI-DEC in designing context-adaptive and user-centered AI explanations for real-world applications. Our contributions are the following:

- The AI-DEC, a card-based design method to engage end-users in the design process of AI explanations;
- Using the AI-DEC to co-design AI explanations with workers from various domains who possess empirical experience engaging with AI systems daily;
- Insights derived from worker design solutions on information needs for user-centered AI explanations in workplaces.

2 RELATED WORK

2.1 Advancements in Human-Centered AI Research and Design Approaches

An expanding body of research has concentrated on enhancing the human-centered design and interpretability of AI systems. Previous studies have employed diverse methods, such as design probes [117, 135], design workshops [6, 103, 120], prototyping [89, 115, 121], Wizard of Oz [131], design workbooks and toolkits [17, 42], participatory design [75], co-design [7, 14, 145], research through design [94, 114], and end-user evaluations [44, 84, 112, 148], resulting in guidelines for creating human-centered AI systems. Noteworthy contributions include the integration of human decision-making theory into AI explanations [134], emphasizing the consideration of situational factors in AI explanations [36], utilizing uncertainty for transparency [16], supporting end-user participation in AI decision-making [75], and providing guidelines for communicating AI trustworthiness [78]. These contributions collectively advance the goal of making AI systems more interpretable, thereby promoting their effective integration into various domains while addressing concerns regarding transparency, trust, and user engagement.

Simultaneously, another line of research has emerged, focusing on developing design tools and approaches to aid designers in crafting human-centered AI. Educational materials have been devised to enhance designers' understanding of AI system capabilities and technical intricacies. Interactive machine learning tools facilitate designers in comprehending technical capabilities and simulating user-system interactions [127]. Guidelines on addressing failures and managing uncertainty have been proposed. Design methods to comprehend user needs and foster collaboration with users have also been introduced. Examples include a "Question Bank" categorizing explanation needs and related user queries Liao et al. [77], a scenario-based design method addressing context-specific user requirements in complex settings Wolf [139], and a repository of research results and tools from the DARPA XAI program to aid designers in incorporating XAI techniques Hu et al. [55]. Additionally, data probes [122, 135] have been developed to facilitate AI co-design sessions with end-users. While work on understanding the needs for the user-centered design of AI systems has been actively conducted, there are limited design tools and methods developed to directly incorporate and support end-users in the design process of human-centered AI systems.

2.2 Card-based Design Methods in HCI

Design cards have been recognized as valuable approaches in HCI and design research [48, 54, 97]. Roy and Warren [110] categorized the main purpose of 155 sets of design cards into systematic design methods, human-centered design, or domain-specific design. Notable examples include IDEO Method Cards [58, 59], Envisioning

Cards from the Value Sensitive Design Lab [43], and the Design with Intent toolkit [82, 83]. The appeal of design cards lies in their simplicity, tangibility, and ease of use, making them an accessible tool to infuse information and inspiration into the design process [13, 53, 85, 141]. The main advantages of using design cards involve deepening design knowledge, fostering creativity, engaging stakeholders, and refining design practices [24, 53, 95, 110]. A thorough literature review of design cards by Hsieh et al. [54] identified seven different types of design knowledge that design cards are used to present: Creative Inspiration, Human Insights, Material & Domain, Methods & Tooling, Problem Definition, Team Building, and Values in Practice. However, the authors also highlight the need to overcome perceived barriers of designers that design cards lack usefulness and alignment with practical requirements. To optimize the effectiveness of design cards, there is a call for developing cards with usability and applicability in mind, incorporating clear guidelines, concrete examples, well-defined keywords, and flexibility for designer customization. Moreover, efforts should focus on supporting design stages beyond ideation, specifically in prototyping, implementation, and evaluation. Employing the advantages of intuitive operation, versatility, and tangible interaction, the AI-DEC utilizes design cards as a method to enhance end-user usability and applicability across various contexts.

3 DEVELOPMENT OF THE AI-DEC

In this section, we discuss the development procedure of the AI-DEC. Section 3.1 explains how the structure of the AI-DEC was built from a theoretical model of communication strategy. Subsequently, Section 3.2 discusses the procedure of selecting the design elements of the AI-DEC.

3.1 Communication Model as a Design Framework for AI Explanations

In settings where humans interact with one another, communication acts as the binding element that holds together various distributed channels [91, 92]. Its crucial role lies in facilitating the smooth flow of information, coordination, and decision-making among the different entities involved. Ineffective communication can lead to problems within these channels, causing individuals to feel excluded or frustrated due to misunderstandings [93]. The importance of communication extends to the deployment of AI systems in real-world environments where they engage with end-users [57]. The interaction between the AI system and end-users encompasses complex policies, numerous factors, and training processes along a pathway that is often not communicated to end-users due to the opaque nature of AI systems. Therefore, it is essential to design AI explanations with a communication principle in mind. Crafting AI explanations that center around communication strategies tailored to the deployment environment and end-user needs can enhance understanding, adoption, and satisfaction among end-users. This approach can also help address skepticism and distrust that might hinder the successful integration of AI systems during deployment, promoting transparency, effective functionality, and the seamless integration of the AI system.

The AI-DEC's structure is based on the theoretical model of communication strategies proposed by Mohr and Nevin [91] for organization and marketing channels. This model for channel communication reflects the diversity of channel settings and elicits the role of communication to attain enhanced levels of channel outcomes. In this context, a channel refers to the "channel of distribution," which is the pathway or route through which goods or services move from the producer or manufacturer to the end consumer. It encompasses the various intermediaries, such as wholesalers, retailers, and other middlemen, who facilitate the movement of products along the distribution process. The authors propose a contingency theory, in which communication strategy moderates the impact of channel conditions (structure, climate, and power) on channel outcomes (coordination, satisfaction, commitment, and performance). When a communication strategy matches the channel conditions, channel outcomes will be enhanced in comparison with the outcomes when a communication strategy mismatches channel conditions. Specifically, the communication strategy consists of a combination of communication dimensions of context, modality, frequency, and direction [41]. *Content* describes the "main message that is transmitted — or what is said." *Modality* represents the "medium of communication or the method used to transmit the information." *Frequency* involves the "amount of communication and/or duration of contact between organizational members." Finally, *direction* refers to the "vertical and/or horizontal movement of communication within the organizational hierarchy."

3.2 Design of the AI-DEC

From the communication model of Mohr and Nevin [91], we adopt its four-dimension structure, including content, modality, frequency, and direction, along with their associated definitions, to create a card-based design tool for AI explanation design. The AI-DEC incorporates this structured approach as a card deck, where each dimension consists of cards containing specific design elements.

To develop the design cards, we conducted a focused review on user-centered design in human-computer/AI interaction, interpretable ML, explainable AI, and design literature. This review was designed to be comprehensive but not exhaustive, aiming to gather a robust set of design elements that can inspire end-users imaginations. Table 1 lists the design elements incorporated into the design cards for each communication dimension, detailing the descriptions of each element and the literature sources from which they were derived. Based on the literature chosen, the design cards encompassed a variety of explanation techniques, including interpretation methods applied to machine learning model outputs or computational systems, human communication and reasoning strategies, data visualization techniques, and more. As the selection of the design elements is not exhaustive, each dimension included a "Design Your Own!" card where users can freely create their design element. Figure 2 shows an example of a design card. The complete set of the AI-DEC can be found in Figure 5 in the Appendix.

The objective of the AI-DEC is to be utilized in design studies, incorporating end-users into the process of crafting user-centric AI explanations and fostering structured, creative thinking. According to the classification of card-based design tools by Roy and Warren [110], the AI-DEC falls under "human-centered design," where the cards help "designers to focus on the users of a product, service, or system, considering their needs, wishes, and requirements."

Table 1: *AI-DEC Design Elements* — Description of design elements and related works in HCI, XAI, ML, and design that develop, use, or evaluate these elements. In each dimension, the "Design your own" card allows users to create their own design elements.

Design Element	Description	Related Works
	Dimension: Content	
System Metrics	Measures model precision, quantifies feature influence, assesses stability, and analyzes model complexity	[33, 80, 86]
Endorsement	Validates information through peer and expert approval, using reliable sources and established knowledge	[49, 75, 142]
Counterfactual	Shows how altering variables or hypothetical changes impacts model outcomes	[21, 27, 88, 136]
Causation	Identifies and clarifies the causal links or decision-making rules between input features and model outputs	[12, 27, 45, 130, 136]
Subgoal & Breakdown	Dissects complex models into manageable subgoals, individual components, or hierarchical relationships	[2, 32]
Compare & Contrast	Compares similar and differing cases to highlight decision-making patterns, anomalies, and critical factors	[72, 90, 109, 146]
Examples	Uses specific instances and use cases to illustrate model operations	[64, 66, 128]
	Dimension: Modality	
Metaphor	Utilizes familiar concepts to simplify and clarify complex ideas	[34, 50, 67, 96, 104]
Badges	Represents status, achievements, affiliations, skills, or recognition	[68, 149]
Lists	Organizes information into structured, consecutive formats	[40, 65, 76]
Abstraction & Summary	Condenses complex model information into more manageable forms, often excluding intricate details	[26, 69, 108]
Natural Language	Uses everyday language to communicate complex concepts	[2, 19, 73, 142]
Icon & Pictures	Provides visual representations of model elements (*e.g.,* data, results, and concepts)	[99, 105, 150, 151]
Diagrams & Graphs	Visually depicts model structures, decision processes, variable relationships, trends, and patterns	[11, 18, 100]
	Dimension: Frequency	
Once at a Specific Timing	Delivers at strategically chosen moments (*e.g.,* before/after predictions, periodically, triggered by events, during critical decisions or educational needs)	[106, 107]
Progressive Disclosure	Gradually reveals information to manage complexity and cognitive load by strategic needs or user prompts	[118, 119]
On-demand	Provides information responsive to users' requests	[20, 132]
	Dimension: Direction	
One-way	Supports unidirectional information delivery from the source to the recipient, without necessitating any interaction or response	[106, 107]
Two-way	Supports interactive communication that allows recipients to engage, inquire, and participate actively	[15, 101]
Multi-way	Supports a dynamic interactive environment and information exchange not only within the system but also among multiple users	[36, 113]

When using the AI-DEC, end-users choose cards from the four communication dimensions to construct their desired explanation from the AI system. The final explanation design comprises a combination of cards from each dimension, representing a design solution tailored to user information or interaction needs. This design solution encompasses the explanation's content, mode of delivery, timing or frequency of presentation, and necessary interactivity or user intervention. An example of a user's AI explanation design



Figure 2: *Example of AI-DEC Design Card — Left:* The figure depicts a sample design card in the AI-DEC. The card is color-coded to represent its category within the communication dimensions. It includes a comprehensive description of the design element, a visual example, and examples of other use cases. *Right:* The complete AI-DEC set is displayed. For further details, please see Figure 5 in the Appendix.

using the AI-DEC is shown in Figure 4. A printable copy of the full AI-DEC is accessible in the supplementary materials.[1]

4 CO-DESIGNING AI EXPLANATIONS WITH THE AI-DEC

To understand the effectiveness and usability of the AI-DEC, we conducted co-design sessions with end-users of AI systems to create AI explanations tailored to their specific needs and contexts. We chose a co-design approach to demonstrate examples of how the AI-DEC can be used in real-world applications that incorporate AI systems. Furthermore, this approach aligns with the AI-DEC's goal of facilitating collaborative and inclusive design. It actively engages participants in the design process while providing valuable insights into user interactions and perceptions of the AI-DEC, helping us understand user engagement and identify areas for improvement.

4.1 Study Procedure

The co-design study comprised two parts: understanding *what* workers wanted to design for, and observing their use of the AI-DEC to generate design solutions on *how* they wanted AI explanations to be communicated, in order to satisfy their information and interaction needs. Initially, to understand what to design for, participants reviewed the "Question Bank" by Liao et al. [77]. This bank includes a taxonomy of user needs for AI systems, which we used to help workers easily identify their explanation needs. Utilizing the Question Bank, workers identified five explanation categories that represented their most crucial explanation needs, based on their interactions with the AI system in their workplace.

Based on their selection, participants then developed AI explanations using the AI-DEC. Participants freely selected cards from each dimension (*i.e.,* content, modality, frequency, and direction) as many as they wanted, or created their own design element using the "Design your own!" card. Once participants finished their explanation design, they were asked to describe their design solution in detail and envision how it would fit into their workflow. After the design session, participants answered interview questions to evaluate the AI-DEC. The study lasted approximately one hour.

4.2 Participants

We recruited knowledge workers who interact with decision-support AI systems as an integral part of their daily workflow. To understand the applicability and usefulness of the AI-DEC across various contexts, 16 participants were recruited from three different work domains: health (6), finance (5), and management (5). The three domains were chosen as they are among the most utilized areas where decision-support AI systems are applied in real-world workplaces [28, 29]. Participants were recruited through university mailing lists and study flyers posted in workplaces in the United States. A pre-screening survey determined eligibility based on workers' domain, workplace tasks of interacting with a decision-support AI system, and period of interaction experience with a minimum of three months. Each study session took place in person with one participant and one researcher. Participants were provided with a $50 USD payment upon completion of the study. Participants (9 male, 7 female) were aged 21–47 (M = 31.1, SD = 8.2). Demographic data is included in Table 2. In presenting our findings, we denote participant work domain with "H" (H1–H6) for healthcare, "M" (M1–M5) for management, and "F" (F1–F5) for finance.

4.3 Analysis

All user studies were held in person and video recorded using Zoom. Session recordings were then transcribed using the Otter.ai automated speech-to-text generation tool and manually checked by the research team. The transcriptions were coded following the guidelines developed by Clarke and Braun [31] and McDonald et al. [87]. We conducted Thematic Analysis (TA) using the transcriptions and participants' design solutions using the AI-DEC[2]. The first author was familiarized with the data by facilitating the studies and

[1]The supplementary materials can be found at https://osf.io/fqpy2/?view_only=97dcc454d9dd455c94dd03178c984bbe

[2]The initial interview phase was analyzed in a separate study. This paper focuses on the co-design sessions of the research.

Table 2: *Participant Information* — Demographics and background information.

ID	Age	Gender	Race	Job	System Use
			Domain: Health		
H1	34	Female	White	Radiologist	Computer Aided Detection
H2	26	Female	White	Acute Care Physician	Decision Support Healthcare Software
H3	26	Female	White	Emergency Room Physician	Decision Support Healthcare Software
H4	30	Female	White	Primary Care Physician	Decision Support Healthcare Software
H5	27	Male	White	Family Medicine Physician	Decision Support Healthcare Software
H6	46	Male	White	Radiologist	Computer Aided Detection
			Domain: Management		
M1	20	Female	Hispanic	Multinational Chain Crew Member	Automated Scheduling System
M2	23	Male	White	Multinational Chain Crew Member	Automated Scheduling System
M3	30	Male	Hispanic	Human Resources Recruiting Director	Applicant Tracking System
M4	41	Female	White	Marketing Manager	Advertisement Management Software
M5	37	Male	White	Supply Chain Analyst	Automated Scheduling System
			Domain: Finance		
F1	25	Male	Hispanic	Banker	Automated underwriting system
F2	35	Male	White	Loan Officer	Automated underwriting system
F3	47	Male	White	Banker, Branch Manager	Automated underwriting system
F4	31	Male	White	Banker, Branch Manager	Automated underwriting system
F5	21	Female	Black	Banker	Automated underwriting system

initially generating codes. Through iterative discussions with the whole research team, categories of grouped codes were generated and refined. The categories were then further clustered and reflected upon with the entire research team to derive themes emerging from our study data. After all candidate themes were discussed and reviewed, the final themes are reported as findings.

5 FINDINGS

In this section, we present the findings of the co-design session. In Section 5.1, we describe how the workers' explanation needs differed across domains depending on the AI system's role in the workplace and worker values. In Section 5.2, we present the AI explanation designs created by workers using AI-DEC. Depending on the unique information and interaction needs, these designs aimed to support various levels of worker performance and autonomy. Finally, Section 5.3 describes the workers' evaluations of the usefulness, shortcomings, and potential improvements for future use of the AI-DEC.

5.1 Workplace AI Systems and their Role

In this section, we provide an overview of the AI systems participants use in their everyday work settings. This understanding provides essential context for understanding the participants' design motivation and the findings. These workplace-incorporated systems leverage AI techniques, such as machine learning, data analysis, and reasoning algorithms. The primary objective of these systems is to provide decisions or decision-making support to workers. Despite the workplace AI systems having similar functionalities of decision-support, they had different roles and responsibilities in each work domain. Furthermore, the different roles of the AI system led to workers feeling different levels of autonomy. Factors including workplace characteristics, worker's autonomy afforded by the AI system's role, and worker values shaped diverse information and interaction needs among workers.

Healthcare. Workers in the healthcare domain use Clinical Decision Support Systems (CDSS) and AI systems that support the digitization, management, and sharing of patient medical records while streamlining clinical workflows. The system also provides clinical decision-support tools, offering workers alerts, reminders, and evidence-based guidelines for lab tests, treatments, and diagnosis at the point of decision-making. The computer-aided diagnosis (CAD) systems help workers interpret medical images, particularly in detecting and analyzing potential abnormalities, patterns, or lesions, by leveraging rule-based algorithms and ML techniques.

The AI system in the healthcare domain had an assistive role, leaving the final decision-making to workers. Workers also had follow-up tasks using the output of the AI system in addition to making the final decision, which involved explaining their final decision to various stakeholders in the workplace.

Finance. Workers in the finance domain use Automated Underwriting Systems (AUS) to analyze loan applications based on financial information (*e.g.,* credit reports, digital transaction records,

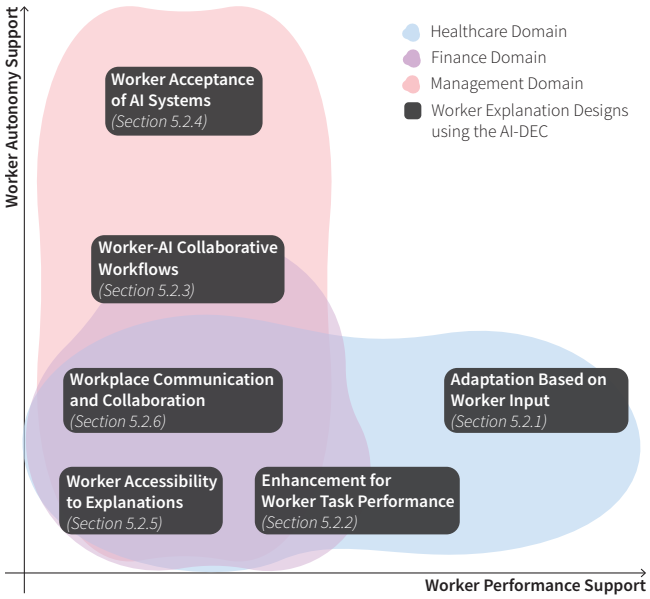


Figure 3: *Findings Summary* — The graph shows different goals of the explanations that workers designed using the AI-DEC. The goals varied in the level of support that workers desired for their autonomy and performance across work domains. Healthcare workers focused explanation designs on enhancing their performance. The finance workers aimed to design explanations that improved both their performance and autonomy. In management, workers prioritized explanations that supported their autonomy.

prior loan history, income, etc.) and generate decisions on approving or denying mortgages, personal loans, small business loans, auto loans, and student loans. During customer interactions, workers gather financial information from the customer, enter it into the system, and receive a definitive decision from the system with no or limited warnings for risks (*e.g.,* "multiple high-risk factors").

The role of the AI system in the finance domain was more decisive than in healthcare, as it held the authority to approve customer applications. This shift in decision-making power led workers to perceive a reduction in autonomy, limiting their ability to apply expertise as they did before the introduction of AI systems. However, workers continued to fulfill essential duties such as explaining the AI system's outcome to customers and advising financial plans.

Management. Workers from multinational chains use an automated system to schedule shifts and positions for various tasks. The AI system automatically generates shift schedules based on factors like performance, skill set, and generated profit. Workers in manufacturing companies use centralized software programs to automatically manage resources and scheduling, including allocating the required labor type, labor time, and tasks to workers while providing information about available resources and task processing. An advertising manager uses a social media management platform that monitors and analyzes social media to recommend the best times for individuals and businesses to promote advertisements and manage multiple accounts. A worker in a Human Resources department uses automation software to manage and streamline the recruitment and hiring process, particularly for candidate screening to analyze resumes and assess qualifications.

The AI systems in the management domain played the most authoritative role compared to the healthcare and finance domain, as it made the final decision for workers regarding decision outcomes, worker tasks, and worker schedules. Workers had minimal or no opportunities to independently make decisions or apply their expertise, resulting in a low level of autonomy provided to workers.

5.2 Workers' AI Explanation Designs with the AI-DEC

In this section, we report workers' explanation designs using the AI-DEC. From our analysis, we identified six themes in the workers' explanation designs, which diverged across different levels of supporting worker performance and autonomy. In this work, we refer to performance as the effectiveness, efficiency, and quality of work carried out by workers in their domains, while autonomy refers to the degree of independence, self-governance, and decision-making authority granted to workers within their roles.

Using these definitions, we identified six explanation themes from workers' designs aimed at enhancing performance, autonomy, or both. The explanation design themes include: (1) adaptation based on worker input; (2) enhancement for worker task performance; (3) worker-AI collaborative workflows; (4) worker acceptance of AI systems; (5) worker accessibility to explanations; and (6) workplace communication and collaboration. A summary of the findings is shown in Figure 3. The complete set of workers' explanation designs can be found in Figure 6 in the Appendix.

5.2.1 Adaptation Based on Worker Input. In the healthcare domain, workers developed AI explanations that allowed for worker input in response to the AI system's output to support worker performance.

Input to Improve Performance of AI System. Using the AI-DEC, healthcare workers (H1, H3, H5, H6) designed AI explanations to improve the system's decision-making performance by incorporating their feedback, particularly regarding error corrections and adjustments to decision-making features. For example, one worker (H3) described *"So the goal here is to prevent me from mistakes. So with the output, there needs to be reasoning for my sensemaking to determine if it [AI system] is correct, we use graphs and metrics, but I also want a breakdown...[selecting cards] But most importantly there needs to be iterative feedback, so I want two-way (direction), and I want a list of what was fixed."* They emphasized that the AI system's output accuracy was directly linked to their own task performance, leading to select design features that enabled continuous improvement based on error correction from worker feedback.

Additionally, workers (H5, H6) sought validation of whether the AI system adhered to values and practices similar to those of human workers in decision-making processes. For example, one worker (H5) used the AI-DEC to devise explanations that conveyed system limitations and risks associated with error correction and worker awareness. The explanation design comprised causation and limitations (content), diagrams and badges (modality), occurring

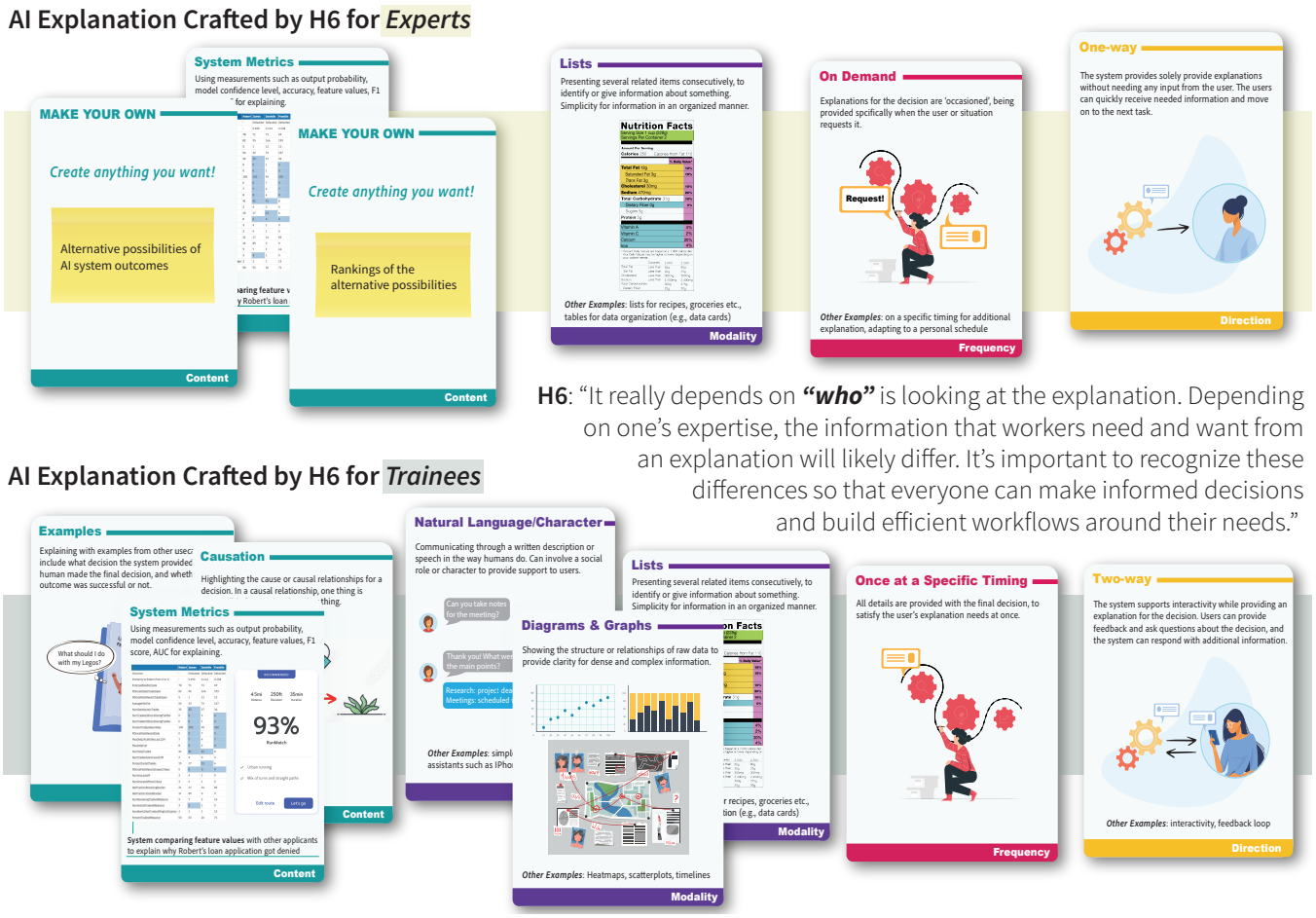


Figure 4: *Worker's AI Explanation Design Using the AI-DEC* — The figure shows an AI explanation crafted by a worker (H6). The worker developed two distinct types of explanations for the worker's varying levels of expertise. *Top:* for experts, the AI explanation functioned as an extra eye. *Bottom:* for trainees, it served as an educational tool, providing more detailed information. Details can be found in Section 5.2.1.

whenever necessary (frequency), and supporting two-way communication for interactive feedback (direction). In response to the limitations outlined in the explanations, workers envisioned correcting misaligned features to mitigate potential adverse outcomes and increase workers' awareness of risks associated with the AI system's outputs.

Input for Expertise-based Explanations. Workers (H1, H3, H6) also used the AI-DEC to design explanations that provided different information support based on the worker's level of expertise. For example, H6 explained *"My design needed to involve multiple versions, because even within my team, explanation needs and uses are different depending on one's expertise. It really depends on "who" is looking at the explanation."* For experts, the explanation aimed to serve as an extra eye, whereas for trainees, it functioned as an educational tool, offering more detailed information. Consequently, one worker (H6) developed two types of explanations using the AI-DEC, when explaining the rationale behind the AI system's outputs. For experts, the worker selected alternative possibilities, metrics, and rankings (content), presented in a list format (modality), on-demand (frequency), and provided one-way communication (direction). Designs focused on ruling out other possibilities and providing concise support for accuracy. Conversely, for novices, the worker selected causation, metrics, and examples (content), presented through lists, diagrams, and natural language (modality), delivered at specific timings at the end of each AI system output (frequency), and facilitating two-way communication for further questions (direction.) The design focused on supporting detailed investigation on the factors and procedures of the final decision. The importance of tailored explanations based on the worker's expertise level was emphasized to ensure that all workers could learn and make informed decisions.

5.2.2 Enhancement for Worker Task Performance. Workers in the healthcare and finance domains designed AI explanations to improve worker performance by helping them plan and execute follow-up tasks using the output of the AI system.

Stakeholder Explanations. Workers (H2, H3, H4, F2, F3, F4) used the AI-DEC to create explanations that helped them clearly explain the final decision to various stakeholders in the workplace. The design elements varied depending on the intended audience, specifically whether explaining to domain experts like co-workers or to domain novices such as customers and patients. For example, one worker (H3) used the AI-DEC to create two distinct versions of explanations depending on whether the audience was domain experts or novices. For domain experts, more technical language involving metrics, counterfactuals, and examples (content) was selected to be presented through intuitive lists and diagrams (modality), be available on demand (frequency), and be presented one-way (direction) as a "report" of the worker's decision-making procedure. Such explanations aimed to convey factual details that support the worker's final decision and factors of alternative options regarding "why not," for transparent communication and preventing misunderstandings between workers.

Conversely, for novices, designs were aimed to easily support high-level understanding. The design entailed causation, breakdown, and comparison (content) for abstracted but informative descriptions, conveyed via metaphors and diagrams (modality) for easier understanding, to be available on demand (frequency) in need of support, and in one-way communication (direction). Explanations for novices clarified outcome rationales while simplifying complex details with interpretable design elements.

Action Planning. Additionally, the AI-DEC was used to devise explanations to support workers (H1–H4, H6, F1–F5) in strategizing follow-up actions based on the AI system's outputs. They noted that a crucial aspect of their job often involved determining the next steps using the AI system's output to accomplish workplace tasks or objectives (*e.g.,* gaining customer approval or planning patient treatment). Thus, workers crafted AI explanations that offer practical guidance, including suggestions on actions required to achieve alternative outcomes. For example, one worker (H1) designed *"recommendations for the "safest" next steps based on prior experience or probability,"* while another worker (F2) designed *"test functions" to experiment with different features for varied outcomes."* This worker's (F2) explanation design enabled workers to experiment with alternative inputs to observe the impact on outcomes. Selected design cards included causation and counterfactuals (content), lists (modality), available on demand (frequency), and facilitating interactive two-way communication (direction). The design rationale was explained as, *F2: "So testing with counterfactuals enables me to ask "What if I made this change? Can I increase my loan amount? Would I still be in the parameters for the inspection waiver?" Along with compare and contrast, show me the difference. This gives me the full picture I need to move forward."*

Another worker in the healthcare domain designed explanations to clearly understand all possible outcomes, aiding decision-making in their high-stakes environment. They mentioned, *H6: "Knowing the possible outcomes, rankings or probabilities becomes much more useful in my workplace where mistakes can be bad. So I selected a list of possible outcomes. Can you give me those as a ranking, or probabilities of each one, and show me the distributions of those where the patient falls? To help me be more sure."* Accordingly, they used the "design your own!" card to include a list of possible outcomes ranked by probability.

5.2.3 Worker-AI Collaborative Workflows. Workers in the finance and management domains designed AI explanations to improve worker autonomy when engaging with workplace AI systems. Designed explanations focused on promoting cooperative opportunities between workers and the AI system.

Opportunities for Worker Expertise. Workers (F2, F3, F5, M2–M5) used the AI-DEC to design explanations that fostered collaboration between AI systems and workers, aiming to augment worker autonomy by providing chances for workers to exercise their expertise. One worker (M1) commented, *"There are many nuances and contextual factors to consider in daily jobs, where human experience is important. So I chose not to let the AI system handle everything [in my design,] and included a human touch, whether it's [AI explanation] asking what to do or how it did, or actually sharing some responsibilities."* Thus workers selected design cards to support worker understanding and provide opportunities for worker intervention, specifically envisioned for where the AI explanation sought worker expertise assistance when encountering unforeseen or abnormal situations (*e.g.,* bankruptcies, natural disasters, and construction projects). One worker (M2) designed explanations to prompt worker intervention when particularly the AI system faced uncertainty. This explanation design encompassed causation and breakdown (content), summary and list (modality), when encountering uncertainty (frequency), and enabled interactive two-way communication (direction). This explanation aimed to provide information on the tasks completed by the AI system and the obstacles encountered, offering context for workers to intervene and make informed decisions.

5.2.4 Worker Acceptance of AI Systems. Workers in the management domain designed explanations to support worker acceptance, aiming to address the lack of autonomy provided by the current workplace AI systems. Such acceptance was an important factor in preventing workers from being reluctant to adhere to the managerial roles and task allocations performed by the AI system.

Support for Onboarding of AI Systems. Using the AI-DEC, workers (M1, M2, M3, M5) crafted explanations that varied in design throughout the deployment timeline. One worker noted, *M3: "Explanations need to change depending on how accustomed we are to the system, and how much we know or trust it. Otherwise, it might be too little or too much information."* Workers expressed their desire for the design of explanations to evolve based on their growing familiarity and rapport with the system. For example, one worker (M3) chose "progressive disclosure (direction)" to be applied not in a single explanation, but across different stages of the AI system's deployment. They designed different explanation versions for both the introduction and active-use stages. During the initial onboarding process, the explanation was designed to focus on establishing rapport and fostering workers' understanding of the AI system's capabilities and value. Thus, explanation designs encompassed causation of the AI system's output and examples demonstrating the AI system's capabilities in similar use cases along with counterfactuals it considered to show its capabilities (content), conveyed

through both summary and extensive lists for detailed understanding (modality), enabling two-way interactive communication to address worker questions for concerns or confusions (direction) any time required on demand (frequency). However, during the active use stage, the explanation design entailed more simplified design elements, involving metrics and causation (content), and just summaries or graphs for intuitive understanding (modality), but also supporting user information needs interactively (direction) when needed (frequency). Through different designs, the worker envisioned the explanation to gradually unveil less detailed information over time during the deployment, as rapport was built through the introductory explanations.

5.2.5 Worker Accessibility to Explanations. In all domains, workers created explanations that included design elements that were readily understandable to laypeople. Workers emphasized the significance of accessibility in explanation design to help workers comprehend and effectively apply it to their workplace tasks.

Accessible Design Elements. Workers (H1, H3–H6, F2–F5, M1, M2, M4) selected design cards that focused on easing the understanding and improving the relatedness to workers' daily work experiences to enhance their comprehension and usability of the AI explanation. For example, one worker in the finance domain (F2) described the need for such explanations as *"My job is to educate borrowers and advise them on how to move forward, but I can't do that if the system doesn't give me meaningful stuff. That's why my explanation gives me specific, financial information that is useful to me, measures that I use every day to do my job."* This worker's explanation design involved combining the metrics (content) and list (modality) card to include the Debt to Income (DTI) ratio and credit scores, with a list of saving strategies and other loan options.

Another worker (M2) from the management domain designed their explanations to reflect the importance of community among workers, emphasizing firsthand experiences and information from colleagues. The design featured interaction examples from employees at various chain locations and used metaphors to articulate the perceived benefits and implications of integrating the AI system. The explanation provided endorsements through examples from others using the AI system under similar circumstances (content), utilize metaphors for relativeness (modality), be available on demand (frequency), through one-way communication (direction).

One worker (H1) in the healthcare domain designed explanations that aligned with their unique, image-focused workflow practices to help their understanding. They utilized the "design your own!" card for color-coding to represent the AI system's uncertainty level, due to its intuitiveness and faster comprehension compared to text. Their explanation design included metrics, examples, and comparisons (content), diagrams and color-coding to support effective worker performance (modality), available on demand (frequency), and facilitated two-way interactive communication (direction).

5.2.6 Workplace Communication and Collaboration. Workers across all domains developed explanations to facilitate workplace communication and collaboration. Workers emphasized the importance and everyday occurrence of collaborative and communicative interactions in the workplace, as workers regularly cooperated through different responsibilities to achieve shared workspace goals. Moreover, workers also socially interacted and communicated with each other daily to discuss workers' final decisions, offer mentorship, gain insights, or validation.

Support for Workplace Collaboration. Workers (H3–H5, F3, F4, M3, M5) designed explanations to enhance collaboration and achieve workplace goals. These explanations focused on monitoring progress, identifying bottlenecks in the pipeline, and adjusting task plans accordingly. For instance, one worker (M5) selected cards that described others' progress (content), included icons within the work chain and summaries of issues (modality), were provided at specific times, such as when obstacles occurred (frequency), and supported multi-way communication (direction). These shared explanations aimed to streamline collaboration and offer guidelines for addressing conflicts. H2 elaborated on such explanation needs as *"Currently, finding where a bottleneck might be is difficult because we don't know when, where, why, and who. It's unclear if someone ignored information from the system, if it malfunctioned, or if another factor was at play. However, if the AI explanation could pinpoint who took what actions and identify any concerns or issues they had, we can act and address them. This teamwork is crucial because it impacts the outcome on patients."*

Additionally, to further enhance collaboration, healthcare workers (H1, H3, H5, H6) incorporated elements into the explanations that promoted reliability and consensus towards the AI system. Specifically, workers opted for endorsements (content) and badges (modality), noting that societal consensus was crucial in fostering trust in their workplace. Such design elements ensured that all team members could trust and effectively use the system, thereby reinforcing collaborative efforts and improving operational efficiency.

Support for Workplace Communication. Workers (H1, H3–H5, F1, F5, M1, M2, M4) designed AI explanations to facilitate communication among workers, enhancing social learning and mentoring. These explanations incorporated results from others' interactions and diverse opinions to validate decisions and build trust. Specifically, one worker (F5), who had recently started their position, designed explanations to incorporate information from past interactions between other workers and the AI system, particularly from higher in the hierarchy and more experienced workers in the workplace hierarchy. They explained their design purpose as *F5: "It is often uncomfortable to ask my manager every time I want to check something, at the same time I am concerned if I am doing anything wrong, especially if I am learning. So I want the explanation to do it for me, help me learn, and validate as I go so I can develop my skills."* The explanation design included counterfactual options that have been used by others in cases with similar circumstances, a breakdown of how to reach the goal, and examples of decisions made by higher hierarchy workers (content), conveyed through natural language (modality), provided on demand (frequency), and enabling multi-way communication (direction). Workers described that explanations supporting social communication helped facilitate continuous learning and maintain social connections, which were key to building a supportive and innovative work environment.

5.3 Worker Feedback of the AI-DEC

In this section, we present the workers' assessment of the AI-DEC gathered at the end of the co-design session.

5.3.1 Advantages of the AI-DEC for User-centered AI Design. Workers (H1–H3, H6, M1–M5, F1, F2, F4, F5) noted that AI-DEC effectively translated abstract preferences into practical design solutions. Its four-dimension structure provided constructive guidance, allowing for a systematic design process that addressed a wide range of information needs. One worker explained this perspective as *H3: "The four categories guided my thinking step-by-step, making me analyze details that I might have overlooked had I simply been asked about my preferences for explanations."*

Additionally, the tangible aspects of prototyping the AI explanations were appreciated by workers (H2–H6, M1–M3, M5, F1, F2, F4, F5). Constructing, deconstructing, and restructuring communication strategies facilitated iterative reflection, testing, and adjustment, leading to designs that were practically sufficient and tailored to their unique workplace environments.

Finally, workers (H1–H3, H6, M2–M5, F2, F5) described the card deck format of AI-DEC as particularly beneficial for its modularity, which supported quick iteration and customization based on feedback and changing requirements. This flexibility not only enhanced the relevance and effectiveness of AI interactions across diverse settings but also encouraged workers to explore creative combinations of design elements, fostering innovative solutions that might not have been possible in more rigid formats.

5.3.2 Shortcomings and Improvements for the AI-DEC. Workers (H3, H6, M3, M4) indicated that using the AI-DEC to design worker-centered explanations for unfamiliar AI systems might be difficult due to a lack of understanding of the AI system's capabilities. Those in our study, familiar with their current workplace AI systems, adeptly tailored solutions to meet their needs. However, one worker (M3) noted the potential difficulty in crafting explanations for a new AI system without understanding its capabilities or how it integrates into existing workflows. Workers (H3, H6, M3) suggested that understanding the AI system's capabilities and limitations could be facilitated by providing users with brief lists or summaries of these details before initiating the design process.

Additionally, workers (H5, M5) observed that certain design descriptions, such as "abstraction" and "examples," were excessively wordy and lacked intuitiveness. They proposed utilizing both sides of each card: one side to display additional examples and visual designs for better comprehension and application, and the opposite side to contain detailed descriptions of the design element, offering comprehensive insights. Similarly, other workers (F3, M1) stressed the need to include more examples on each card, showcasing both general and specific use cases relevant to AI systems.

For future use, workers (H2, H4, M1, M3, F1) envisioned employing AI-DEC as a tool to collaborate with AI designers and engineers during the design phase of workplace AI systems. They noted their frequent exclusion from the pre-deployment process, despite being the primary users over time. Workers described that the AI-DEC can serve dual purposes through prototyping and sketching tailored AI explanations, as *H5: "the AI-DEC can act as a middleman to connect expectations and needs with us and the people who develop the AI, it could be a win-win for everyone."* Workers further elaborated that on the one hand, the AI-DEC can help workers clarify their requirements and desires through hands-on prototyping. On the other hand, it can provide engineers and designers with a clear understanding of user information and interaction needs, fostering effective discussions about the possibilities and limitations of the AI system.

6 DISCUSSION

We introduce the AI-DEC, a card-based design method developed to engage end-users and address their needs in the AI explanation design process. Grounded in communication theory encompassing the structure of content, modality, frequency, and direction, the AI-DEC facilitates the development of user-centered communication strategies for AI explanations. Through co-design sessions, we evaluated the effectiveness of the AI-DEC. Our findings suggest that the AI-DEC effectively supported workers in crafting AI explanations suitable for different levels of performance and autonomy needs. These needs varied depending on the AI system's role in their workplace and the values of the workers. Drawing insights from workers' designs and feedback, we present implications for utilizing the AI-DEC to create user-centered AI experiences.

6.1 Leveraging the AI-DEC with End-users

When to use the AI-DEC. Existing research has shown that XAI techniques can often fall short for end-users requiring domain-specific insights [5, 39, 147], leading to investigations into the contextual dynamics of deployment such as the AI lifecycle [34, 77, 79] and user informational or situational needs [8, 37, 38]. A user-centered approach, prioritizing an understanding of users and their environmental characteristics, has been identified as crucial for effective AI system design [39, 153]. Given these insights, we see an opportunity to leverage the AI-DEC to enhance end-user involvement, particularly in scenarios that require effective communication and collaboration with users.

Firstly, the AI-DEC facilitates communication with users about necessary design features for AI explanations, enabling the identification of specific environmental and individual needs. These discussions highlight the unique aspects of the environment, the specific tasks for which the AI system will be used, and the explanation or interaction needs that support users' real-world objectives. For instance, in our study, workers realized that they value the inclusion of "why not" factors in explanations to colleagues and sought similar features in their AI explanations. To facilitate these discussions, external design methods or frameworks such as employing a taxonomy of user needs [9, 77], storyboarding [124] use cases, or task-based discussions [70] can be utilized along with the AI-DEC.

The AI-DEC can also act as an effective collaboration tool during various design stages for AI systems, involving a range of stakeholders. In our study, workers envisioned AI-DEC as a probe to enhance communication and collaboration between AI designers and engineers. These probes were envisioned to create a common understanding of user needs, capabilities, and features, as well as the explanations that the AI system should provide. Physical prototypes produced with the AI-DEC were seen to be particularly useful

in identifying misunderstandings and encouraging stakeholders to seek consensus through iterative redesigns. For instance, the "content" cards were noted by workers to effectively outline the AI system's technical capabilities, helping them envision optimal ways to use the AI systems to meet their needs. This collaborative approach not only uncovers new design possibilities but also helps avert the risk of deployment failures of AI explanations by integrating user perspectives from the outset.

How to use the AI-DEC. Building on our findings, we recognize the AI-DEC's potential to shape not only individual explanations but also the broader interaction experience between users and AI systems. Although our study required workers to design explanations for specific needs (*i.e.,* categories from the Question Bank [77]), we observed them expanding their designs to create distinct versions for different interaction points. For instance, workers crafted explanations that varied in levels of worker involvement based on the uncertainty detected in the AI system's decision-making. Other explanations tailored the content and modality to match the user's expertise, revealing different levels of detail at various stages to support decision-making. Furthermore, workers considered how explanations should adapt across the AI system's deployment phases (*e.g.,* introductory versus active use). Thus, the AI-DEC can be used to help end-users holistically ideate the interaction experience, defining the role, responsibilities, and communication strategies that the AI system should embody.

Designers can also extend the design elements of AI-DEC to suit workplace requirements or particular preferences. While retaining the established four-dimension framework of content, modality, frequency, and direction, the elements in the cards can be adjusted to include various aspects that designers find valuable or wish to experiment with in relation to their AI projects. This modification enables designers to tailor the AI-DEC cards according to the needs of users, specific deployment domains, problem scenarios, or desired design features. Such adaptability of the AI-DEC also helps address the limitations that can exist with design cards, where the content or elements of the cards lack relatedness to application contexts [54]. Nonetheless, expanding the design cards might introduce trade-offs, such as an increased cognitive load on users when generating explanations. In these instances, designers should balance flexibility and complexity. For example, as complexity increases, designers may simplify the design objectives where feasible and offer guidance and tools [*e.g.,* 77, 134] to aid users in navigating the designs effectively.

6.2 Application of the AI-DEC

Previous work highlights the challenges of seamlessly integrating AI systems into real-world applications, particularly due to difficulties involving end-users in AI projects and understanding the diverse needs and expectations of various stakeholders [22, 102, 137]. We aim to address these difficulties by using the AI-DEC to involve end-users in the design process of prospective AI systems. In this study, we deployed the AI-DEC in workplaces such as healthcare, finance, and management. While the AI-DEC has proven valuable in these professional settings, we believe its potential extends further, potentially aiding the design of AI explanations in a broader range of applications. This is particularly relevant given the increasing research on deploying AI systems in community-level and non-workplace environments [133]. For example, consider its application in smart cities [125], where AI systems manage traffic, distribute resources, and ensure public safety [126]. In such complex systems, stakeholders—from city planners to local residents—have varied informational needs and concerns. City officials might need detailed explanations of traffic flow predictions to plan infrastructure changes, whereas residents could benefit from simpler, more direct explanations about how traffic adjustments affect their daily commutes. The AI-DEC can empower teams to design these varied types of explanations effectively. It guides designers in identifying which aspects of the outputs from the AI system need emphasis and how to communicate them clearly to different stakeholders. Using the cards, designers can determine the optimal level of technical detail for different audiences, ensuring explanations are accessible yet informative. This tailored approach not only enhances transparency but also fosters trust and engagement, making the integration of AI systems more acceptable and supported by the public.

Additionally, the AI-DEC can be utilized to generate knowledge in user experience (UX) and design research. In UX research, designers can conduct surveys and user studies—either in-person or virtual—utilizing AI-DEC to explore users' information needs and encourage them to construct ideal explanations and interaction examples. For virtual studies, a digital version of AI-DEC can allow users to prototype on an online storyboard (*e.g.,* Miro [1]). The examples generated, encompassing AI explanations and interactions, can then inform the designs of the final product.

The importance of design research in HCI is increasingly recognized, driven by methodologies such as research through design [152] and design-focused UX research [129]. In this context, we believe that the AI-DEC is valuable in generating new knowledge about AI design requirements and in creating artifacts that represent solutions to these challenges. Given the often variable and unclear design requirements or objectives for AI explanations in different deployment settings, the AI-DEC serves as a useful tool in aiding designers to elicit diverse perspectives on human-AI interaction issues, ideate on potential solutions, and engage in reflective iteration to refine these issues and solutions. Importantly, the artifacts produced during these design activities serve as exemplars, offering insights into how future AI explanations and systems should be designed for successful integration. These insights not only enrich our current understanding but also pave the way for future research directions in human-centered AI design.

7 LIMITATIONS & FUTURE WORK

Our study has several limitations to consider. Firstly, the AI-DEC's scope and design features are not exhaustive and require further validation through comparative evaluations to become a standardized design method. Future work can involve collaboration with design experts to verify and compare AI-DEC against other design tools or methodologies. Additionally, recruiting a larger number of participants in each domain proved challenging, partly due to the busy schedules of experts and the limited availability of workplaces experienced in using AI systems extensively in their daily operations. Moreover, our work involved three domains utilizing AI

systems in real-world applications, excluding other domains such as education and government. Future work should conduct studies with larger pools of participants (*e.g.*, both end-users and designers) and a more diverse number of domains to evaluate the efficacy of utilizing the AI-DEC with HCI research methods and the design process of AI systems. Finally, the qualitative and exploratory nature of our research restricts the generalizability of our findings, as they may not apply universally to all user requirements and AI system application contexts.

8 CONCLUSION

In this work, we present the AI-DEC, a card-sorting-based design method developed specifically to meet the informational and organizational needs of end-users for effective AI explanations. The AI-DEC directly incorporates end-users in the design process of AI explanations, allowing them to design AI explanations that fit their personalized needs. We evaluated the AI-DEC through a co-design session with 16 workers from three different domains—healthcare, finance, and management—to understand its efficacy and applicability in designing AI explanations that better support user needs. Our findings indicate that the AI-DEC helped workers develop explanation designs that supported varying levels of worker performance and autonomy, which differed depending on the AI system's role and worker values. Finally, we present implications for designers using the AI-DEC to build user-centered AI explanations and human-centered AI systems.

ACKNOWLEDGMENTS

We thank the anonymous reviewers for their constructive feedback. This research was supported by the National Science Foundation through awards 1925043, IIS-1925043, IIS-1939606, DGE-2125858, CCF-2217721; Good Systems, a UT Austin Grand Challenge for developing responsible AI technologies;[3] and Swedish Research Council. Any opinions, findings, conclusions, or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

REFERENCES

[1] 2011. Miro. https://miro.com. Accessed: 2024-04-29.
[2] Michael Ahn, Anthony Brohan, Noah Brown, Yevgen Chebotar, Omar Cortes, Byron David, Chelsea Finn, Chuyuan Fu, Keerthana Gopalakrishnan, Karol Hausman, et al. 2022. Do as i can, not as i say: Grounding language in robotic affordances. *arXiv preprint arXiv:2204.01691* (2022).
[3] Michael J Ahn and Yu-Che Chen. 2020. Artificial intelligence in government: potentials, challenges, and the future. In *The 21st annual international conference on digital government research*. 243–252.
[4] Ashraf Alam. 2023. Harnessing the Power of AI to Create Intelligent Tutoring Systems for Enhanced Classroom Experience and Improved Learning Outcomes. In *Intelligent Communication Technologies and Virtual Mobile Networks*. Springer, 571–591.
[5] Ahmed Alqaraawi, Martin Schuessler, Philipp Weiß, Enrico Costanza, and Nadia Berthouze. 2020. Evaluating saliency map explanations for convolutional neural networks: a user study. In *Proceedings of the 25th international conference on intelligent user interfaces*. 275–285.
[6] Renate Andersen, Anders I Mørch, and Kristina Torine Litherland. 2022. Collaborative learning with block-based programming: investigating human-centered artificial intelligence in education. *Behaviour & Information Technology* 41, 9 (2022), 1830–1847.
[7] Á Aranda-Muñoz, Ulrika Florin, Yuji Yamamoto, Yvonne Eriksson, and Kristian Sandström. 2022. Co-Designing with AI in Sight. *Proceedings of the Design Society* 2 (2022), 101–110.
[8] Alejandro Barredo Arrieta, Natalia Díaz-Rodríguez, Javier Del Ser, Adrien Bennetot, Siham Tabik, Alberto Barbado, Salvador García, Sergio Gil-López, Daniel Molina, Richard Benjamins, et al. 2020. Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. *Information fusion* 58 (2020), 82–115.
[9] Vijay Arya, Rachel KE Bellamy, Pin-Yu Chen, Amit Dhurandhar, Michael Hind, Samuel C Hoffman, Stephanie Houde, Q Vera Liao, Ronny Luss, Aleksandra Mojsilović, et al. 2019. One explanation does not fit all: A toolkit and taxonomy of ai explainability techniques. *arXiv preprint arXiv:1909.03012* (2019).
[10] Amid Ayobi, Katarzyna Stawarz, Dmitri Katz, Paul Marshall, Taku Yamagata, Raul Santos-Rodriguez, Peter Flach, and Aisling Ann O'Kane. 2021. Co-designing personal health? Multidisciplinary benefits and challenges in informing diabetes self-care technologies. *Proceedings of the ACM on Human-Computer Interaction* 5, CSCW2 (2021), 1–26.
[11] Andrew J Barnes, Yaniv Hanoch, Talya Miron-Shatz, and Elissa M Ozanne. 2016. Tailoring risk communication to improve comprehension: Do patient preferences help or hurt? *Health Psychology* 35, 9 (2016), 1007.
[12] Sander Beckers. 2022. Causal explanations and xai. In *Conference on Causal Learning and Reasoning*. PMLR, 90–109.
[13] Tilde Bekker and Alissa N Antle. 2011. Developmentally situated design (DSD) making theoretical knowledge accessible to designers of children's technology. In *Proceedings of the SIGCHI conference on human factors in computing systems*. 2531–2540.
[14] Jesse Josua Benjamin, Christoph Kinkeldey, Claudia Müller-Birn, Tim Korjakow, and Eva-Maria Herbst. 2022. Explanation Strategies as an Empirical-Analytical Lens for Socio-Technical Contextualization of Machine Learning Interpretability. *Proceedings of the ACM on Human-Computer Interaction* 6, GROUP (2022), 1–25.
[15] Astrid Bertrand, Tiphaine Viard, Rafik Belloum, James R Eagan, and Winston Maxwell. 2023. On Selective, Mutable and Dialogic XAI: a Review of What Users Say about Different Types of Interactive Explanations. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. 1–21.
[16] Umang Bhatt, Javier Antorán, Yunfeng Zhang, Q Vera Liao, Prasanna Sattigeri, Riccardo Fogliato, Gabrielle Melançon, Ranganath Krishnan, Jason Stanley, Omesh Tickoo, et al. 2021. Uncertainty as a form of transparency: Measuring, communicating, and using uncertainty. In *Proceedings of the 2021 AAAI/ACM Conference on AI, Ethics, and Society*. 401–413.
[17] Oloff C Biermann, Ning F Ma, and Dongwook Yoon. 2022. From tool to companion: Storywriters want AI writers to respect their personal values and writing strategies. In *Designing Interactive Systems Conference*. 1209–1227.
[18] Sandra M Brown, Julie O Culver, Kathryn E Osann, Deborah J MacDonald, Sharon Sand, Andrea A Thornton, Marcia Grant, Deborah J Bowen, Kelly A Metcalfe, Harry B Burke, et al. 2011. Health literacy, numeracy, and interpretation of graphical breast cancer risk estimates. *Patient education and counseling* 83, 1 (2011), 92–98.
[19] Tom Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared D Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, et al. 2020. Language models are few-shot learners. *Advances in neural information processing systems* 33 (2020), 1877–1901.
[20] Adrian Bussone, Simone Stumpf, and Dympna O'Sullivan. 2015. The role of explanations on trust and reliance in clinical decision support systems. In *2015 international conference on healthcare informatics*. IEEE, 160–169.
[21] Ruth MJ Byrne. 2019. Counterfactuals in Explainable Artificial Intelligence (XAI): Evidence from Human Reasoning.. In *IJCAI*. 6276–6282.
[22] Averill Campion, Mila-Gasco Hernandez, Slava Mikhaylov Jankin, and Marc Esteve. 2020. Managing artificial intelligence deployment in the public sector. *Computer* 53, 10 (2020), 28–37.
[23] Tara Capel and Margot Brereton. 2023. What is Human-Centered about Human-Centered AI? A Map of the Research Landscape. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. 1–23.
[24] Gabriela Carneiro, Gil Barros, and Carlos Zibel Costa. 2012. ilo Cards: A tool to support the design of interactive artifacts. (2012).
[25] Chen-Fu Chien, Stéphane Dauzère-Pérès, Woonghee Tim Huh, Young Jae Jang, and James R Morrison. 2020. Artificial intelligence in manufacturing and logistics systems: algorithms, applications, and case studies. , 2730–2731 pages.
[26] Jihye Choi, Jayaram Raghuram, Ryan Feng, Jiefeng Chen, Somesh Jha, and Atul Prakash. 2023. Concept-based explanations for out-of-distribution detectors. In *International Conference on Machine Learning*. PMLR, 5817–5837.
[27] Yu-Liang Chou, Catarina Moreira, Peter Bruza, Chun Ouyang, and Joaquim Jorge. 2022. Counterfactuals and causability in explainable artificial intelligence: Theory, algorithms, and applications. *Information Fusion* 81 (2022), 59–83.
[28] Michael Chui. 2017. Artificial intelligence the next digital frontier. *McKinsey and Company Global Institute* 47, 3.6 (2017).
[29] Michael Chui, James Manyika, Mehdi Miremadi, Nicolaus Henke, Rita Chung, Pieter Nel, and Sankalp Malhotra. 2018. Notes from the AI frontier: Insights from hundreds of use cases. *McKinsey Global Institute* 2 (2018).
[30] Douglas Cirqueira, Dietmar Nedbal, Markus Helfert, and Marija Bezbradica. 2020. Scenario-Based Requirements Elicitation for User-Centric Explainable AI: A Case in Fraud Detection. In *International cross-domain conference for machine learning and knowledge extraction*. Springer, 321–341.

[3] https://goodsystems.utexas.edu

[31] Victoria Clarke and Virginia Braun. 2014. Thematic analysis. In *Encyclopedia of critical psychology*. Springer, 1947–1952.

[32] Devleena Das, Been Kim, and Sonia Chernova. 2023. Subgoal-based explanations for unreliable intelligent decision support systems. In *Proceedings of the 28th International Conference on Intelligent User Interfaces*. 240–250.

[33] Anupam Datta, Shayak Sen, and Yair Zick. 2016. Algorithmic transparency via quantitative input influence: Theory and experiments with learning systems. In *2016 IEEE symposium on security and privacy (SP)*. IEEE, 598–617.

[34] Shipi Dhanorkar, Christine T Wolf, Kun Qian, Anbang Xu, Lucian Popa, and Yunyao Li. 2021. Who needs to know what, when?: Broadening the Explainable AI (XAI) Design Space by Looking at Explanations Across the AI Lifecycle. In *Proceedings of the 2021 ACM Designing Interactive Systems Conference*. 1591–1602.

[35] Najmeddine Dhieb, Hakim Ghazzai, Hichem Besbes, and Yehia Massoud. 2020. A secure ai-driven architecture for automated insurance systems: Fraud detection and risk measurement. *IEEE Access* 8 (2020), 58546–58558.

[36] Upol Ehsan, Q Vera Liao, Michael Muller, Mark O Riedl, and Justin D Weisz. 2021. Expanding explainability: Towards social transparency in ai systems. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. 1–19.

[37] Upol Ehsan, Samir Passi, Q Vera Liao, Larry Chan, I Lee, Michael Muller, Mark O Riedl, et al. 2021. The who in explainable ai: How ai background shapes perceptions of ai explanations. *arXiv preprint arXiv:2107.13509* (2021).

[38] Upol Ehsan and Mark O Riedl. 2020. 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*. Springer, 449–466.

[39] Upol Ehsan, Philipp Wintersberger, Elizabeth A Watkins, Carina Manger, Gonzalo Ramos, Justin D Weisz, Hal Daumé Iii, Andreas Riener, and Mark O Riedl. 2023. Human-Centered Explainable AI (HCXAI): Coming of Age. In *Extended Abstracts of the 2023 CHI Conference on Human Factors in Computing Systems*. 1–7.

[40] Pardis Emami-Naeini, Janarth Dheenadhayalan, Yuvraj Agarwal, and Lorrie Faith Cranor. 2021. An informative security and privacy "nutrition" label for internet of things devices. *IEEE Security & Privacy* 20, 2 (2021), 31–39.

[41] Richard V Farace, Peter R Monge, and Hamish M Russell. 1977. *Communicating and organizing*. Vol. 1077. Addison-Wesley Reading, MA.

[42] Jules Françoise, Baptiste Caramiaux, and Téo Sanchez. 2021. Marcelle: composing interactive machine learning workflows and interfaces. In *The 34th Annual ACM Symposium on User Interface Software and Technology*. 39–53.

[43] Batya Friedman and David Hendry. 2012. The envisioning cards: a toolkit for catalyzing humanistic and technical imaginations. In *Proceedings of the SIGCHI conference on human factors in computing systems*. 1145–1148.

[44] Andrew Garbett, Ziedune Degutyte, James Hodge, and Arlene Astell. 2021. Towards understanding people's experiences of ai computer vision fitness instructor apps. In *Designing interactive systems conference 2021*. 1619–1637.

[45] Amirata Ghorbani, James Wexler, James Y Zou, and Been Kim. 2019. Towards automatic concept-based explanations. *Advances in neural information processing systems* 32 (2019).

[46] Steven M Goodman, Ping Liu, Dhruv Jain, Emma J McDonnell, Jon E Froehlich, and Leah Findlater. 2021. Toward user-driven sound recognizer personalization with people who are d/deaf or hard of hearing. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 5, 2 (2021), 1–23.

[47] Hongyan Gu, Jingbin Huang, Lauren Hung, and Xiang'Anthony' Chen. 2021. Lessons learned from designing an AI-enabled diagnosis tool for pathologists. *Proceedings of the ACM on Human-Computer Interaction* 5, CSCW1 (2021), 1–25.

[48] Bruce Hanington and Bella Martin. 2019. *Universal methods of design expanded and revised: 125 Ways to research complex problems, develop innovative ideas, and design effective solutions*. Rockport publishers.

[49] Katharine E Henry, Rachel Kornfield, Anirudh Sridharan, Robert C Linton, Catherine Groh, Tony Wang, Albert Wu, Bilge Mutlu, and Suchi Saria. 2022. Human–machine teaming is key to AI adoption: clinicians' experiences with a deployed machine learning system. *NPJ digital medicine* 5, 1 (2022), 97.

[50] Jonathan Hey, Julie Linsey, Alice M Agogino, and Kristin L Wood. 2008. Analogies and metaphors in creative design. *International Journal of Engineering Education* 24, 2 (2008), 283.

[51] Hui-Ru Ho, Edward M Hubbard, and Bilge Mutlu. 2024. " It's Not a Replacement:" Enabling Parent-Robot Collaboration to Support In-Home Learning Experiences of Young Children. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*. 1–18.

[52] Wayne Holmes and Ilkka Tuomi. 2022. State of the art and practice in AI in education. *European Journal of Education* 57, 4 (2022), 542–570.

[53] Eva Hornecker. 2010. Creative idea exploration within the structure of a guiding framework: the card brainstorming game. In *Proceedings of the fourth international conference on Tangible, embedded, and embodied interaction*. 101–108.

[54] Gary Hsieh, Brett A Halperin, Evan Schmitz, Yen Nee Chew, and Yuan-Chi Tseng. 2023. What is in the Cards: Exploring Uses, Patterns, and Trends in Design Cards. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. 1–18.

[55] Brian Hu, Paul Tunison, Bhavan Vasu, Nitesh Menon, Roddy Collins, and Anthony Hoogs. 2021. XAITK: The explainable AI toolkit. *Applied AI Letters* 2, 4 (2021), e40.

[56] Yaxin Hu, Hajin Lim, Hailey L Johnson, Josephine M O'Shaughnessy, Lisa Kakonge, Lyn Turkstra, Melissa Duff, Catalina Toma, and Bilge Mutlu. 2023. Investigating Day-to-day Experiences with Conversational Agents by Users with Traumatic Brain Injury. In *Proceedings of the 25th International ACM SIGACCESS Conference on Computers and Accessibility*. 1–15.

[57] Yaxin Hu, Yuxiao Qu, Adam Maus, and Bilge Mutlu. 2022. Polite or direct? Conversation design of a smart display for older adults based on politeness theory. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*. 1–15.

[58] IDEO. 2003. IDEO Method Cards: 51 ways to inspire design. https://www.ideo.com

[59] IDEO. 2014. IDEO Nature Cards. https://www.ideo.com

[60] Fei Jiang, Yong Jiang, Hui Zhi, Yi Dong, Hao Li, Sufeng Ma, Yilong Wang, Qiang Dong, Haipeng Shen, and Yongjun Wang. 2017. Artificial intelligence in healthcare: past, present and future. *Stroke and vascular neurology* 2, 4 (2017).

[61] Jinglu Jiang, Surinder Kahai, and Ming Yang. 2022. Who needs explanation and when? Juggling explainable AI and user epistemic uncertainty. *International Journal of Human-Computer Studies* 165 (2022), 102839.

[62] Annika Kaltenhauser, Verena Rheinstädter, Andreas Butz, and Dieter P Wallach. 2020. " You Have to Piece the Puzzle Together" Implications for Designing Decision Support in Intensive Care. In *Proceedings of the 2020 ACM Designing Interactive Systems Conference*. 1509–1522.

[63] Atreyi Kankanhalli, Yannis Charalabidis, and Sehl Mellouli. 2019. IoT and AI for smart government: A research agenda. , 304–309 pages.

[64] Mark T Keane and Eoin M Kenny. 2019. How case-based reasoning explains neural networks: A theoretical analysis of XAI using post-hoc explanation-by-example from a survey of ANN-CBR twin-systems. In *Case-Based Reasoning Research and Development: 27th International Conference, ICCBR 2019, Otzenhausen, Germany, September 8–12, 2019, Proceedings 27*. Springer, 155–171.

[65] Patrick Gage Kelley, Joanna Bresee, Lorrie Faith Cranor, and Robert W Reeder. 2009. A" nutrition label" for privacy. In *Proceedings of the 5th Symposium on Usable Privacy and Security*. 1–12.

[66] Eoin M Kenny, Courtney Ford, Molly Quinn, and Mark T Keane. 2021. Explaining black-box classifiers using post-hoc explanations-by-example: The effect of explanations and error-rates in XAI user studies. *Artificial Intelligence* 294 (2021), 103459.

[67] 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.

[68] Mallory C Kidwell, Ljiljana B Lazarević, Erica Baranski, Tom E Hardwicke, Sarah Piechowski, Lina-Sophia Falkenberg, Curtis Kennett, Agnieszka Slowik, Carina Sonnleitner, Chelsey Hess-Holden, et al. 2016. Badges to acknowledge open practices: A simple, low-cost, effective method for increasing transparency. *PLoS biology* 14, 5 (2016), e1002456.

[69] Been Kim, Martin Wattenberg, Justin Gilmer, Carrie Cai, James Wexler, Fernanda Viegas, et al. 2018. Interpretability beyond feature attribution: Quantitative testing with concept activation vectors (tcav). In *International conference on machine learning*. PMLR, 2668–2677.

[70] Callie Y Kim, Christine P Lee, and Bilge Mutlu. 2024. Understanding Large-Language Model (LLM)-powered Human-Robot Interaction. In *Proceedings of the 2024 ACM/IEEE International Conference on Human-Robot Interaction*. 371–380.

[71] Tzu-Sheng Kuo, Hong Shen, Jisoo Geum, Nev Jones, Jason I Hong, Haiyi Zhu, and Kenneth Holstein. 2023. Understanding Frontline Workers' and Unhoused Individuals' Perspectives on AI Used in Homeless Services. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. 1–17.

[72] Jokin Labaien, Ekhi Zugasti, and Xabier De Carlos. 2020. Contrastive explanations for a deep learning model on time-series data. In *International Conference on Big Data Analytics and Knowledge Discovery*. Springer, 235–244.

[73] Himabindu Lakkaraju, Stephen H Bach, and Jure Leskovec. 2016. Interpretable decision sets: A joint framework for description and prediction. In *Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining*. 1675–1684.

[74] Christine P Lee, Bengisu Cagiltay, and Bilge Mutlu. 2022. The unboxing experience: Exploration and design of initial interactions between children and social robots. In *Proceedings of the 2022 CHI conference on human factors in computing systems*. 1–14.

[75] Min Kyung Lee, Daniel Kusbit, Anson Kahng, Ji Tae Kim, Xinran Yuan, Allissa Chan, Daniel See, Ritesh Noothigattu, Siheon Lee, Alexandros Psomas, et al. 2019. WeBuildAI: Participatory framework for algorithmic governance. *Proceedings of the ACM on Human-Computer Interaction* 3, CSCW (2019), 1–35.

[76] Tianshi Li, Kayla Reiman, Yuvraj Agarwal, Lorrie Faith Cranor, and Jason I Hong. 2022. Understanding challenges for developers to create accurate privacy nutrition labels. In *Proceedings of the 2022 CHI Conference on Human Factors in*

Computing Systems. 1–24.
[77] Q Vera Liao, Daniel Gruen, and Sarah Miller. 2020. Questioning the AI: informing design practices for explainable AI user experiences. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. 1–15.
[78] Q Vera Liao and S Shyam Sundar. 2022. Designing for Responsible Trust in AI Systems: A Communication Perspective. *arXiv preprint arXiv:2204.13828* (2022).
[79] Q Vera Liao and Kush R Varshney. 2021. Human-centered explainable ai (xai): From algorithms to user experiences. *arXiv preprint arXiv:2110.10790* (2021).
[80] Brian Y Lim and Anind K Dey. 2010. Toolkit to support intelligibility in context-aware applications. In *Proceedings of the 12th ACM international conference on Ubiquitous computing*. 13–22.
[81] Houjiang Liu, Anubrata Das, Alexander Boltz, Didi Zhou, Daisy Pinaroc, Matthew Lease, and Min Kyung Lee. 2023. Human-centered NLP Fact-checking: Co-Designing with Fact-checkers using Matchmaking for AI. *arXiv preprint arXiv:2308.07213* (2023).
[82] Daniel Lockton. 2013. *Design with intent: a design pattern toolkit for environmental and social behaviour change*. Ph. D. Dissertation. Brunel University School of Engineering and Design PhD Theses.
[83] Dan Lockton, David Harrison, and Neville A Stanton. 2010. The Design with Intent Method: A design tool for influencing user behaviour. *Applied ergonomics* 41, 3 (2010), 382–392.
[84] Owen HT Lu, Anna YQ Huang, Danny CL Tsai, and Stephen JH Yang. 2021. Expert-authored and machine-generated short-answer questions for assessing students learning performance. *Educational Technology & Society* 24, 3 (2021), 159–173.
[85] Andrés Lucero, Peter Dalsgaard, Kim Halskov, and Jacob Buur. 2016. Designing with cards. *Collaboration in creative design: Methods and tools* (2016), 75–95.
[86] Scott M Lundberg and Su-In Lee. 2017. A unified approach to interpreting model predictions. *Advances in neural information processing systems* 30 (2017).
[87] 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 (11 2019), 1–23. https://doi.org/10.1145/3359174
[88] Tim Miller. 2019. Explanation in artificial intelligence: Insights from the social sciences. *Artificial intelligence* 267 (2019), 1–38.
[89] Swati Mishra and Jeffrey M Rzeszotarski. 2021. Designing interactive transfer learning tools for ML non-experts. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. 1–15.
[90] Brent Mittelstadt, Chris Russell, and Sandra Wachter. 2019. Explaining explanations in AI. In *Proceedings of the conference on fairness, accountability, and transparency*. 279–288.
[91] Jakki Mohr and John R Nevin. 1990. Communication strategies in marketing channels: A theoretical perspective. *Journal of marketing* 54, 4 (1990), 36–51.
[92] Jakki Mohr and Robert Spekman. 1994. Characteristics of partnership success: partnership attributes, communication behavior, and conflict resolution techniques. *Strategic management journal* 15, 2 (1994), 135–152.
[93] Jakki J Mohr and Ravipreet S Sohi. 1995. Communication flows in distribution channels: Impact on assessments of communication quality and satisfaction. *Journal of retailing* 71, 4 (1995), 393–415.
[94] Cecily Morrison, Edward Cutrell, Martin Grayson, Anja Thieme, Alex Taylor, Geert Roumen, Camilla Longden, Sebastian Tschiatschek, Rita Faia Marques, and Abigail Sellen. 2021. Social Sensemaking with AI: Designing an Open-ended AI experience with a Blind Child. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. 1–14.
[95] Florian Mueller, Martin R Gibbs, Frank Vetere, and Darren Edge. 2014. Supporting the creative game design process with exertion cards. In *Proceedings of the sigchi conference on human factors in computing systems*. 2211–2220.
[96] Dave Murray-Rust, Iohanna Nicenboim, and Dan Lockton. 2022. Metaphors for designers working with AI. (2022).
[97] Camille Nadal, Shane McCully, Kevin Doherty, Corina Sas, and Gavin Doherty. 2022. The TAC toolkit: supporting design for user acceptance of health technologies from a macro-temporal perspective. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*. 1–18.
[98] Ashok Kumar Reddy Nadikattu. 2021. Influence of Artificial Intelligence on Robotics Industry. *International Journal of Creative Research Thoughts (IJCRT), ISSN* (2021), 2320–2882.
[99] Yasmina Okan, Rocio Garcia-Retamero, Edward T Cokely, and Antonio Maldonado. 2015. Improving risk understanding across ability levels: Encouraging active processing with dynamic icon arrays. *Journal of Experimental Psychology: Applied* 21, 2 (2015), 178.
[100] Yasmina Okan, Rocio Garcia-Retamero, Mirta Galesic, and Edward T Cokely. 2012. When higher bars are not larger quantities: On individual differences in the use of spatial information in graph comprehension. *Spatial Cognition & Computation* 12, 2-3 (2012), 195–218.
[101] Long Ouyang, Jeffrey Wu, Xu Jiang, Diogo Almeida, Carroll Wainwright, Pamela Mishkin, Chong Zhang, Sandhini Agarwal, Katarina Slama, Alex Ray, et al. 2022. Training language models to follow instructions with human feedback. *Advances in Neural Information Processing Systems* 35 (2022), 27730–27744.
[102] Andrei Paleyes, Raoul-Gabriel Urma, and Neil D Lawrence. 2022. Challenges in deploying machine learning: a survey of case studies. *ACM computing surveys* 55, 6 (2022), 1–29.
[103] Hyanghee Park, Daehwan Ahn, Kartik Hosanagar, and Joonhwan Lee. 2022. Designing fair AI in human resource management: Understanding tensions surrounding algorithmic evaluation and envisioning stakeholder-centered solutions. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*. 1–22.
[104] James Pierce and Carl DiSalvo. 2018. Addressing network anxieties with alternative design metaphors. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. 1–13.
[105] Melanie Price, Rachel Cameron, and Phyllis Butow. 2007. Communicating risk information: the influence of graphical display format on quantitative information perception—accuracy, comprehension and preferences. *Patient education and counseling* 69, 1-3 (2007), 121–128.
[106] Marco Tulio Ribeiro, Sameer Singh, and Carlos Guestrin. 2016. " Why should i trust you?" Explaining the predictions of any classifier. In *Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining*. 1135–1144.
[107] Marco Tulio Ribeiro, Sameer Singh, and Carlos Guestrin. 2016. Model-agnostic interpretability of machine learning. *arXiv preprint arXiv:1606.05386* (2016).
[108] Marco Tulio Ribeiro, Sameer Singh, and Carlos Guestrin. 2018. Anchors: High-precision model-agnostic explanations. In *Proceedings of the AAAI conference on artificial intelligence*, Vol. 32.
[109] Marcel Jurriaan Robeer. 2018. *Contrastive explanation for machine learning*. Master's thesis.
[110] Robin Roy and James P Warren. 2019. Card-based design tools: A review and analysis of 155 card decks for designers and designing. *Design Studies* 63 (2019), 125–154.
[111] Dennis Schlegel, Kajetan Schuler, and Jens Westenberger. 2023. Failure factors of AI projects: results from expert interviews. *International journal of information systems and project management: IJISPM* 11, 3 (2023), 25–40.
[112] Ralf Schmälzle and Shelby Wilcox. 2022. Harnessing artificial intelligence for health message generation: The folic acid message engine. *Journal of Medical Internet Research* 24, 1 (2022), e28858.
[113] Tjeerd AJ Schoonderwoerd, Wiard Jorritsma, Mark A Neerincx, and Karel Van Den Bosch. 2021. Human-centered XAI: Developing design patterns for explanations of clinical decision support systems. *International Journal of Human-Computer Studies* 154 (2021), 102684.
[114] Anna Marie Schröder and Maliheh Ghajargar. 2021. Unboxing the algorithm: Designing an understandable algorithmic experience in music recommender systems. In *15th ACM Conference on Recommender Systems (RecSys 2021), Amsterdam, The Netherlands, September 25, 2021*.
[115] Hugo Scurto, Baptiste Caramiaux, and Frédéric Bevilacqua. 2021. Prototyping machine learning through diffractive art practice. In *Designing Interactive Systems Conference 2021*. 2013–2025.
[116] Mohammad Ahmad Sheikh, Amit Kumar Goel, and Tapas Kumar. 2020. An approach for prediction of loan approval using machine learning algorithm. In *2020 International Conference on Electronics and Sustainable Communication Systems (ICESC)*. IEEE, 490–494.
[117] Chaehan So. 2020. Human-in-the-loop design cycles–A process framework that integrates design sprints, agile processes, and machine learning with humans. In *Artificial Intelligence in HCI: First International Conference, AI-HCI 2020, Held as Part of the 22nd HCI International Conference, HCII 2020, Copenhagen, Denmark, July 19–24, 2020, Proceedings 22*. Springer, 136–145.
[118] Aaron Springer and Steve Whittaker. 2019. Progressive disclosure: empirically motivated approaches to designing effective transparency. In *Proceedings of the 24th international conference on intelligent user interfaces*. 107–120.
[119] Aaron Springer and Steve Whittaker. 2020. Progressive disclosure: When, why, and how do users want algorithmic transparency information? *ACM Transactions on Interactive Intelligent Systems (TiiS)* 10, 4 (2020), 1–32.
[120] Logan Stapleton, Min Hun Lee, Diana Qing, Marya Wright, Alexandra Chouldechova, Ken Holstein, Zhiwei Steven Wu, and Haiyi Zhu. 2022. Imagining new futures beyond predictive systems in child welfare: A qualitative study with impacted stakeholders. In *Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency*. 1162–1177.
[121] Hariharan Subramonyam, Colleen Seifert, and Eytan Adar. 2021. Protoai: Model-informed prototyping for ai-powered interfaces. In *26th International Conference on Intelligent User Interfaces*. 48–58.
[122] Hariharan Subramonyam, Colleen Seifert, and Eytan Adar. 2021. Towards a process model for co-creating AI experiences. In *Designing Interactive Systems Conference 2021*. 1529–1543.
[123] Jiao Sun, Q Vera Liao, Michael Muller, Mayank Agarwal, Stephanie Houde, Kartik Talamadupula, and Justin D Weisz. 2022. Investigating Explainability of Generative AI for Code through Scenario-based Design. In *27th International Conference on Intelligent User Interfaces*. 212–228.
[124] Khai N Truong, Gillian R Hayes, and Gregory D Abowd. 2006. Storyboarding: an empirical determination of best practices and effective guidelines. In *Proceedings*

of the 6th conference on Designing Interactive systems. 12–21.

[125] Huey-Ru Tsai, Yasser Shoukry, Min Kyung Lee, and Vasumathi Raman. 2017. Towards a socially responsible smart city: dynamic resource allocation for smarter community service. In *Proceedings of the 4th ACM International Conference on Systems for Energy-Efficient Built Environments*. 1–4.

[126] Zaib Ullah, Fadi Al-Turjman, Leonardo Mostarda, and Roberto Gagliardi. 2020. Applications of artificial intelligence and machine learning in smart cities. *Computer Communications* 154 (2020), 313–323.

[127] Philip Van Allen. 2018. Prototyping ways of prototyping AI. *Interactions* 25, 6 (2018), 46–51.

[128] Jasper van der Waa, Elisabeth Nieuwburg, Anita Cremers, and Mark Neerincx. 2021. Evaluating XAI: A comparison of rule-based and example-based explanations. *Artificial Intelligence* 291 (2021), 103404.

[129] Arnold POS Vermeeren, Virpi Roto, and Kaisa Väänänen. 2016. Design-inclusive UX research: design as a part of doing user experience research. *Behaviour & Information Technology* 35, 1 (2016), 21–37.

[130] Giulia Vilone and Luca Longo. 2021. A quantitative evaluation of global, rule-based explanations of post-hoc, model agnostic methods. *Frontiers in artificial intelligence* 4 (2021), 717899.

[131] Sruthi Viswanathan, Cecile Boulard, Adrien Bruyat, and Antonietta Maria Grasso. 2022. Situational Recommender: Are You On the Spot, Refining Plans, or Just Bored?. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*. 1–19.

[132] Jerold W Wallis and Edward H Shortliffe. 1984. Customized explanations using causal knowledge. *Rule-Based Expert Systems: The MYCIN Experiments of the Stanford Heuristic Programming Project* (1984), 371–388.

[133] Ruyuan Wan, Adriana Alvarado Garcia, Devansh Saxena, Catalina Vajiac, Anna Kawakami, Logan Stapleton, Haiyi Zhu, Kenneth Holstein, Heloisa Candello, and Karla Badillo-Urquiola. 2023. Community-driven AI: Empowering people through responsible data-driven decision-making. In *Companion Publication of the 2023 Conference on Computer Supported Cooperative Work and Social Computing*. 532–536.

[134] Danding Wang, Qian Yang, Ashraf Abdul, and Brian Y Lim. 2019. Designing theory-driven user-centric explainable AI. In *Proceedings of the 2019 CHI conference on human factors in computing systems*. 1–15.

[135] Qiaosi Wang, Ida Camacho, Shan Jing, and Ashok K Goel. 2022. Understanding the Design Space of AI-Mediated Social Interaction in Online Learning: Challenges and Opportunities. *Proceedings of the ACM on Human-Computer Interaction* 6, CSCW1 (2022), 1–26.

[136] Greta Warren, Mark T Keane, and Ruth MJ Byrne. 2022. Features of Explainability: How users understand counterfactual and causal explanations for categorical and continuous features in XAI. *arXiv preprint arXiv:2204.10152* (2022).

[137] Joyce Weiner. 2022. *Why AI/data science projects fail: how to avoid project pitfalls*. Springer Nature.

[138] Jens Westenberger, Kajetan Schuler, and Dennis Schlegel. 2022. Failure of AI projects: understanding the critical factors. *Procedia computer science* 196 (2022), 69–76.

[139] Christine T Wolf. 2019. Explainability scenarios: towards scenario-based XAI design. In *Proceedings of the 24th International Conference on Intelligent User Interfaces*. 252–257.

[140] Christine T Wolf and Kathryn E Ringland. 2020. Designing accessible, explainable AI (XAI) experiences. *ACM SIGACCESS Accessibility and Computing* 125 (2020), 1–1.

[141] Christiane Wölfel and Timothy Merritt. 2013. Method card design dimensions: A survey of card-based design tools. In *Human-Computer Interaction–INTERACT 2013: 14th IFIP TC 13 International Conference, Cape Town, South Africa, September 2-6, 2013, Proceedings, Part I 14*. Springer, 479–486.

[142] Qian Yang, Yuexing Hao, Kexin Quan, Stephen Yang, Yiran Zhao, Volodymyr Kuleshov, and Fei Wang. 2023. Harnessing biomedical literature to calibrate clinicians' trust in AI decision support systems. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. 1–14.

[143] Qian Yang, Aaron Steinfeld, Carolyn Rosé, and John Zimmerman. 2020. Re-examining whether, why, and how human-AI interaction is uniquely difficult to design. In *Proceedings of the 2020 chi conference on human factors in computing systems*. 1–13.

[144] Kun-Hsing Yu, Andrew L Beam, and Isaac S Kohane. 2018. Artificial intelligence in healthcare. *Nature biomedical engineering* 2, 10 (2018), 719–731.

[145] Angie Zhang, Alexander Boltz, Jonathan Lynn, Chun-Wei Wang, and Min Kyung Lee. 2023. Stakeholder-Centered AI Design: Co-Designing Worker Tools with Gig Workers through Data Probes. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. 1–19.

[146] Wencan Zhang and Brian Y Lim. 2022. Towards relatable explainable AI with the perceptual process. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*. 1–24.

[147] Yunfeng Zhang, Q Vera Liao, and Rachel KE Bellamy. 2020. Effect of confidence and explanation on accuracy and trust calibration in AI-assisted decision making. In *Proceedings of the 2020 conference on fairness, accountability, and transparency*. 295–305.

[148] Chengbo Zheng, Dakuo Wang, April Yi Wang, and Xiaojuan Ma. 2022. Telling stories from computational notebooks: Ai-assisted presentation slides creation for presenting data science work. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*. 1–20.

[149] Gabe Zichermann and Joselin Linder. 2010. *Game-based marketing: inspire customer loyalty through rewards, challenges, and contests*. John Wiley & Sons.

[150] Brian J Zikmund-Fisher, Angela Fagerlin, and Peter A Ubel. 2010. A demonstration of "less can be more"in risk graphics. *Medical Decision Making* 30, 6 (2010), 661–671.

[151] Brian J Zikmund-Fisher, Holly O Witteman, Mark Dickson, Andrea Fuhrel-Forbis, Valerie C Kahn, Nicole L Exe, Melissa Valerio, Lisa G Holtzman, Laura D Scherer, and Angela Fagerlin. 2014. Blocks, ovals, or people? Icon type affects risk perceptions and recall of pictographs. *Medical decision making* 34, 4 (2014), 443–453.

[152] John Zimmerman, Jodi Forlizzi, and Shelley Evenson. 2007. Research through design as a method for interaction design research in HCI. In *Proceedings of the SIGCHI conference on Human factors in computing systems*. 493–502.

[153] Douglas Zytko, Pamela J. Wisniewski, Shion Guha, Eric PS Baumer, and Min Kyung Lee. 2022. Participatory design of AI systems: Opportunities and challenges across diverse users, relationships, and application domains. In *CHI Conference on Human Factors in Computing Systems Extended Abstracts*. 1–4.

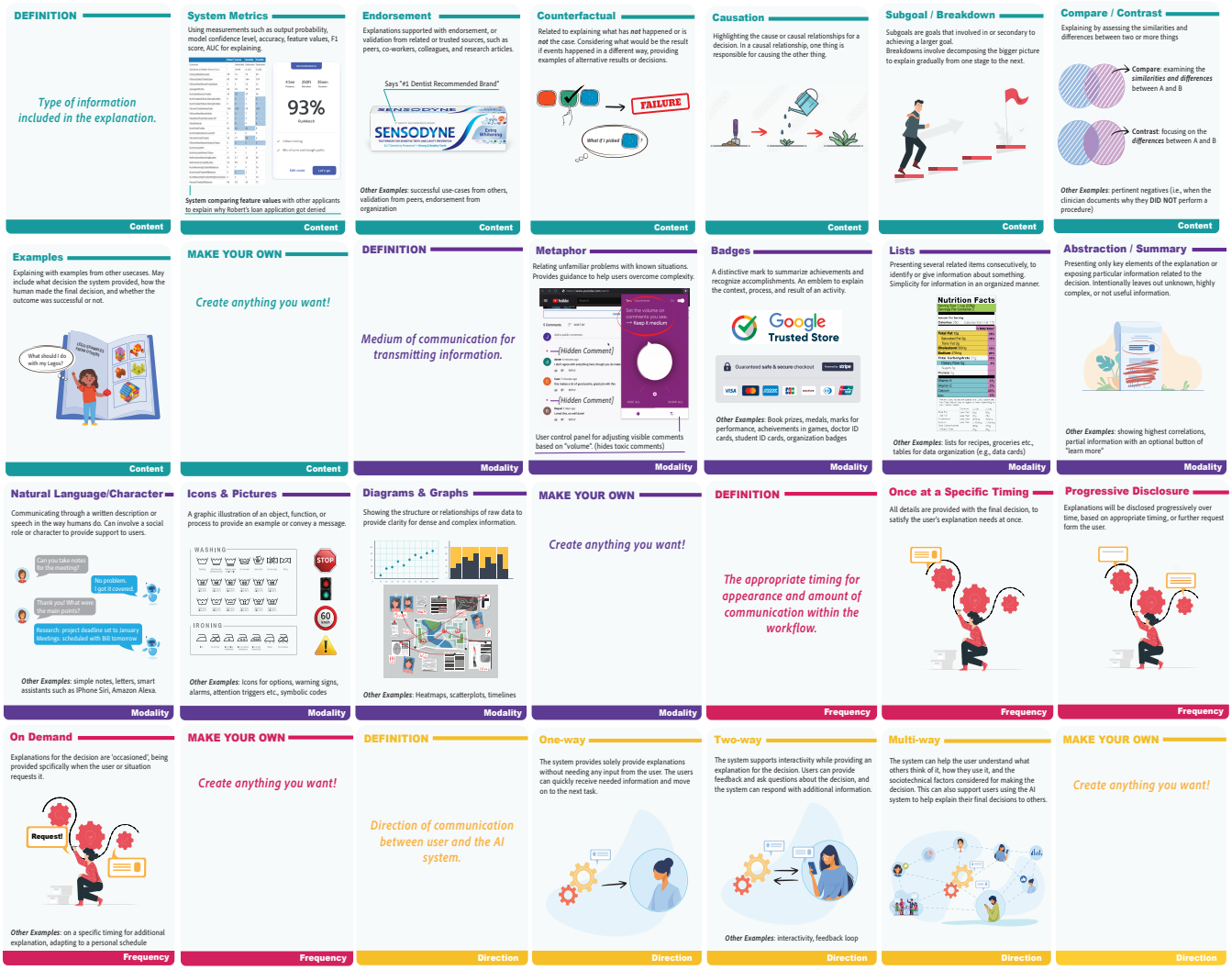


Figure 5: *Complete Set of AI-DEC* — The AI-DEC consists of four types of cards, content, modality, frequency, and direction. In each facet, there are cards with design elements derived from existing design research in ML, HCI, human-robot interaction (HRI), and XAI. When utilizing the AI-DEC, end-users will select cards from the four communication facets to assemble their preferred AI explanation. The resulting explanation design will consist of a combination of cards from each facet, forming a tailored design solution to address user information or interaction requirements. This design solution will entail the content of the explanation, the mode of delivery, the timing or frequency of presentation, and the required level of user interactivity.

PID	Selected Question Bank Category	AI Explanation Design with AI-DEC
H1	performance	compare, example, metrics, badges, endorsement, icons, once, one, multi-way
	actions	metrics, endorsement, example, diagrams, progressive, multi-way
	input	metrics, example, compare, diagrams, "color-coding", on demand, two-way
	output	metrics, compare, example, diagrams, icons, once, one-way
	how	metrics, compare, example, diagram, on demand, two-way
H2	performance	example, endorsement, icons, lists, progressive, multi-way
	actions	subgoal, list, on demand, two-way
	how	caustion, summary, language, on demand, multi-way
	output	example, lists, diagram, on demand, multi-way
	why	metrics, language, on demand, multi-way
H3	action	subgoal, counterfactual, list, on demand, two-way
	why	metrics, counterfactual, examples, lists, diagrams, on demand, two-way
		causation, breakdown, compare, metaphor, diagrams, on demand, one-way
	why not	counterfactual, metrics, list, diagram, on demand, two-way
	social	metrics, examples, badges, diagram, language, on demand, multi-way
	input	example, endorsement, summary, diagram, once, one-way
H4	input	metrics, compare, diagrams, summary, lists, once, one-way
	actions	metrics, summary, diagrams, on demand, one-way
	social	icons, diagrams, metrics, "when occurrs", one-way
	output	causation, metrics, diagrams, once, one-way
	why	causation, examples, compare, summary, icons, on demand, one-way
	how	metrics, summary, diagrams, on demand, one-way
H5	performance	"limitations", diagrams, badges, "whenever occurs", two-way
	input	compare, diagram, language, on demand, two-way
	how to still be that	metrics, causation, diagram, progressive disclosure, two-way
	how	summary, compare, diagram, progressive disclosure, on demand, multi-way
	why	causation, breakdown, language, diagram, on demand, progressive disclosure, one-way
H6	performance	causation, "limitations", badges, diagrams, once, on demand, two-way
	output	"alternative possibilities", metrics, "rankings", list, on demand, one-way
		causation, metrics, examples, lists, diagrams, natural language, once at a specific timing, two-way
	input	metrics, compare, badges, list, diagrams, on demand, two-way
M1	input	subgoal, causation, examples, language, on demand, multi-way
	social	compare, language, icons, progressive, multi-way
	actions	metrics, diagram, list, on demand, two-way
	how	causation, examples, summary, language, progressive, two-way
	performance	counterfactual, summary, compare, progressive, icons, language, multi-way
M2	performance	example, metaphor, endorsement, on demand, one-way
	social	examples, language, on demand, multi-way
	input	causation, breakdown, summary, list, "when encountering untrained", two-way
	why	compare, counterfactual, language, on demand, multi-way
	output	metrics, diagram, once, one-way
M3	social	metrics, subgoals, natural language, multi-way
	input	metrics, subgoal, compare, badges, language, metaphor, on demand, two-way, multi-way
	output	causation, metrics, endorsement, diagrams, lists, two-way
	performance	metrics, endorsement, lists, diagrams, badges, on demand, one-way
	how to still be this	examples, metrics, counterfactual, causation, compare, language, on demand, two-way
M4	why	example, summary, diagram, on demand, two-way
	input	subgoal, metrics, list, on demand, progressive, one-way
	how	causation, subgoal, "interrelationship", "highlight", list, on demand, multi-way
	performance	example, language, "flagging", two-way

M5	why	causation, list, once, multi-way
	social	"descriptions of others progress", summary, icons, diagram, once, multi-way
	how	compare, diagram, list, on demand, two-way
	why not	compare, list, on demand, two-way
	output	examples, list, metaphor, on demand, two-way
	performance	metrics, list, once, one-way
F1	input	compare, metrics, diagram, icons, progressive, two-way
	how to be that	subgoal, causation, list, on demand, two-way
	actions	examples, "results from similar past interaction results", language, progressive, multi-way
F2	how to be that	causation, counterfactual, lists, on demand, two-way
	action	subgoal, examples, compare, language, list, progressive, two-way
	input	metrics, compare, summary, diagrams, lists, on demand, one-way
	social	language, summary, on demand, multi-way
	performance	endorsement, icons, badges, on demand, one-way
F3	performance	compare, metrics, example, diagram, list, progressive, two-way
	social	example, icons, metrics, diagram, lists, progressive, multi-way
	what if	compare, counterfactual, metaphor, language, on demand, two-way
	why	compare, counterfactual, diagram, language, on demand, two-way
	how to be that	metrics, compare, counterfactual, diagram, list, on demand, two-way
F4	performance	metrics, lists, once, one-way
	input	example, language, on demand, multi-way
	social	examples, "results of others", lists, on demand, one-way
	output	subgoal, metaphor, on demand, multi-way
	actions	subgoal, diagram, on demand, multi-way
F5	why	compare, causation, metrics, diagram, language, progressive disclosure, two-way
	how to be that	counterfactual, breakdown, examples, "past interactions from workers in higher hierarchy", language, on demand, multi-way
	how	metrics, example, "past interactions from workers in higher hierarchy", summary, language, list, progressive disclosure, on demand, multi-way
	how to still be this	summary, icons, "ranking", as needed, one-way
	input	compare, examples, "past interactions from workers in higher hierarchy", badges, metaphor, as needed, multi-way

Figure 6: *Full Set of Participant Design Solution* — The table presents the complete set of participants' explanation designs utilizing the AI-DEC during the co-design session.