The Report Committee for Vanessa Sanchez

Certifies that this is the approved version of the following report:

Insights from Participatory Workshops: Enhancing Human-Centered Design of Responsible AI Tools and a Proposed RAI Resources Cataloging System

SUPERVISING COMMITTEE:

Min Kyung Lee, Supervisor

Kenneth R. Fleischmann, Co-supervisor

Insights from Participatory Workshops: Enhancing Human-Centered Design of Responsible AI Tools and a Proposed RAI Resources Cataloging System

by

Vanessa Sanchez

Report

Presented to the Faculty of the Graduate School of
The University of Texas at Austin
in Partial Fulfillment
of the Requirements
for the Degree of

Master of Science in Information Studies

The University of Texas at Austin

August 2024

Dedication

To those who have inspired and supported me on my journey and continue to do so.

Epigraph

What starts here changes the world.

—University of Texas at Austin

Acknowledgments

Thanks to all those who contributed to my ability to write this report, be it through an inspiring course, mentorship, helping me recognize my strengths, sharing resources, mental check-ins, supportive messages, a friendly discussion over coffee, IRB advice, or shared learning moments while collaborating on a project that became part of the foundation for this paper. In particular, I want to acknowledge Dr. Min Kyung Lee, Dr. Kenneth R. Fleischmann, Dr. John Neumann, Dr. Elliott Hauser, Chelsea Collier, Tina Lassiter, Angie Zhang, Whitney Nelson, Jonathan Lynn, and The University of Texas School of Information.

Preface

Readers, this document represents a great deal of time and effort. Enjoy.

—Vanessa Sanchez

Abstract

Insights from Participatory Workshops: Enhancing Human-Centered Design of Responsible AI Tools and a Proposed RAI Resources Cataloging System

Vanessa Sanchez, MSIS
The University of Texas at Austin, 2024

SUPERVISORS: Min Kyung Lee, Kenneth R. Fleischmann

The proliferation of AI ethics principles, frameworks, and toolkits has led to a saturation of resources, many of which are reported as ineffective in practice (Mittelstadt 2019; Crawford 2021; Munn 2023). Professionals need the ability to identify and access relevant responsible AI (RAI) resources that translate abstract principles into tailored processes, tangible tools, and measurable success metrics. By integrating the situated perspectives of various RAI stakeholders, the design and cataloging of RAI tools can be significantly improved. This paper addresses the existing challenges of ineffective centralized cataloging and the inaccessible design of RAI tools. The study explores the RAI experiences of professionals across multiple sectors, focusing on what they value in RAI tools, as revealed through participatory workshops. Data was collected from a survey of 14 participants and in-depth interviews with 12 of the participants, conducted remotely via Zoom. The participants included AI ethicists, start-up consultants, a lawyer, engineers, a scientist, UX professionals, and researchers from various industries. Analysis of the interview transcripts identified three key themes: (1) Participants struggle to determine which RAI tool to use; (2) There is a strong need for customized solutions and better coordination in RAI practices; and (3) Participants emphasize the importance of inclusiveness and accessibility in the design of RAI tools. These findings will provide valuable insights for making RAI resources more accessible, useful, and usable for professionals aiming to enhance their RAI practices.

Table of Contents

List of Tables

List of Figures

Chapter One: Introduction

Introduction

Problem Statement

This paper focuses on the inaccessibility of responsible artificial intelligence (RAI) resources and tools, including the lack of effective centralized cataloging. To address this set of problems, this paper will explore the RAI experiences of practitioners across various sectors who were participants in the study to examine what they find valuable in RAI resources and tools. Through transcript analysis and the outputs of participatory one-on-one workshops, this paper will examine not only what criteria are relevant to participants when seeking RAI tools and resources, but also the human factors that influence these criteria.

Translating Ethical AI Principles into Everyday Practice

AI ethics principles, frameworks and toolkits have become copious and are reported as ineffective (Mittelstadt 2019; Crawford 2021; Munn 2023). At best, they are well-meaning attempts at AI governance from too high a perspective to be useful to AI teams or AI customers. At worst, they are sly attempts at ethics-washing to protect public image and avoid regulation (van Maanen 2022; Ochigame 2019). Professionals need the ability to find relevant RAI resources that translate abstract principles into tailored processes, tangible tools, and measurable metrics for success. RAI resources and tools can better support RAI practices when they are created for specific industries and professional contexts while also being inclusive and accessible to the various RAI stakeholders in an organizational ecosystem.

The Relevance of Practitioners Beyond AI Developers

There has already been growing recognition of the relevance of stakeholders beyond the development team or broader product team to the practice of RAI. Practitioners in legal roles, finance roles, executive roles, data security roles, and vendor roles (to name a few) are potentially relevant to the successful implementation of RAI operations and, more concretely, day-to-day RAI practices. An organization is an ecosystem of people, processes, and information. If AI is part of that ecosystem, its development or use will have a ripple effect across the organization. This makes RAI a relevant issue to everyone. To develop an effective RAI strategy, it's essential to first fully understand the organizational environment and all the people who may intersect with RAI practices and tools. Each RAI stakeholder brings a different perspective that could be relevant to a successful strategy.

Impact on Stakeholders, Fields, and Industries

This study will be relevant to all industries involved with AI, AI governance organizations, and RAI government agencies and NGOs. This study may be relevant to startups in particular, as they are perhaps at the greatest disadvantage with regards to affordability of RAI consultants or in-house AI ethicists. The emergent themes of the study will be useful for understanding how to make RAI resources more accessible, useful, and usable for a wide range of practitioners seeking tools to support their RAI practices, especially within start-ups, small teams, sole proprietorships, and for non-technical practitioners.

Study Rationale

Why This Study is Needed

A search on Google Scholar—keywords "findability and accessibility of RAI tools", "findability of RAI tools", "activity theory and RAI tools," "situated action and RAI tools", "RAI resource catalogs"—reveals no prior work regarding findability or accessibility in the design of RAI tools and resources, yet this was cited as a pain point by several participants. There are stakeholders proactively

seeking RAI resources and tools but struggle in knowing where to look, what to look for, and how to evaluate the resources. It's conceivable that in the future, RAI audits may become more commonplace, increasing the need for RAI resources and tools among practitioners, AI governance teams, in-house AI ethicists, and 3rd party RAI consultants. RAI resource databases exist, but are not centralized, comprehensive, nor effectively designed to support users in evaluating the resources.

What This Paper Aims to Contribute

This paper will address definition, discoverability, relevance, accessibility, customization, and evaluation of RAI tools. It aims to contribute a human-centered design perspective in the design of RAI tools and the design of a free centralized RAI resources library. This paper also aims to contribute discussion on how situated perspectives influence a professional's entry point to RAI tools; their experience of recognizing and defining their needs; their experience of finding, accessing, and applying the tools in daily practice; and finally, their ways of evaluating the effectiveness of the tool. By necessity, this paper will also attempt to define what is meant by RAI resources and RAI tools, respectively.

Study Objectives

Research Questions

- **RQ1:** What RAI issues do professionals with different situated perspectives want addressed and why?
- **RQ2:** How do professionals want RAI tools to be designed and implemented and RAI practices to be measured?
- **RQ3:** What stakeholders and human factors are relevant to the design of RAI tools and how they might be cataloged and why?

Desired Outcomes

- Greater accessibility of relevant RAI tools for practitioners wanting such tools to support their practices, especially sole proprietors, startup teams, AI governance teams, AI ethicists, and RAI consultants
- Usability insights for creators of RAI resources and tools, including RAI government agencies, NGOs, and AI governance marketplace players
- Design insights for agencies or organizations who could theoretically pull together a free centralized RAI resources library

Overview of Methods

This study used a multi-step methodology, beginning with the development of a research protocol and securing IRB exempt status for human subjects research. Recruitment targeted U.S.-based professionals involved in AI/ML across various industries, aiming for a diverse participant pool. Participants were recruited through recruiting advertisements posted across multiple online platforms. Data collection involved 12 remote 1x1 Zoom sessions that included interviews with a workshop component, focusing on responsible AI practices. Analysis methods included thematic analysis of survey data, workshop outputs, and transcripts, supported by generative AI tools (ChatGPT and Otter.AI) to enhance accuracy and extract insights. The study prioritized participant privacy, ensuring confidentiality throughout the research process.

Thesis Statement

By considering human factors and situated perspectives of various RAI stakeholders, RAI tools can be designed and cataloged with greater success, resulting in enhanced accessibility and usability for people seeking RAI tools to support RAI practices.

Chapter Two: Background

Related Work

Overview

Research on information-seeking behavior and human factors reveals that cognitive styles, search experience, and even personality can significantly influence how users interact with information systems. The following studies show the need for adaptable, exploratory, personalized interfaces that cater to different perspectives and goals. They also highlight the importance of understanding user contexts and the impact of intrinsic vs. external motivators on user engagement and satisfaction.

Information-Seeking Behavior

In a study that explored how cognitive styles and online search experiences impact search performance and navigational patterns, Kim (2001) suggests that increased online search experience can impact navigational styles and improve search performance. This led to recommendations for enhancing web interfaces and user training programs. Dörk et al. (2011) introduced a new optimistic perspective on information seeking inspired by the leisurely and observant city wanderer, focusing on the personal, engaged nature of interacting with information. Advocating for explorability in the design of information systems, this study highlights the need for better ways of representing levels of details and abstractions. Limberg (1999) examined the relationship between information seeking behaviors and learning outcomes among high school students. The findings of this study showed overlap in the ways students perceived and engaged with information seeking and subject matter, suggesting that variations in information seeking closely correlate with different ways of understanding content.

Human Factors in Information-Seeking

In a paper on situated action models, activity theory, and distributed cognition, Nardi (Ed.). (1996) states, *"It has been recognized that system design will benefit from explicit study of the context in which users work. The unaided individual divorced from a social group and from supporting artifacts is no longer the model user."* Kim's work (2001) clearly demonstrates the impact of human factors (cognitive style and experience with online searches) on user behavior and their interaction with an informational website. Al-Samarraie et al. (2017) studied the impact of personality traits on eye-movement across three information-seeking tasks: factual, exploratory, and interpretive. The results indicated that personality influences how people search for information online. According to Santosa et al. (2005), intrinsic motivators enhance user engagement with an information system whereas situational motivators have a stronger positive impact on user satisfaction, revealing how users assign psychological importance to a system.

The Relevance of Theories to RAI Tool Design and Cataloging

By incorporating insights from these prior works, designers of RAI tools and online libraries can create systems that are more personalizable, intuitive, and engaging. Critical components to a cataloging system include accommodating various levels of search experience with flexibility for the users. The system should also allow exploratory navigation options and support diverse cognitive styles. Designers should also be mindful of the varying contexts in which users exist and how that may impact their search goals, search behaviors, and the tools that will best serve their needs. The ability to personalize the user interface or the RAI tools themselves could improve performance for users. These studies highlight the importance of making information systems adaptable to the cognitive needs and motivations of different users.

Implementing Responsible AI Practices is Tough

Challenges in Implementing Responsible AI in Organizations
There are numerous gaps and societal concerns regarding responsible AI practices in industry (Anagnostou et al., 2022), the urgency of which has been increasingly acknowledged by researchers, technologists, and policymakers (Dignum 2019). In a study by Heger et al. (2020), researchers identified organizational barriers to RAI adoption. They found that despite the toolkits and metrics aimed at AI practitioners, those practitioners were frequently constrained in RAI implementation by organizational culture and leadership priorities, underscoring the need for an RAI maturity model. Similarly, Rakova et al. (2021) have identified organizational structures that support or hinder RAI efforts.

Gaps Between Ethical Principles and Practical Application
Schneiderman (2021) advocated for a combination of AI ethics and human-centered User Experience Design (UXD) to create practical AI applications, bridging the gap between ethics and practice. The article provides a diagram of concentric, contextual rings identifying concrete technical practices at the team level, management strategies at the organization level, and independent oversight at the industry level. Morley et al. (2020) also proposed a tangible approach to RAI practices, having developed a typology to help developers apply ethics principles at each stage of the machine learning (ML) development pipeline.

Ethics-Washing and Superficial Engagement
In a scrutinizing analysis of industry players, de Laat (2021) discusses how "the language of ethics is being instrumentalized for self-serving corporate ends" to avoid the "looming threat of increasing governmental regulation." As ethics becomes a commodity and facilitator in relationships between industry, government agencies, and academia, AI ethics-washing becomes a mask for true intentions (Ochigame 2019).

Synthesis of Problems and Connection to Context
RAI systems and information both require a nuanced understanding of human contexts, needs, and behaviors to create solutions that are usable, effective, and satisfying. Both fields must navigate the intricacies of human factors and organizational structures to achieve successful outcomes. From another angle, an effectively designed information system may be able to support RAI operations, as practitioners and professionals require informational resources and RAI tools to support them in the tangible implementation of these practices.

Proposed Solutions for Effective RAI Tools

Operationalizing Ethical AI Principles
There have been several interventional approaches to addressing the problem of operationalizing AI ethics, such as literature reviews of publicly available AI ethics tools, methods, and research to translate principles into action (Morley et al 2020) and theoretical papers discussing operational needs, such as public trust and product/process support (Zhu et al 2022). There have been practical guidelines for different contexts, such as various levels of governance that are flexible enough for different organizational structures (Schneiderman 2020). Cramer et al. (2018) developed industry-specific guidelines for addressing algorithmic bias in voice UI and recommended best practices in consumer-facing product development.

Human-Centered Design Approaches

Lee et al. (2020) argues that as AI reshapes decision-making in organizations and government, it has become critical to align AI systems with diverse human values and real-world contexts. Lee et al. brings a focus to workshops in research methods. In a related inquiry, Capel and Brereton (2023) explore the meaning of human-centered AI and advocate for increased collaboration between AI and human-computer interaction (HCI) researchers, like Lee et al. Among efforts for human-centered usability to enable RAI practices is the wide-spread creation of toolkits for use within specific or various contexts, such as an interactive toolbox that helps systematically sort, locate, visualize and compare sets of principles based on their contexts (Canca 2020). There have also been usability studies with industry practitioners to assess the effectiveness of some tools, such as ML Fairness toolkits (Richardson et al 2021).

Enhancing Accessibility and Usability

Smith et al. (2023) introduced a card-based AI literacy toolkit to enhance participation of non-technical audiences in an AI co-design workshop, highlighting the importance of involving end-users and non-developers in AI discussions despite their limited technical knowledge. Dominguez & Stoyanovich (2023) similarly address the gap in AI literacy initiatives by examining case studies and advocating for a stakeholder-first approach that opens the conversation up to non-technical stakeholders who may have been only interested in the social implications of AI.

Synthesis of Solutions

The literature on operationalizing ethical AI principles shows a clear shift towards inclusive human-centered approaches. This is integral to bridging the gap between principles and practice. It also connects with the related work on information-seeking behavior and human factors. By meeting the top-down approach (abstract principles and policies) with a bottom-up approach (real-world contexts, human behaviors, and lived experiences), RAI systems can likely be implemented with greater success.

Gaps and Opportunities

Critical Gaps

With the abundance of information on AI, RAI, and ethical AI–and the ever-growing assortment of tools and toolkits–there may be a need for a centralized cataloging system that can provide all RAI stakeholders, technical and non-technical, easy access to the information or tool they need. What would they want in such a cataloging system? What kinds of tools would be useful to them? What do they consider an RAI tool?

Implications for the Study

This study can examine these questions to learn more about what RAI stakeholders (also called practitioners, professionals, and participants) experience in their RAI practices, what kind of supports they want, and what open questions they might have. The prior studies that utilized participatory workshops and several case studies provide promising guideposts for the approach. The related work on information-seeking behavior and cognition will come into play during the analysis of cognitive entry points to RAI practice and tools among stakeholders.

Chapter Three: Methods

Study Design

Procedure

The procedure for this study included developing a research protocol and procuring an exempt status from the Institutional Review Board (IRB) for human subjects research. The research protocol defined the study context, goals, research questions, methods, desired outcomes, risks to participants, and methods for mitigating risks. The research plan defined a participant quota in the interest of equitable recruitment and to inform the design of the recruitment survey. Channels for dispatching the call for participants were identified and a Google Survey form was created through a researcher's secure UT Mail account. An email language template was created in preparation for responding to survey participants who would be selected for 1x1 interviews. Supplemental documents were prepared with additional information about the study and a Calendly link was prepared to enable participants to schedule their sessions. Study instruments were designed to facilitate the interview sessions and participatory workshops. Sessions were conducted remotely via Zoom for no more than 1-hour each. The sessions were divided into four sections: (1) An introduction, (2) gathering background information about the participant, (3) discussing responsible AI in the participant's professional context, and (4) brainstorming RAI tools and resources, prioritizing ideas through the allocation of play money followed by identifying relevant stakeholders that would intersect with those ideas. At the end of each session after recording stopped, participants were asked to provide feedback on the experience, including recruitment, scheduling, and the session. Following the sessions, Zoom recordings and transcripts were downloaded for analysis. Transcripts were edited for accuracy and to remove identifiable or sensitive information. Workshop outputs and transcripts/recordings were analyzed for themes and insights.

Participant Quota

The study populations targeted for recruitment were professionals based in the U.S. that were working in any industry space in which (a) data was collected, labeled or trained for AI/ML purposes, (b) AI models were developed, (c) AI-powered applications were developed, (d) AI-powered services were provided, or (d) AI tools were used for productivity-enhancement purposes. Target participants included people in the functional roles of executive leadership, operations, administration, product, customer service, sales, legal, technical and other. It was expected that roles and role titles of participants would vary. The study welcomed participants from any industry, professional background, and seniority level. In summary, the participant quota considered the following participant characteristics, mostly aiming for an even mix to provide a widely representative sampling for the study:

- U.S. Citizen or Lives/Has Lived in U.S. while working for an organization based in the United States or U.S. Territories
- Participants' industries (any, but particular interest in healthcare, finance, transportation, public sector, and non-profit)
- Degree to which participants experience or directly observe difficulties in translating AI ethics into RAI practice
- Kind of organizations where participants have worked (Data training/labeling, AI model development, AI product development, AI used to provide a service, AI tools for productivity)
- Participants' professions, years of experience, and functional roles within their organizations

Recruitment

The recruitment survey contained 12 questions designed to help the research team meet the participant quota and was built and deployed through Google Survey via the lead researcher's secure UT Mail

account. The language of the recruitment posting included information about the nature of the study, the goals, and disclosed that the study was part of the lead researcher's graduate capstone course. The advertisement graphics were created in Adobe Illustrator by the lead researcher in formats suitable for LinkedIn, Slack, Discord, Facebook, and as an email attachment. The visuals were designed to catch attention with bold colors, simple graphics, concise and informative copywriting, and a clear call to action. The UT logo was included on the cover page to signal the researcher's institution, which was also mentioned in the recruitment advertising language. A posting log was created and maintained to track where the call for participants was posted and the status of each post. The advertisement was distributed on the lead researcher's personal LinkedIn and across 21 LinkedIn groups, 9 Facebook groups, 8 Discord channels, 3 Slack groups, and 11 direct LinkedIn messages within the researcher's professional network. Alt text descriptions were included in the LinkedIn postings. Qualifying survey respondents would be contacted via email with more information and the opportunity to schedule their session. To compensate participants for their time, a short report of the initial findings most relevant to the participant would be offered.

Instruments

This study utilized 8 study instruments:

- Recruiting advertisement language
- Advertisement graphics
- Screener survey
- Email language
- Participant information sheet
- Consent form
- Interview script
- Miro workshop template

Mitigating Risk to Participants

Participants were provided a description of the study, its purposes, the activities they would be asked to participate in within the recruitment advertisement and screener survey, the follow-up scheduling email, the participant information sheet, and in the introduction of the 1x1 sessions. They were given the opportunity to decline to participate within the screener survey, the follow-up scheduling email, and in the introduction of the 1x1 sessions. The process for protecting the privacy of participants and confidentiality of participant data included:

- Creating, deploying and managing the recruitment survey through a researcher's UT Mail Google account
- Informing participants of the collection of their data in the beginning of the recruitment survey and asking for their consent prior to continuing the survey
- Storing identifiable data, including recordings, survey responses and any other collected data on UT devices and clouds, including UT Box, that are only accessible by the research team; Storing only de-identified data on non-UT devices and clouds, such as personal devices.
- Informing participants in the follow-up scheduling email that the research team wished to record their session for internal notetaking purposes only and that all recordings are strictly confidential
- Emailing participants a consent form ahead of their session date, enabling them to indicate if they consented to their session being recorded or not
- Reminding participants about their consent to record at the start of each session and the researcher being prepared to default to taking hand-written notes
- Ensuring personally and organizationally identifiable information was not included in the research findings

The study only included interactions involving survey, interview procedures, and observation of benign behaviors during the interviews and workshops.

Analysis Methods

Survey
Personally identifiable information was removed from survey results prior to analysis. From the survey data, analysis was performed on participant industries and professions, years of experience, the relationships to AI in professional settings, and the perceived frequency of participants having trouble translating RAI policies, principles, and/or guidelines into practice within their professional contexts. This data was cross-analyzed with transcript analysis and workshop outputs to glean contextual insights about human factors that impact RAI practices.

Transcript
Transcripts were saved to the researcher's UT Zoom Cloud account, downloaded and saved to UT Box, and copies were created as Word documents for formatting and easier readability. These were also saved to UT Box. Transcripts were cleaned through comparison with the audio files and personally or organizationally identifiable information was scrubbed. Transcripts were also analyzed by AI tools to facilitate automated summaries, extract key quotes, and help with generating categorization and insights.

Workshop
The Miro workshop outputs included digital sticky notes arranged in thematic groupings and digital play money allocated across the groupings of "new ideas for tangible enablers" of RAI practices, indicating participant prioritizations. The prioritized ideas were then analyzed and thematically grouped to form categories representing what participants believe are the most valuable RAI tools and resources requiring the most funding.

Thematic Analysis
Thematic analysis was performed on the survey data, workshop outputs, and transcripts individually and in cross-analysis.

> **Survey data:** Responses were entered into a table created in the researchers' Notion workspace and reordered according to participant codes. Personally identifiable information and qualifying questions where everyone answered "yes" were removed. This data was analyzed for the represented industries, professions, and/or roles of participants, years of experience in their professions, and for participants' mental models regarding AI as informed by a cross analysis of the kind of AI-related work environments they've experienced vs. their perceived frequency of having trouble in translating RAI principles, policies, and/or guidelines into practice within their professions or professional roles.
>
> **Workshop outputs:** Two groups of sticky notes were generated during sessions regarding the participant's unique context: (1) Tools and resources that the participant thought already worked well in supporting RAI practices; and (2) participant ideas for new tools or resources they'd want for supporting RAI practices. For the second group of new ideas, sticky notes were grouped into themes as determined by the participants during their sessions.
>
> **Transcripts:** Regarding participants' professional backgrounds and familiarity with AI, transcripts were analyzed for supporting quotes and findings were summarized. Regarding RAI in participants' professional practice, transcripts were analyzed for (1) how RAI shows up in participants' professional contexts, (2) what participants think works well now in RAI practices,

(3) what participants seek regarding AI and RAI, and (4) ideas on how to measure RAI practices. For question one, supporting quotes were extracted then summarized and keywords were identified. For question two, supporting quotes were extracted and the key ideas were identified along with keywords and themes. For question three, supporting quotes were extracted and interpreted into brief descriptions and key concepts, the cognitive activity indicated by each quote was identified as either information-seeking, outcome-seeking, RAI tool-seeking, or concern, and the subject and sub-category of the statement was identified. For question four, supporting quotes were extracted and interpreted into brief descriptions, and key ideas were identified.

Cross-analysis: The tables that were created from thematic analyses of survey data, workshop outputs and transcripts were cross-analyzed to find deeper insights on the following questions:

1. Who are the participants and how does this impact the study's results? Why does this matter?
2. How does RAI show up across these professional contexts?
3. What RAI issues do professionals want addressed? How does this differ between professional contexts?
4. What are participants' ideas on how RAI tools and resources should be designed and implemented? How can their impact be made measurable? How does this differ between professional context?
5. What stakeholders and human factors are relevant to the design of RAI tools?
6. How might these human factors translate to the design of an RAI resources cataloging system?

Generative AI Assistance

ChatGPT and Otter.AI were used to support data analysis and writing. Any use of these tools included researcher oversight and verification of accuracy. Transcripts, sections of transcripts, and spreadsheets were uploaded to ChatGPT-4. From transcript information, the lead researcher extracted quotes, keywords, key ideas or themes, and summaries. From spreadsheets, further thematic analysis was performed including cross-analysis between participant backgrounds, workshop outputs, and responses to interview questions. Otter.AI was found to provide more accurate transcription than Zoom with automatic AI-powered summaries and keywords. The tool also provided the ability to visually review the transcripts word for word against the audio, slow down the playback, make edits, and to easily separate and label the speakers. Participants were labeled using their coded unique identifiers. Cleaned transcripts from Otter.AI were downloaded as PDFs for analysis.

Chapter Four: Results

Participant Backgrounds

Overview of Backgrounds

This analysis examines the diverse industries and professional experiences of participants as it relates to AI, focusing on how their familiarity with AI influences their ability to translate RAI principles into practice. The participants come from a diverse range of industries and professions, such as software development, healthcare, and AI ethics to name a few, reflecting the multidisciplinary nature of AI involvement. Experience levels range from early-career to seasoned professionals. The study reveals that those with higher AI familiarity, such as AI engineers and AI researchers, often face significant challenges due to the complexity of the systems they work with. Those with moderate familiarity, like policy strategists, encounter occasional difficulties in adapting RAI frameworks to specific contexts. Participants with lower AI familiarity report fewer challenges, as their roles typically involve less intensive AI engagement. Most participants reported working in multiple industries or at the intersection of different fields, indicating a cross-functional approach to their work with AI. This diversity underscores the varied contexts in which AI is applied and the breadth of expertise required to navigate its ethical and practical challenges. The findings underscore the need for advanced, context-specific tools to support professionals in navigating the ethical complexities of AI.

Participant Industries, Professions and Years of Experience

The recruitment survey gathered information about participants' backgrounds, including any industries and professions wherein they were involved with AI and their years of experience. Most participants reported multiple industries or a cross-section of professions. As shown in **Figure 4.1**, the industries and professions represented by the 14 participants, which includes 2 survey respondents who were not interviewed, included the following: Software and tech development (9); User experience (4); FinTech (3); Healthcare / HealthTech (3); Consulting (3); RAI / AI Ethics (3); Marketing (2); AI Research / AI Engineering (2); SAAS (2); Eco / Environment (1); Human Resources (1); Job Hunting (1); Academic Research (1); and Information Science (1). Survey respondents also reported their years of experience in their professions, resulting in an even distribution across the following categorical buckets: 4 people had 1-4 years of experience; 3 people had 5-9 years of experience; 4 people had 10-19 years of experience; and 3 people had 20 years of experience or more. Due to the vague nature of the survey questions, the years of experience reported do not necessarily correlate with respondents' time engaging with AI within any organization nor with their overall years of work experience.

Relationships to AI and Perceived Frequency of Having Trouble Translating RAI Principles into Practice

In the survey, participants reported which kind of organizations they have worked at according to the relationship to AI. As shown in **Figure 4.2**, Most participants had worked in environments where AI was used to enhance productivity while the fewest participants had worked in environments where AI/ML models were developed. Most participants reported rarely having trouble translating ethical AI principles into practice in the survey. Those that reported having frequent trouble worked in software/tech development, healthcare / healthtech, RAI consulting / AI ethics, and AI research / engineering. Only one participant reported regularly having trouble and that participant worked in cross-sections of software / tech development, fintech, marketing, and HR. Zero participants reported never having trouble.

Fig 4.1 *- Mapping of represented industries/professions of participants and reported years of experience.*

Challenges in Translating RAI Principles into Practice and the Role of AI Relationships

Participants who frequently encounter challenges in implementing RAI practices tend to be in roles where AI decisions carry significant ethical implications, reflecting part of their situated perspectives within complex operational contexts. These roles often require advanced, customized tools and resources that align with specific ethical and operational contexts. A key factor influencing the frequency and nature of these challenges is the participant's relationship to AI, specifically their familiarity with AI concepts.

Analysis reveals no correlation between the type of organization (which indicates place in the AI lifecycle) where participants have worked and their perceived frequency of challenges in translating RAI principles into practice. Participants from organizations heavily involved in AI development or those with stringent regulatory environments, such as tech companies and healthcare, reported more frequent difficulties in translating RAI principles into practice. This suggests that the complexity and ethical stakes in these settings may amplify the challenges of integrating RAI. Conversely, participants from organizations where AI is used more peripherally, such as in consulting or less regulated industries, reported fewer challenges. These findings highlight how organizational context and the role of AI within it significantly influence the ease with which RAI principles are implemented. Overall, the distribution of how participants reported experiencing difficulty in translating ethical AI principles into practice vs. the type of organization where they have worked (indicating place in AI lifecycle) did not show a strong pattern. Instead, it appears that most participants don't experience much trouble translating ethical AI principles into practice, which may be due to their levels of familiarity with AI and the roles they play in their professional contexts.

Participants with a strong familiarity with AI, such as machine learning engineers and AI researchers, reported frequent challenges in translating RAI principles into practice. This could be due to the technical complexity and depth of work they do related to AI systems, which require sophisticated tools and a nuanced understanding of RAI. These professionals frequently encounter obstacles as they navigate the ethical landscapes of AI technologies. Participants in roles that have moderate familiarity with AI, such as policy strategist or ecologist, generally report occasional challenges. They typically rely on existing frameworks and guidelines but struggle with adapting these tools to specific contexts, especially when AI intersects with their broader concerns, such as policy or environmental regulation. Their challenges often arise from the need to bridge the gap between general RAI principles and the unique demands of their specific domains. Those with lower AI familiarity, such as individuals new to AI or in non-technical roles, tend to report fewer direct challenges. For these participants, existing tools and guidelines are often sufficient to navigate the ethical landscape, as their responsibilities involve less complex AI applications.

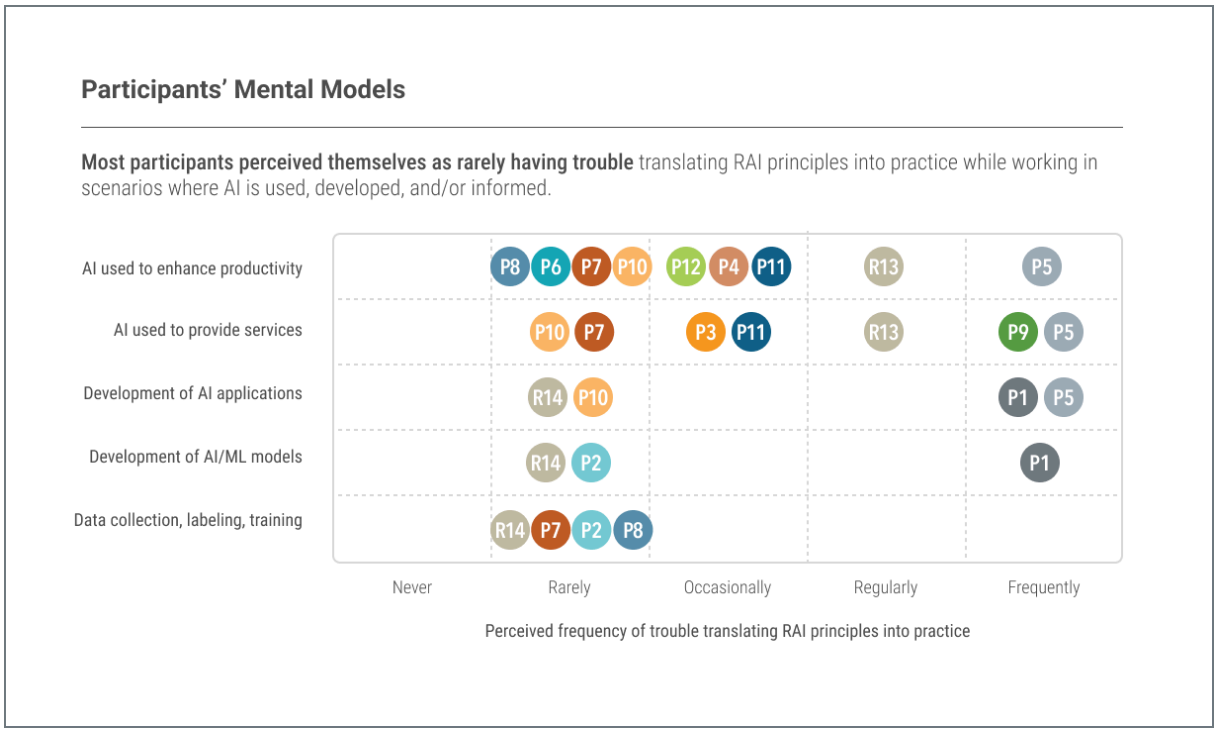


Fig 4.2 *- Mapping of participant relationships to AI suggested by cross-referencing types of organizations where participants reported working (place in AI lifecycle) against how participants reported frequency of having trouble translating RAI principles into practice.*

The challenges of translating RAI practices into action are deeply influenced by the situated perspectives of participants, shaped by both the ethical weight of their roles and their familiarity with AI. As indicated in **Figure 4.3**, those in high-impact, AI-intensive roles encounter the most significant challenges, highlighting the need for more advanced, contextually relevant tools to support their work.

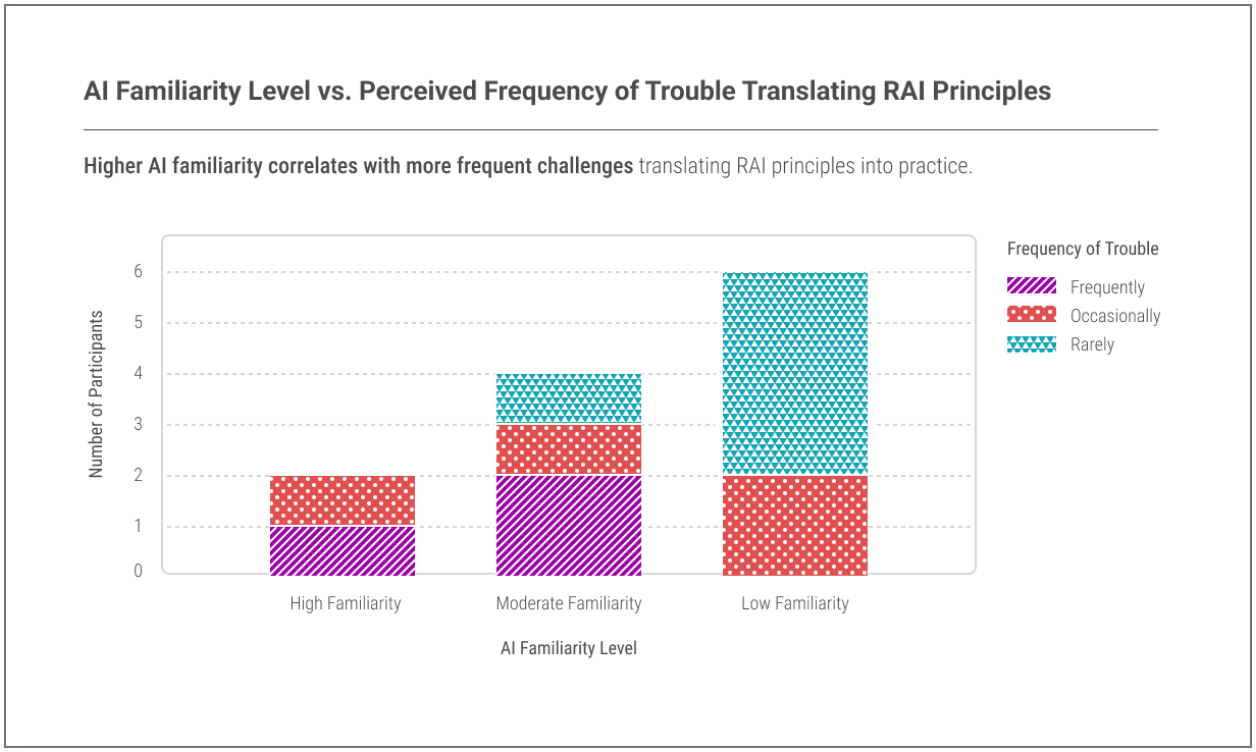


Fig 4.3 *- Categorization of participants' levels of familiarity with AI expressed during interviews vs. perceived frequency of trouble translating RAI principle into practice as reported in the survey.*

Situated Perspectives of AI Stakeholders and the Responsible AI Issues They Want to Address

Overview

The functional roles of participants in the study highlight a diverse range of perspectives on RAI, influenced by their professional backgrounds and the interdisciplinary nature of their work. Participants' professions were categorized into five broad groups: Leadership/entrepreneurship, science/academia, research/ethics, policy/strategy, and engineering/R&D, reflecting their varied contributions to AI. Industry-based roles were also categorized, including techno-ethical, legal-tech, eco-tech, health-tech, and fin-tech roles, each combining specific domain expertise with AI. Additionally, participants' roles were analyzed based on their strategic, operational, technical, and support functions, as well as their involvement with different sectors (public, private, non-profit, academic) and AI lifecycle stages (design, deployment, monitoring, ethics, operations, end-user application). These roles, combined with participants' situated perspectives, informed their concerns and priorities in RAI, particularly around governance, ethical data practices, bias mitigation, and the practical implementation of AI strategies.

Functional Roles and Situated Perspectives

The functional roles of participants, which contribute to informing their situated perspectives on RAI, can be broken down in several ways. This paper examines the following breakdowns: Professions combined into 5 categories, industry-based interdisciplinary roles combined into 5 categories, interdisciplinary roles, strategic vs. operational vs. technical roles vs. support roles, sector-based roles, and roles based on AI lifecycle stages.

Professions of participants can be generally combined into 5 categories:

- Leadership/entrepreneurship, including roles like Startup Founder and UX Director
- Science/Academia, including roles like Ecologist and Informatics Researcher
- Research/Ethics, including roles like Head of AI Research and Ethics
- Policy/Strategy, including roles like Policy Development Strategist
- Engineering/R&D, including roles like ML Engineer and UX Designer

Industry-based interdisciplinary roles of participants:

- **Techno-ethical** roles involve responsibilities that bridge technology and ethics, ensuring that AI development is aligned with ethical principles.
- **Legal-tech** roles combine legal expertise with technical understanding, particularly in areas like intellectual property, privacy, and AI regulation.
- **Eco-tech** roles exist at the intersection of environmental science and technology, where AI is used to help address ecological challenges.
- **Health-tech** roles involve using AI to innovate in medical practices, patient care, and health data management.
- **Fin-tech** roles involve using AI to innovate in financial practices, customer experiences, and financial data management.

Strategic leadership, operational, technical, and support roles of participants:

- **Strategic leadership roles** typically involve participants with over 10 years of experience in their profession, often in higher-level decision-making positions responsible for the strategic direction of AI initiatives.
- **Operational roles** generally involve managing day-to-day operations or those involved in the administrative aspects of AI deployment, ensuring that AI systems are implemented effectively.
- **Technical roles** explicitly involve the development and implementation of AI/ML systems, hands-on work with algorithms, data, and engineering solutions.
- **Support roles** provide auxiliary support, including technical support, training, and documentation for AI systems, facilitating the work of others in AI.

Sector-based roles of participants:

- **Those working in or with the public sector** focus on ensuring that AI technologies align with regulatory frameworks and public policy goals, prioritizing transparency, accountability, and public trust in the deployment of AI systems.
- **Those working in or with the private sector** are primarily concerned with the commercial viability and competitive advantage of AI technologies, balancing innovation and speed-to-market with ethical considerations and regulatory compliance.
- **Those working in or with the non-profit sector** emphasize the ethical use of AI to address societal challenges, often advocating for equitable access to AI technologies and ensuring that AI solutions benefit underserved or marginalized communities.
- **Those working in or with the academic sector** are focused on advancing theoretical and applied research related to AI, contributing to the development of new knowledge, ethical frameworks, and educating future professionals in responsible AI practices.

Participant roles based on AI lifecycle stages:

- **AI design and development:** Participants involved in the conceptualization, design, and technical development of AI systems. Participants working in technology and startups face challenges related to the early stages of AI tool development, emphasizing the need for ethical considerations from the outset.

- **AI deployment and integration:** Those who oversee the integration of AI into existing systems, ensuring that it functions effectively within its intended environment.
- **AI monitoring and evaluation:** Professionals responsible for ongoing monitoring, evaluation, and improvement of AI systems post-deployment.
- **AI ethics and compliance:** Focused on ensuring that AI systems adhere to ethical standards and regulatory compliance throughout their lifecycle.
- **AI operations and support:** Participants who provide operational support for the ongoing maintenance of AI systems, ensuring that AI solutions remain functional, up-to-date, and aligned with organizational needs.
- **AI end-user application:** End-users or operational staff using AI tools and implementing those tools within their specific domain to solve practical problems, focusing on the application of AI outputs in real-world scenarios.

RAI Issues Participants Want Addressed

Participants expressed interest in AI governance, corporate responsibility, data governance, ethical data collection, bias mitigation, and the practical implementation of RAI strategies. Illustrating the range of situated perspectives of AI stakeholders, participants expressed interest across: (1) Technical management and AI governance; (2) Ethical and regulatory considerations; (3) User-centric and practical AI development; (4) Strategic frameworks and implementation; (5) Bias and fairness in AI; and (6) AI in specific professional roles and applications. See **Table 1**.

Themes in RAI Subjects of Interest Among Participants		
Technical Management and AI Governance		
AI Governance	Technical Understanding	Security and Compliance
AI Performance	AI Output Reliability	-
Ethical and Regulatory Considerations		
Ethical AI Implementation	AI Principles Adoption	Data Ethics Information
Transparency and Explainability	Ethics and Regulation	-
User-Centric and Practical AI Development		
User-Centric AI Development	User Engagement	AI Feedback Mechanism
Understanding AI Integration	User Representation in AI	-
Strategic Frameworks and Implementation		
Strategic Frameworks and Methodologies	Implementation Strategy	Organizational Policy Development
Bias and Fairness in AI		
AI Minimizing Expert Input	Bias Mitigation	Ethical Outcomes

AI in Specific Professional Roles and Applications		
AI Ethics in Professional Role	AI in Healthcare Applications	Visualization of Complex Information

Table 1 *- Six emergent themes from subject analysis of AI and RAI issues that participants wanted addressed in their unique contexts.*

Responsible AI Tools and Resources Valued by Participants

Overview

Participants place value transparency, safety, and ethical compliance in RAI practices. They emphasized the importance of clear communication, organizational readiness, and receiving external validation through audits. Among existing tools and resources considered useful in supporting RAI practices, participants valued model cards and API guardrails, effective visual communication aids, clear guidelines, educational initiatives, model validation, and third-party audits. Participants brainstormed new tools they would value, which included AI governance systems, testing and validation mechanisms, and community-driven workshops. Emergent themes from allocation of hypothetical funds revealed the desire for tools that support governance and ethical oversight, rigorous testing, validation and compliance, and customization with context-specific solutions. Participants placed highest priority on (1) establishing robust ethical frameworks, (2) operational tools and process optimization, and (3) training, education and transparency.

Existing Tools and Resources Valued by Participants

Participants value tools and resources that enhance transparency, safety, and ethical compliance in AI practices. These elements are particularly important in contexts where clear communication, organizational readiness, and external validation are crucial. The emphasis on transparency and structured guidance reflects the need for trust and accountability in AI systems, especially in environments where AI decisions have significant ethical and societal implications. The existing tools and resource valued by participants were:

- **Model Cards and API Guardrails** - These tools are appreciated for their role in enhancing transparency and safety in AI development. Particularly in startup environments, where resources might be limited, standardized tools like model cards offer structured guidance. The use of API guardrails also ensures that AI systems operate within defined ethical boundaries, which is crucial in environments where rapid deployment and innovation are common.
- **Effective Communication through Visuals** - In fields where complex data or concepts must be communicated to diverse audiences, visual tools are valued for their clarity and impact. This is particularly important in sectors like science or public communication, where the ability to convey technical information in an accessible way can drive responsible AI practices by ensuring that all stakeholders understand the implications of AI use.
- **Clear Guidelines and Educational Initiatives** - In roles focused on AI ethics and research, clear guidelines are valuable for navigating the complex ethical landscapes of AI. Educational initiatives are also valued for their role in disseminating responsible AI practices across teams and organizations, ensuring that everyone involved in AI development is aligned with ethical standards.
- **Model Validation and Third-Party Audits** - In policy development and legal contexts, validating AI models and employing third-party audits are crucial for ensuring accountability.

These practices help maintain trust in AI systems by providing external verification of their ethical and functional integrity, which is essential in regulating industries or sectors where compliance is a key concern.

- **Organizational Readiness for RAI Principles** - Participants involved in engineering or technical roles emphasize the need for organizational readiness to adopt RAI principles. This involves not just the cultural readiness of an organization but also the infrastructure needed to implement and support responsible AI practices. In fast-paced development environments, ensuring that the organization is prepared to integrate these principles can be a significant challenge. In some cases where AI tools are used for productivity, organizations may block the AI tools until ethical alignment is ensured.
- **Transparency in AI Processes** - Transparency is highly valued by participants working in roles related to research and development. Ensuring transparency in how AI systems operate and make decisions is crucial for building trust with users and other stakeholders. This focus on transparency is particularly important in environments where the ethical implications of AI decisions are closely scrutinized.
- **Community-Driven, Action-Oriented Workshops** - Bringing together communities in workshops or conferences to discuss and implement RAI principles is seen as valuable, particularly by those in academic or collaborative environments. These gatherings provide a platform for sharing best practices, learning from others' experiences, and collectively advancing the state of responsible AI.
- **Tools Providing Sourcing for AI Responses** - Participants who emphasize the importance of data-sourcing in AI outputs typically work in areas where the reliability and credibility of information are paramount. Tools that provide clear sourcing for AI-generated content help ensure that the information is trustworthy and that the AI system is operating transparently.
- **Centralized RAI Resources** - Having a designated place to reference responsible AI practices would greatly aid strategic implementation. This centralization of resources supports consistent application of RAI principles across projects and teams as everyone would have access to the same guidelines and tools. If the resources are organized in a particular way, this system of organization may also help unify the conversation and practice of RAI within an organization.

New Tools Wanted by Participants

Participants brainstormed additional tools and resources they would like to further support them in RAI practices and allocated $1,000 of play money across their ideas. Some of the ideas are already mentioned in the previous section, however in this exercise, each participant was focused on tools and resources that would be new to them in their context. The new tools and resources brainstormed by participants are represented in Table 4.2.

P-Code	Item Identified by Participant	Allocated Cash
P1	AI governance tools to distribute the work across existing tools	$500
	AI testing tools geared towards responsible AI	$400
	AI ethics training platforms	$100
	Concrete operational frameworks and standards (free)	$0
P2	AI-usage peer reviews	$300

	Something to manage and assess self-generating workflows	$250
	Stronger safeguards against shortcomings	$150
	Automated validation of findings / similar cases	$150
	Something to ensure linear validation process is effective and saves money	$150
P3	ISO standards	$500
	Clear regulation with praxis orientation	$250
	Post-implementation literature	$250
P4	A tool that helps synthesize different frameworks	$500
	Grading rubric for companies	$300
	Explainability for impacted stakeholders	$200
P5	Roles to be implemented (hiring of AI ethics consultancy)	$500
	A template we can follow (hiring of AI ethics consultancy)	$250
	Streamlined way to find well-rounded candidates for AI ethics auditing / researcher (hiring of AI ethics consultancy)	$250
P6	A specific feedback system from the AI	$500
	An organization system that ethically holds my information	$250
	A way to know if we're abusing AI	$250
P7	I need a baseline; ChatGPT plugins that identify if data is restricted	$350
	I need a baseline; A way to cross-reference GoogleAI principles against company mission/values to generate our own AI principles	$350
	I need a baseline; Checkpoints/toll gates with what we need to do	$300
P8	Context-specific tools; Evaluative tool for ChatGPT	$600
	Context-specific tools; Reporting	$400
P9	A tool or set of tools for implementation that's specifically for me	$400
	Learning resources	$300
	Something to support community connection and peer mentorship	$200
	Something to support team communications	$100

P10	A set of prompts to cross-check AI assistants	$500
	Community / user outreach and equitable design of RAI solutions	$450
	Ethical workspace templates	$50
	A pledge list (free)	$0
P11	Collaborative framework document	$500
	A way to evaluate the output	$200
	Gen-AI model that lets me input core values, ethics	$150
	Organizational core AI values	$150
P12	User-controlled privacy and safety settings	$300
	A way to assess hallucinations	$300
	External reviews	$150
	Scenario testing / usability testing	$150
	Flags / user generated reports	$100

Table 2 *- The result of participants' brainstorming on what new tools or resources they would like to support them in RAI practices. Each row item represents an idea or a synthesized cluster of ideas as grouped by the participant and includes the amount of play cash allocated to that idea.*

Emergent Themes in Prioritization of Tools and Resources

The following shows emergent themes based on the cash allocations provided by participants, allowing each line item to contribute to more than one thematic category. This flexible analysis reveals a comprehensive approach to responsible AI practices, which will be relevant to the design of an RAI resources cataloging system. The themes that emerged in order of receiving the most allocations were (1) Governance and ethical oversight with $2,350; (2) Testing, validation, and compliance with $1,850; (3) Operational tools and process optimization with $1,700; (4) Community-building and collaboration with $1,450; (5) Training, education, and awareness at $1,350; (6) Transparency and accountability with $1,300; and (7) Customization and context-specific solutions with $1,150. The categories break down as follows:

Governance and ethical oversight ($2,350) - Participants placed the highest value on establishing governance structures and ethical oversight mechanisms that ensure AI practices align with organizational values and industry standards. The significant investment in governance tools and consultancies indicates that participants view these elements as foundational to responsible AI. The items and allocations that support this category include:

- AI governance tools to distribute the work across existing tools ($500 by P1)
- Hiring of an AI ethics consultancy ($1,000 by P5)
- ISO standards ($500 by P3)
- Collaborative framework documents ($500 by P11)

- A way to cross-reference Google AI principles against company mission/values ($350 by P7)

Testing, validation, and compliance ($1,850) - Next, participants prioritized the need for rigorous testing and validation mechanisms to ensure AI models perform ethically and effectively. This includes tools for testing, validation, and adherence to international standards like ISO. The items and allocations represented by this category are:

- AI testing tools geared towards responsible AI ($400 by P1);
- Automated validation of findings / similar cases ($150 by P2);
- Something to ensure linear validation process is effective and saves money ($150 by P2)
- Context-specific tools; Evaluative tool for ChatGPT ($600 by P8)
- Post-implementation literature ($250 by P3)
- External reviews ($150 by P12)
- Scenario testing/usability testing ($150 by P12)

Operational tools and process optimization ($1,700) - Participants want operational tools that can streamline AI processes, ensure ethical management of information, and facilitate the practical implementation of AI governance and ethics. These tools should integrate seamlessly into existing workflows and support responsible AI at an operational level. The items and allocations that support this category are:

- AI governance tools to distribute the work across existing tools ($500 by P1)
- Something to manage and assess self-generating workflows ($250 by P2)
- Concrete operational frameworks and standards, free ($0 by P1)
- A tool or set of tools for implementation that's specifically for me ($400 by P9)
- An organization system that ethically holds my information ($250 by P6)
- Checkpoints / toll gates with what we need to do ($300 by P7)

Community-building and collaboration ($1,450) - Several participants believed that building strong communities and fostering collaboration are important for advancing responsible AI practices. They emphasized the value of tools that facilitate outreach, mentorship, and collaboration, both within organizations and across the broader AI community. The items and allocations that support this category are:

- Community/user outreach and equitable design of RAI solutions ($450 by P10)
- Something to support community connection and peer mentorship ($200 by P9)
- Collaborative framework document ($500 by P11)
- AI-usage peer reviews ($300 by P2)

Training, education, and awareness ($1,350) - Education and training are seen as critical for the successful implementation of RAI principles. Participants would invest in platforms and resources that build awareness and ensure that stakeholders are well-informed about responsible AI practices. The items and allocations to support this category are:

- AI ethics training platforms ($100 by P1)
- Learning resources ($300 by P9)
- Community / user outreach and equitable design of RAI solutions ($450 by P10)
- A set of prompts to cross-check AI assistants ($500 by P10)

Transparency and accountability ($1,300) - Transparency and accountability are critical concerns for participants, who prioritize tools that provide clear explanations to stakeholders, allow users to control privacy settings, and offer mechanisms for reporting and evaluating AI outputs. These tools help build trust and ensure that AI practices are accountable to all stakeholders. The items and allocations that support this category are:

- Explainability for impacted stakeholders ($200 by P4)
- User-controlled privacy and safety settings ($300 by P12)
- Flags / user-generated reports ($100 by P12)
- A way to evaluate the output ($200 by P11)
- Something to support team communications ($100 by P9)
- Context-specific tools; reporting ($400 by P8)

Customization and context-specific solutions ($1,150) - Participants recognized the need for tools that can be tailored to specific contexts and organizational values. Customizable solutions, such as evaluative tools for specific AI applications and templates that align with core ethics, are seen as essential for ensuring that AI practices are both effective and aligned with organizational goals. The items and allocations that support this category are:

- Evaluative tool for ChatGPT ($600 by P8)
- A way to cross-reference Google AI principles against company mission/values ($350 by P7)
- Gen-AI model that lets me input core values, ethics ($150 by P11)
- Ethical workplace templates ($50 by P10)

Correlations with Participant Backgrounds

Participants in policy development or strategic roles seem to value tools that enhance transparency and accountability. They often emphasize the need for clear guidelines, model validation, and third-party audits to ensure that AI systems are ethically compliant and aligned with organizational or governmental standards. Those in R&D or AI ethics research emphasize the importance of educational initiatives and working guidelines. They value resources like model cards, API guardrails, and community-driven workshops that help build a shared understanding of RAI principles across teams. Technical professionals prioritize operational tools and process optimization. They seek tools that can smoothly integrate into workflows, such as AI governance tools and specific feedback systems from AI, to ensure that RAI practices are both practical and effective.

A recurring theme is the desire for tools that can be customized to specific professional contexts. For example, participants want tools that align AI practices with organizational values, frameworks that synthesize different ethical standards, and AI ethics consultants that can help them navigate unique challenges. There is also a strong demand for educational platforms to build awareness of RAI. This is especially important for participants in research, policy, and technical roles, who need to stay informed about the latest in ethical AI.

How AI Stakeholders Want RAI Tools and Resources Designed and Implemented

Overview

The following themes are taken from the top prioritized ideas as indicated by the hypothetical cash allocations. Participants were asked to identify intersecting stakeholders and to consider what might be needed to successfully implement those ideas. Overall, the themes show that participants desire RAI tools that are user-friendly, inclusive, credible, transparent, context-aware, and aligned with organizational strategies and values.

Ease of Use and Accessibility

Three participants emphasize the importance of making RAI tools easy to use, onboard, and understand. They believe tools should be enjoyable, affordable (if not free), and capable of generating comprehensive

reports that cater to various stakeholders. One participant stresses the need for public accessibility to RAI tools and up-to-date standards.

Inclusivity and Collaboration
One participant calls for inclusivity across technical and professional users, ensuring visibility of different team members as well as equitable usability. Two participants suggest tools should encourage collaboration across diverse roles, including top-down and bottom-up organizational involvement. One participant highlights the importance of community support and shared learning in tool development.

Credibility and Standards Compliance
Two participants focus on credibility through AI-usage peer reviews and ISO standards. The success of these tools hinges on maintaining scientific rigor and integrating expert contributions from mature individuals in ethics and governance.

Transparency and Accountability
Two participants stress transparency, particularly in job application processes where AI is involved, and the need for tools that allow cross-referencing AI principles against company values. One participant calls for user-controlled privacy and safety settings, which would allow for responsible adjustments to personal AI parameters by users.

Context-Specific and Evaluative Tools
Two participants emphasize the need for context-specific tools that can evaluate AI-generated results, raising awareness and literacy around potential harms. Two participants focus on tools that provide confidence scores, traceability, and assessments of AI "hallucinations."

Organizational and Strategic Alignment
One participant highlights the strategic importance of AI ethics consultancies, focusing on the financial and reputational risks of non-compliance. Another participant suggests that successful RAI implementation requires organizational alignment, where all levels understand the importance of responsible AI.

P-Code	Prioritized Ideas	Intersecting Stakeholders	Considerations for Success
P1	AI governance tools to distribute the work across existing tools	Product management, data science, ML, software engineers, information security, management or executive teams to some extent for visibility into the tools	Easy to use and onboard, enjoyable, affordable, transparent and gives visibility to different team members, inclusive for technical and professional users, capable of generating comprehensive reports for a variety of stakeholders
P2	AI-usage peer review	Institutions such as universities, professors and colleagues, other scientists and researchers, journal databases and integration of journal publications	Having the credibility and confidence of the state of science; Maintaining credibility without losing the thoroughness required in traditional peer reviews; Using AI to support and enhance the process without replacing the human element; Developing a system akin to "Wikipedia on steroids," where only credentialed or screened individuals can contribute, ensuring expert input
P3	ISO standards	Technical writers are the professional	Successful implementation of

		role that would most interact with standards development and understanding	standards requires a tempered approach to technology and contributions from mature individuals in ethics, public policy, and governance.
P4	A tool that helps synthesize different frameworks, to give us more consensus around which frameworks and standards to follow (and gives clarity on what responsible AI means)	Public policy governance, researchers, general counsel, and data monitoring boards	Public accessibility and up-to-date standards are crucial for the success of AI tools and frameworks
P5	Hiring of 3rd party AI ethics consultancy [to guide us in] roles to be implemented	AI ethics consultancy, the organization	Emphasizing the financial and reputational risks of not following AI ethics can motivate adoption of responsible practices
P6	A specific feedback system from the AI; A way to know if my job application got rejected because I used AI irresponsibly	Recruiters, hiring managers, and job listing services	Transparency in the reasons for job application rejections, offering clarity and understanding for the job seeker
P7	I need a baseline; A way to cross-reference Google AI Principles against company mission/values to generate our own AI principles ChatGPT plugins that identify if data is restricted	Data scientists, engineers, and legal teams	Testing and adjusting based on performance
P8	Context-specific tools; Evaluative tool for ChatGPT (Plugin or feature to evaluate accuracy of AI generated results; Or when using AI tools in data analysis; Or when using ChatGPT for grading)	Developers familiar with AI's capabilities and limitations, administrative regulations are necessary for guiding the responsible use of AI tools	Awareness and literacy regarding AI tools, education about potential harms and cautious use
P9	A tool or set of tools for implementation that's specifically for me; Workbook/worksheet that's short…20 pages, understandable, clear who it's for and what to use it for; Something that lets me focus on a section, ie, transparency; Implementation guide	Various operational, legal, financial, and development teams within startups, including product, engineering, marketing	Collaborative and iterative development, along with community support to facilitate shared learning and application
P10	A set of prompts to cross-check AI assistants to push/check its boundaries; Frameworks for stopgaps	Diversity and ethical consultants, developers	Collaborative efforts, community building, and acknowledgment of contributors
P11	Collaborative framework document; Top-down and bottom-up	Various roles from C-suite to product teams	Success hinges on all organizational levels understanding why responsible AI is important and involving everyone in the development of AI principles
P12	User-controlled privacy and safety settings; A way for users to peel back safety settings in a responsible way; To change the "temperature" of the AI, like with a	Legal, RAI team, product leadership, lots of user researchers, content designers, developers, CIO (awareness)	Providing developers with easy-to-use toolkits and frameworks can significantly enhance the success of AI initiatives.

	knob; For employee-facing bot. This is something my users have asked for. A way to assess hallucinations; Confidence scores, citations, traceability	Developers, product leadership, product designers, researchers, content creators, CIO (awareness)	Developer toolkits, getting developers the tools they need; Development owns the keys in terms of whether we can [implement]

Table 3 *- Aggregation of participants' top prioritized ideas on RAI tools/resources, those they believe to be the intersecting stakeholders for those ideas, and their considerations for successful implementation.*

How AI Stakeholders Think RAI Practices Could be Measured

Six Themes Emerged

Six themes emerged from participants' ideas on how to measure the success of RAI operations. In this result, an idea can appear in more than one theme for maximum flexibility and comprehensive analysis. The themes that emerged were: (1) Documentation and transparency; (2) Compliance and governance; (3) Cultural and organizational shift; (4) Impact and effectiveness; (5) Ethical and equitable outcomes; and (6) Performance monitoring and feedback. See Table 4.4 for the breakdown.

Themes from Participants' Ideas on How/What to Measure in RAI Operations				
Documentation and Transparency				
Implementation of standards and accountability, documentation	Documentation of decisions and processes	Reporting method of conducting research	Detailed reporting of AI usage	Transparency in AI contribution
Specific statistics of AI involvement	Shared understanding and transparency	Third-party evaluation	Feedback into the tool	Transparent communication
Compliance and Governance				
Compliance with laws and regulations	Governance model implementation	Adherence to identified pillars	Regular updates and public accessibility	Benchmarks / certifications
Security systems for underage data	Tracking data breaches	Monitoring algorithm performance	System reboots	-
Cultural and Organizational Shift				
Cultural shift and continuous engagement	Capability to say "no" to clients	Participation and collaboration	Iterative and collaborative development process	Adoption rates
User feedback, addressing product's ethical stances, ability to respond to / integrate suggestions	Open review system	-	-	-
Impact and Effectiveness				

Impact on understanding	Visual representation of data	Accessibility across disciplines	Integrated quant and qual data approaches	Established metrics
Reducing outliers	Reduction in AI hallucinations	Detecting and mitigating bias	Generating diverse and equitable outputs	Evaluation tools for accuracy
Scenario and usability testing	User sentiment surveys	User reviews	-	-
Ethical and Equitable Outcomes				
Percentage of consensual data	Absence of politically oriented content	Diversity and bias feedback	Reporting of harmful content	Overall intent and good intent
Clear understanding of why	-	-	-	-
Performance Monitoring and Feedback				
Consistency in data themes and trends	Sliding scale	Detecting and mitigating bias	User feedback, product's ethical stance, ability to respond to / integrate suggestions	Scenario and usability testing
User sentiment surveys	User reviews	-	-	-

Table 4 *- Themes from aggregation of participants' ideas on how to measure RAI operations.*

Chapter Five: Discussion and Conclusion

Cognitive Points of Entry to RAI Tools Among Professionals

Overview

The findings of this study align with the literature. As in prior research, the situated perspectives and contexts of participants informed how they responded, what they valued, what they wanted to know about AI, and even their level of engagement. Also, just as with prior research, human-centered contexts within organizations have been defined along with tangible processes and outputs within each context (Schneiderman 2021). Professionals and practitioners both with and without a strong technical background generally showed engagement during their sessions, especially during the participatory workshop portion.

Professional Motivations for Engaging with RAI Tools

Professionals engaging with RAI—whether seeking technical understanding of AI, ethical and regulatory information, specific outcomes, or strategic frameworks—are motivated by specific needs that shape their inquiries. Those focused on technical understanding, such as AI developers, seek deep knowledge of AI's mechanisms, requiring a catalog with detailed resources on architecture, security, and system integration. Ethical and regulatory inquiries, mostly driven by policymakers, ethicists, and legal experts, aim to align AI with ethical standards and legal requirements, necessitating up-to-date guidelines and compliance tools. Outcome-seeking professionals, such as consultants, UX professionals, and end-users, prioritize AI's measurable impact, benefiting from resources on best practices, case studies, and performance-tracking tools. Finally, those interested in strategic frameworks, typically senior executives and RAI consultants, focus on integrating AI into broader strategies, needing adaptable resources that bridge technical, ethical, and business perspectives for long-term success.

Seeking Technical Understanding

Professionals driven by technical understanding seek to build knowledge about AI's operational mechanisms, capabilities, and limitations. They are often professionals in technical roles—such as AI developers, data scientists, and engineers—who need to ensure that AI systems are technically sound, secure, and compliant with industry standards. Their motivation is based in mastering the technical aspects of AI to effectively deploy and manage these systems.

This group may assume that a deep technical understanding of AI is crucial for effective RAI implementation. They appear to believe that without this knowledge, they cannot fully leverage AI's potential or mitigate associated risks. They also seem to operate from the belief that AI systems are complex and require detailed knowledge to avoid errors, inefficiencies, or vulnerabilities. These assumptions drive their need for comprehensive, accurate, and up-to-date technical resources.

The implications of this point of entry to RAI resources relate to resource design, search and filter functions, and knowledge-building. The cataloging system should offer detailed technical documentation, tutorials, and guides that cater to various levels of expertise, from beginners to advanced professionals. Resources should cover topics like AI architecture, algorithms, security protocols, and system integration. Users should also be able to easily filter resources by technical complexity, relevance to specific AI technologies, or focus areas such as machine learning models, data management, or AI deployment strategies. And finally, the system should support continuous learning, offering resources that build upon each other to support users as they progress from foundational knowledge to advanced technical expertise.

Seeking Ethical and Regulatory Information

Professionals engaged in ethical and regulatory inquiry are motivated by the need to ensure that AI systems are developed, deployed, and used in ways that are ethically sound and legally compliant. They could be policymakers, ethicists, legal professionals, or organizational leaders responsible for governance. Their motivation stems from a desire to mitigate risks, avoid legal liabilities, protect societal values, and maintain public trust in AI technologies.

These professionals may assume that ethical considerations are fundamental to the responsible use of AI. They may believe that AI should enhance human well-being, or at least not harm human well-being. They may also believe that an AI misstep could lead to significant harm, both to individuals and society at large. They also seem to assume that adherence to legal and regulatory frameworks is not optional but a critical aspect of AI deployment. Non-compliance is seen as a serious risk that could result in fines, legal actions, and damage to an organization's reputation. These professionals assume that the ethical and regulatory landscape for AI is continually evolving, requiring them to stay informed and adaptable. They recognize that as AI technologies advance, new ethical challenges and regulatory requirements will emerge.

The implications of this point of entry for RAI resources can be related to resource design, comprehensive coverage of a subject, practical tools and frameworks, and up-to-date information. The cataloging system should offer a wide range of resources, including ethical guidelines, regulatory updates, case studies, and practical tools that help users navigate the complexities of AI ethics and compliance. Resources should cover various aspects of ethics and regulation, from data privacy and bias mitigation to transparency, accountability, and the ethical implications of AI decisions. The system should include tools for compliance checks, ethics assessments, and risk mitigation strategies that can be directly applied in organizational contexts. Given the dynamic nature of ethics and regulation, the platform should regularly see updated content and resources to reflect the latest developments, ensuring that users can identify and have access to current and relevant information.

Seeking Outcomes

Professionals who are outcome-seeking are primarily concerned with the practical results of AI implementations. This group includes product managers, business strategists, and operational leaders who need to ensure that AI systems deliver value, meet user needs, and achieve specific organizational goals. Their motivation is driven by the desire to see tangible benefits from AI, such as improved efficiency, customer satisfaction, or competitive advantage—all while practicing responsibly.

These professionals may believe that AI should not just be innovative but must also deliver measurable outcomes that align with business or organizational objectives. They believe that AI's value is ultimately demonstrated through its impact on performance, productivity, and user experience. They may also assume that the success of AI systems depends on how well they serve the end-users, whether those users are customers, employees, or other stakeholders. This fuels their focus on usability, accessibility, and user satisfaction.

The implications of this point of entry to RAI resources can be related to resource alignment, case studies/success stories, and outcome metrics. The cataloging system should provide resources that focus on best practices for achieving specific outcomes with AI, such as user-centric design, performance optimization, and impact measurement. By providing real-world examples where AI has successfully delivered on its promises while adhering to RAI principles, the cataloging system could provide users with models to emulate and "lessons learned." By including tools and resources that help users measure and track the outcomes of their AI projects, the cataloging system could help ensure they can assess the effectiveness and impact of their implementations.

Seeking Strategic Frameworks and Methodologies
Professionals interested in strategic frameworks and methodologies are focused on the long-term planning and systematic implementation of AI and RAI. This group includes people in leadership or strategy positions who are integrating AI/RAI into broader business strategies and need to ensure its alignment with organizational goals and values. They are motivated by the need for structured processes that can guide RAI initiatives from conception to execution to long-term adaptability.

These professionals may assume that AI should be integrated into the organization's overall strategy and not treated as a standalone or isolated initiative. They believe that AI's success is contingent on its alignment with broader organizational objectives and its ability to drive innovation and transformation. They assume that the successful implementation of AI requires rigorous methodologies and frameworks that can be consistently applied across projects and teams. This drives their interest in systematic approaches that reduce risks and increase the chances of success.

The implications of this point of entry to RAI resources can be related to the kind of resources available in the cataloging system and the scalability and adaptability of those resources. The catalog should feature strategic planning frameworks, implementation toolkits, and methodology guides that help organizations incorporate AI into their overall strategy. The cataloging system should provide resources that bridge technical, ethical, and business perspectives, supporting the holistic integration of AI within organizations. Provided resources should allow organizations to tailor frameworks and methodologies to their specific context, industry, and scale.

Human Factors x RAI Factors
Similar to the rungs of the human-tech ladder (Vicente 2003), human factors relevant to the design of a centralized RAI resources online cataloging system include the following:

- Political region (e.g. U.S. vs. E.U.)
- Sector (for-profit, non-profit, government, academia)
- Industry (businesses/practices, relationship to AI lifecycle, regulations)
- Profession (expertise, years of experience, relationship to AI lifecycle, code of conduct/ethics)
- Organization (size, age/maturity, business, relationship to AI lifecycle, leadership objectives, organizational culture, organizational structure)
- Team (size, hierarchy, goals, your role, culture, intersecting or collaborative stakeholders)
- Individual (cognition, behavior, physical)
- Change over time

Influence on Future Research, Policy, or Practice
The findings of this study may influence research methods, sparking greater interdisciplinary approaches to the study of RAI in praxis. As there is a quality assurance aspect to the centralized RAI resource library concept, there is the possibility that future phases of this work will require collaboration with policy bodies, such as NIST or the U.S. Department of State. The intent of this study is to have a positive impact on RAI practice by enabling professionals and practitioners of all kinds to have easier access to highly relevant and quality RAI tools.

Limitations and Future Work

Acknowledgement of Limitations of Study's Approach and Data
This study began as a broad research topic which required extensive literature review. Due to ongoing literature review from November 2023 to July 2024, and after analyzing participant data, research goals evolved after the participant study was completed. Given the vague and sensitive nature of the study

combined with time constraints, it was decided to keep the participant quota broad to ensure recruitment of at least 5 participants. This may have resulted in a sampling biased towards UX professionals, AI ethics / RAI professionals, and professionals within the researcher's personal network. Because many participants were already RAI advocates, it may have been easier to recruit them. AI developers were noticeably absent from the participant pool. Participants also skewed towards startups and sole proprietor consultants. Lack of distinction between RAI and AI ethics during the sessions contributed to participants feeling disoriented and giving vague responses to some questions. The interview protocol changed after the first two sessions, as participants needed a strong introduction to the context of the study. Small improvements to execution of the sessions were made throughout the 12 interviews, as the researcher gathered feedback from each participant at the end of their session. Overall, vague concepts and phrasing made it difficult for participants to respond with confidence.

Challenges Faced During the Research Process

Evolving study goals: The research gap shifted as I continued my literature review after the study was completed, resulting in a slightly different focus. Not knowing what the research goals are from the beginning creates challenges in study design and analysis.

Critical incident vs. confidentiality: Despite the intention of using critical incident technique, wanting to respect organizational confidentiality and participants' possible unease in discussing their employer contributed to the formation of vague questions during interviews.

RAI competitive advantage vs. academic study goals: While one of the study goals was to learn about existing RAI practices for professionals and the gaps they may be experiencing, sharing this information potentially posed a threat to the participants' organizational RAI competitive advantage. It was important to respect informational boundaries.

A ton of data to analyze: The study surpassed its recruitment goal by more than 2x and collected a substantial amount of mostly qualitative data. The process of reviewing and cleaning transcripts against twelve hours of recorded audio, extracting data into tables, analyzing, and cross-analyzing took significantly more time than anticipated. Although not initially planned, this analysis required assistance from AI tools, as mentioned in the Methods chapter of this paper. Even with this assistance, analysis was done carefully and methodically, cell by cell, table by table. In the future, large amounts of qualitative data may call for more advanced methods of analysis.

Future Work

Future work will include a more complete literature review and extensive competitive research on RAI resource libraries and RAI tools. It will also include proposed criteria for a centralized online RAI resources library, including: (1) Persona groups with detailed empathy maps; (2) User journey mapping for identified key user groups; (3) Description of key system affordances; (4) Proposed classification system for the subject: "RAI resources"; (5) Proposed schema for RAI resource records; (6) Early user testing of these artifacts (concept only, no build); (7) Interviews with possible "owner" entities (government, NGOs, academia) of a centralized online RAI resources library to gauge interest and project feasibility.

Summarization of Key Findings

Participants desire RAI tools that are user-friendly, inclusive, credible, transparent, context-aware, and aligned with organizational strategies and values. Despite the abundance of RAI tools and frameworks available, participants struggle to determine which would be appropriate for their unique professional and

organizational contexts. There is a strong demand for standards in AI governance and clear guidance from their industries or organizations. There is also a strong need for tailored and personalized solutions and better coordination in RAI practices, preferably with the help of a third party RAI consultancy.

Contributions to the Field

These findings will provide valuable insights for making RAI resources more accessible, useful, and usable for professionals aiming to enhance their RAI practices. Additionally, this paper contributes ideas on the design of a centralized online RAI resources cataloging system and actionable next steps.

Closing Remarks

Should adherence to RAI standards be part of competitive advantage? What would it mean for the AI industry and those impacted by AI if only large companies can afford an RAI consultant? In considering the experience of the tech start-up, a final statement from one of the participants:

> *"What I've seen is just that kind of…beginnings of, 'we're interested in this, we know we should be doing it. We want to know what framework to do and we want to make sure we can do it in a way that isn't burdensome, that doesn't stop our innovation.'"*

Works Cited

Al-Samarraie, H., Eldenfria, A., & Dawoud, H. (2017). The impact of personality traits on users' information-seeking behavior. Information Processing & Management, 53(1), 237-247.

Anagnostou, M., Karvounidou, O., Katritzidaki, C., Kechagia, C., Melidou, K., Mpeza, E., ... & Peristeras, V. (2022). Characteristics and challenges in the industries towards responsible AI: a systematic literature review. Ethics and Information Technology, 24(3), 37.

Benjamins, R., Barbado, A., & Sierra, D. (2019). Responsible AI by design in practice. arXiv preprint arXiv:1909.12838.

Berman, G., Goyal, N., & Madaio, M. (2024, May). A Scoping Study of Evaluation Practices for Responsible AI Tools: Steps Towards Effectiveness Evaluations. In Proceedings of the CHI Conference on Human Factors in Computing Systems (pp. 1-24).

Boyd, K. (2022, June). Designing up with value-sensitive design: Building a field guide for ethical ML development. In Proceedings of the 2022 ACM conference on fairness, accountability, and transparency (pp. 2069-2082).

Canca, C. (2020). Operationalizing AI ethics principles. Communications of the ACM, 63(12), 18-21.

Capel, T., & Brereton, M. (2023, April). 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 (pp. 1-23).

Cho, S. H., Jon, S., Jin, Y., Jung, J., & Oh, C. (2024, May). Understanding the Dynamics in Creating Domain-Specific AI Design Guidelines: A Case Study of a Leading Digital Finance Company in South Korea. In Extended Abstracts of the CHI Conference on Human Factors in Computing Systems (pp. 1-7).

Cramer, H., Garcia-Gathright, J., Springer, A., & Reddy, S. (2018). Assessing and addressing algorithmic bias in practice. Interactions, 25(6), 58-63.

de Laat, P. B. (2021). Companies committed to responsible AI: From principles towards implementation and regulation?. Philosophy & technology, 34(4), 1135-1193.

Deshpande, A., & Sharp, H. (2022, July). Responsible ai systems: who are the stakeholders?. In Proceedings of the 2022 AAAI/ACM Conference on AI, Ethics, and Society (pp. 227-236).

Dignum, V. (2019). Responsible artificial intelligence: how to develop and use AI in a responsible way (Vol. 2156). Cham: Springer.

Dignum, V. (2023, January). Responsible Artificial Intelligence-From Principles to Practice: A Keynote at TheWebConf 2022. In ACM SIGIR Forum (Vol. 56, No. 1, pp. 1-6). New York, NY, USA: ACM.

Domínguez Figaredo, D., & Stoyanovich, J. (2023). Responsible AI literacy: A stakeholder-first approach. Big Data & Society, 10(2), 20539517231219958.

Dörk, M., Carpendale, S., & Williamson, C. (2011, May). The information flaneur: A fresh look at information seeking. In Proceedings of the SIGCHI conference on human factors in computing systems (pp. 1215-1224).

Hadfield, G. K., & Clark, J. (2023). Regulatory markets: The future of ai governance. arXiv preprint arXiv:2304.04914.

Heger, A., Passi, S., & Vorvoreanu, M. (2020). All the tools, none of the motivation: Organizational culture and barriers to responsible AI work. Cultures in AI.

Jacobs, A. Z. (2021). Measurement as governance in and for responsible AI. arXiv preprint arXiv:2109.05658.

Kafai, Y. B., Proctor, C., Cai, S., Castro, F., Delaney, V., DesPortes, K., ... & Rosé, C. P. (2024). What Does it Mean to be Literate in the Time of AI? Different Perspectives on Learning and Teaching AI Literacies in K-12 Education. In Proceedings of the 18th International Conference of the Learning Sciences-ICLS 2024, pp. 1856-1862. International Society of the Learning Sciences.

Khurana, A., Subramonyam, H., & Chilana, P. K. (2024, March). Why and when llm-based assistants can go wrong: Investigating the effectiveness of prompt-based interactions for software help-seeking. In Proceedings of the 29th International Conference on Intelligent User Interfaces (pp. 288-303).

Kim, K. S. (2001). Implications of user characteristics in information seeking on the World Wide Web. International Journal of Human-Computer Interaction, 13(3), 323-340.

Lee, M. K., Grgić-Hlača, N., Tschantz, M. C., Binns, R., Weller, A., Carney, M., & Inkpen, K. (2020, April). Human-centered approaches to fair and responsible AI. In Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems (pp. 1-8).

Limberg, L. (1999). Three conceptions of information seeking and use. Exploring the contexts of information behaviour, 116-135.

Sadek, M., Constantinides, M., Quercia, D., & Mougenot, C. (2024, May). Guidelines for Integrating Value Sensitive Design in Responsible AI Toolkits. In Proceedings of the

CHI Conference on Human Factors in Computing Systems (pp. 1-20).

Mikalef, P., Conboy, K., Lundström, J. E., & Popovič, A. (2022). Thinking responsibly about responsible AI and 'the dark side'of AI. European Journal of Information Systems, 31(3), 257-268.

Moggridge, B., & Atkinson, B. (2007). Designing interactions (Vol. 17). Cambridge: MIT press.

Morley, J., Floridi, L., Kinsey, L., & Elhalal, A. (2020). From what to how: an initial review of publicly available AI ethics tools, methods and research to translate principles into practices. Science and engineering ethics, 26(4), 2141-2168.

Munir, W. (2014). Saudi Health 2014 presentation human factors [PowerPoint slides]. SlideShare. Retrieved July 29, 2024, from https://www.slideshare.net/slideshow/saudi-health-2014-presentation-human-factors-34953823/34953823

Nardi, B. A. (Ed.). (1996). Context and consciousness: Activity theory and human-computer interaction. MIT Press.

Ochigame,R. (2019). The invention of 'ethical AI': How big tech manipulates academia to avoid regulation. Economies of virtue,49.,Ochigame R. (2019). The invention of 'ethical AI': How big tech manipulates academia to avoid regulation. Economies of virtue 49.

Olatoye, F. O., Awonuga, K. F., Mhlongo, N. Z., Ibeh, C. V., Elufioye, O. A., & Ndubuisi, N. L. (2024). AI and ethics in business: A comprehensive review of responsible AI practices and corporate responsibility. International Journal of Science and Research Archive, 11(1), 1433-1443.

Peters, D., Vold, K., Robinson, D., & Calvo, R. A. (2020). Responsible AI—two frameworks for ethical design practice. IEEE Transactions on Technology and Society, 1(1), 34-47.

Rakova, B., Yang, J., Cramer, H., & Chowdhury, R. (2021). Where responsible AI meets reality: Practitioner perspectives on enablers for shifting organizational practices. Proceedings of the ACM on Human-Computer Interaction, 5(CSCW1), 1-23.

Richardson, B., Garcia-Gathright, J., Way, S. F., Thom, J., & Cramer, H. (2021, May). Towards fairness in practice: A practitioner-oriented rubric for evaluating Fair ML Toolkits. In Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems (pp. 1-13).

Santosa, P. I., Wei, K. K., & Chan, H. C. (2005). User involvement and user satisfaction with information-seeking activity. European Journal of Information Systems, 14(4),

361-370.

Schiff, D., Rakova, B., Ayesh, A., Fanti, A., & Lennon, M. (2020). Principles to practices for responsible AI: closing the gap. arXiv preprint arXiv:2006.04707.

Shneiderman, B. (2021). Responsible AI: Bridging from ethics to practice. Communications of the ACM, 64(8), 32-35.

Smith, F., Sadek, M., & Mougenot, C. (2023, August). Empowering End-users in Co-Designing AI: An AI Literacy Card-Based Toolkit for Non-Technical Audiences. In 36th International BCS Human-Computer Interaction Conference (pp. 13-22). BCS Learning & Development.

Soklaski, R., Goodwin, J., Brown, O., Yee, M., & Matterer, J. (2022). Tools and practices for responsible AI engineering. arXiv preprint arXiv:2201.05647.

Subramonyam, H., Pondoc, C. L., Seifert, C., Agrawala, M., & Pea, R. (2023). Bridging the Gulf of Envisioning: Cognitive Design Challenges in LLM Interfaces. arXiv preprint arXiv:2309.14459.

Sun, J., Liao, Q. V., Muller, M., Agarwal, M., Houde, S., Talamadupula, K., & Weisz, J. D. (2022, March). Investigating explainability of generative AI for code through scenario-based design. In Proceedings of the 27th International Conference on Intelligent User Interfaces (pp. 212-228).

Tankelevitch, L., Kewenig, V., Simkute, A., Scott, A. E., Sarkar, A., Sellen, A., & Rintel, S. (2024, May). The metacognitive demands and opportunities of generative AI. In Proceedings of the CHI Conference on Human Factors in Computing Systems (pp. 1-24).

van Maanen, G. (2022). AI ethics, ethics washing, and the need to politicize data ethics. Digital Society, 1(2), 9.

Wallach, W., & Allen, C. (2008). Moral machines: Teaching robots right from wrong. Oxford University Press.

Wang, Y., Xiong, M., & Olya, H. (2020, January). Toward an understanding of responsible artificial intelligence practices. In Proceedings of the 53rd Hawaii International Conference on System Sciences (pp. 4962-4971). Hawaii International Conference on System Sciences (HICSS).

Wang, Q., Madaio, M., Kane, S., Kapania, S., Terry, M., & Wilcox, L. (2023, April). Designing responsible ai: Adaptations of ux practice to meet responsible ai challenges. In Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems (pp. 1-16).

Zhu, L., Xu, X., Lu, Q., Governatori, G., & Whittle, J. (2022). AI and ethics—Operationalizing

responsible AI. Humanity driven AI: Productivity, well-being, sustainability and partnership, 15-33.

Appendix A: Recruitment Language

I NEED YOUR VOICE (call for participants in ALL INDUSTRIES AND ROLES): Do you personally struggle with closing the gap between abstract responsible AI policies/principles and performing "responsible AI" in the day-to-day operations of your **professional context?** Would you like insights from **professionals like you?**

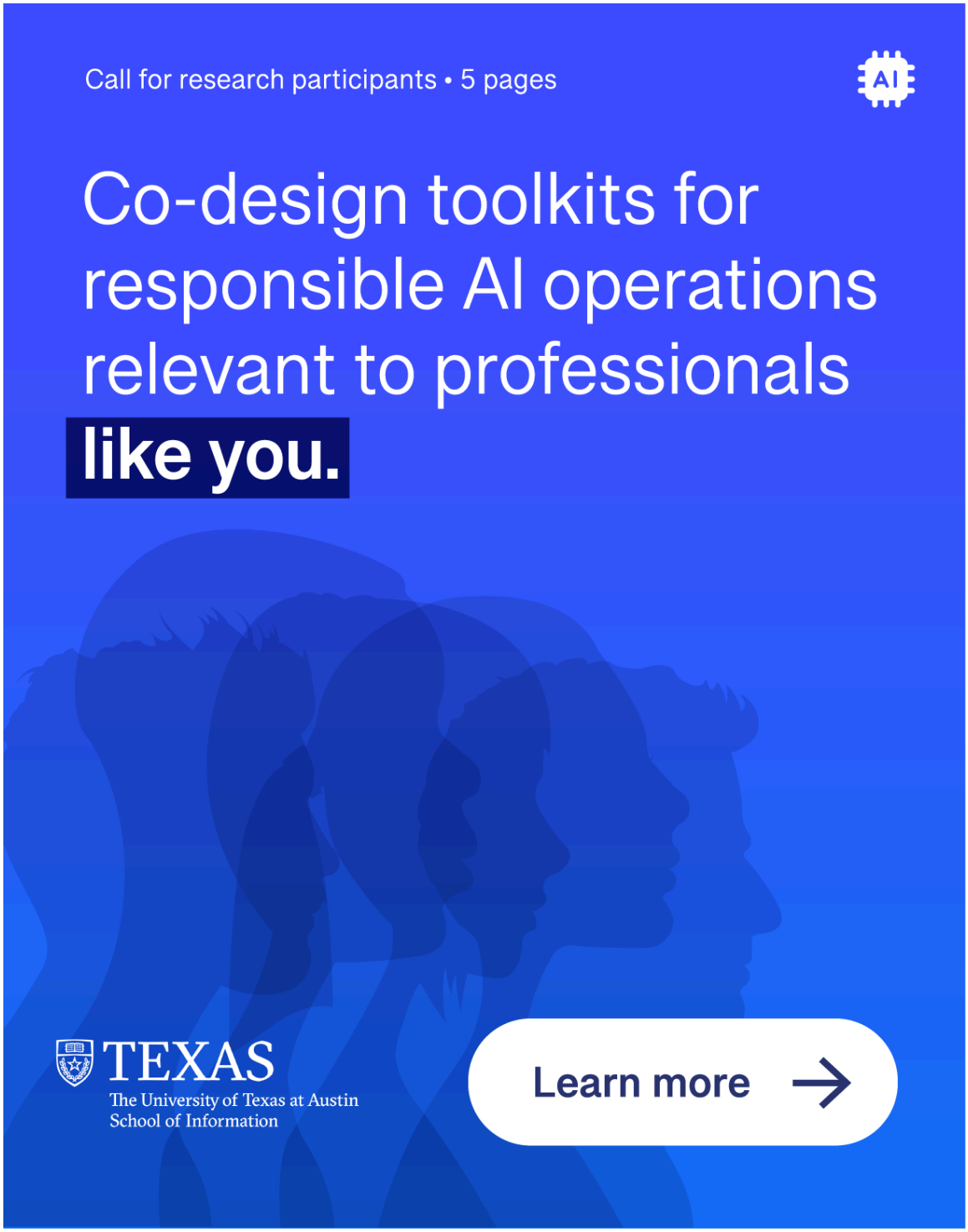
My name is Vanessa Sánchez, and I am a graduate student in the School of Information at The University of Texas at Austin. My research interests are in responsible AI, particularly in understanding the challenges professionals face in operationalizing responsible AI.

For my capstone this Spring, I'm conducting research to learn how professionals in their unique contexts can be better supported and enabled through tangible tools, materials, and assets. I am looking for **U.S.-based research participants** for a **remote 1-hour one-on-one interview/workshop**.

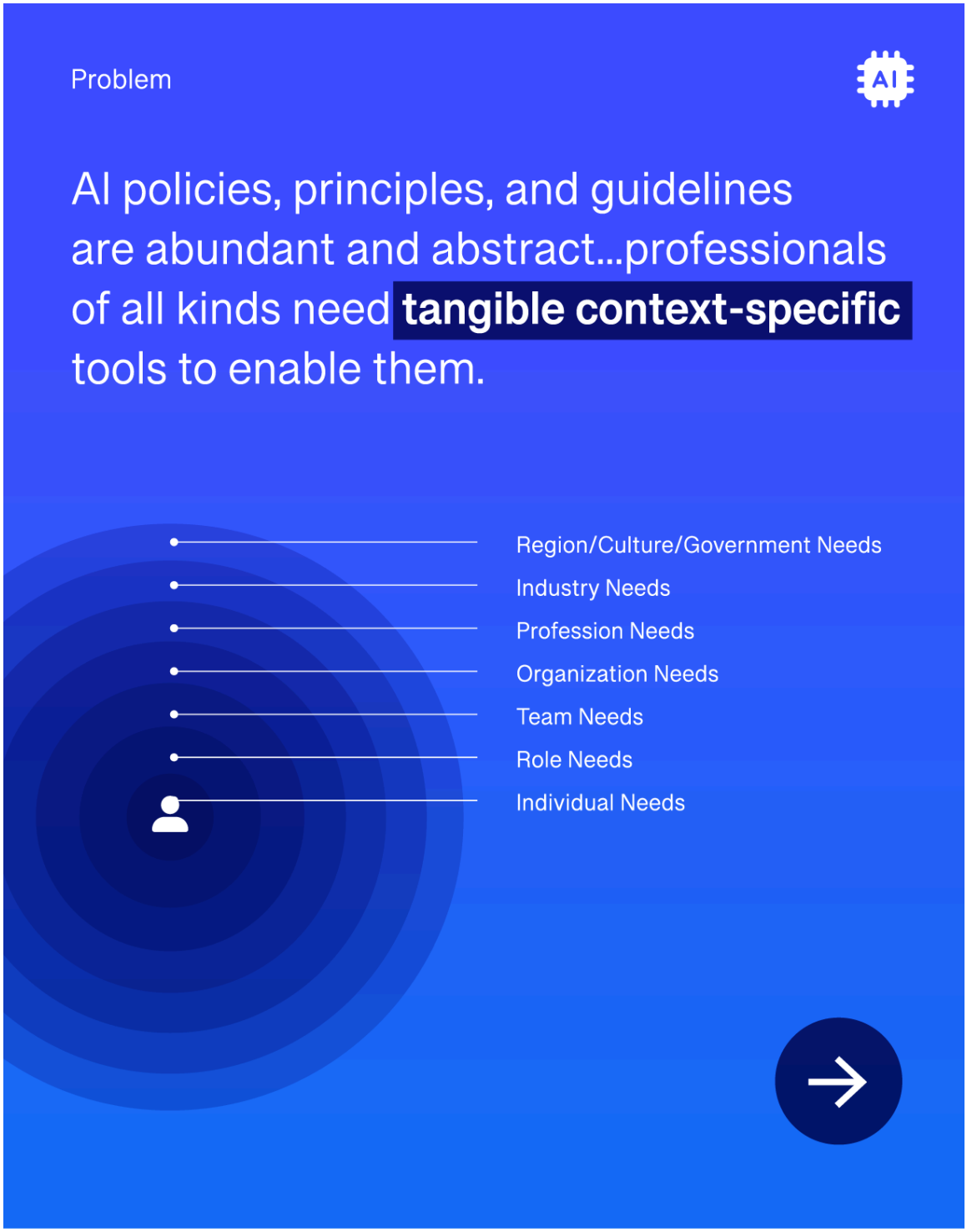
Participants will gain **strategic ideas** on how to tangibly advance responsible AI operations from within their professional role while **contributing to insights** for their profession or professional role. *All personally and organizationally identifiable information will be protected in the findings. All collected data will be securely stored on UT devices and UT cloud accounts and will only be accessible by the research team.*

If you are interested in participating, please fill out this 3-minute survey**:** https://forms.gle/72srk8rh6pnNmux9A. Thanks, and we look forward to co-designing with you!

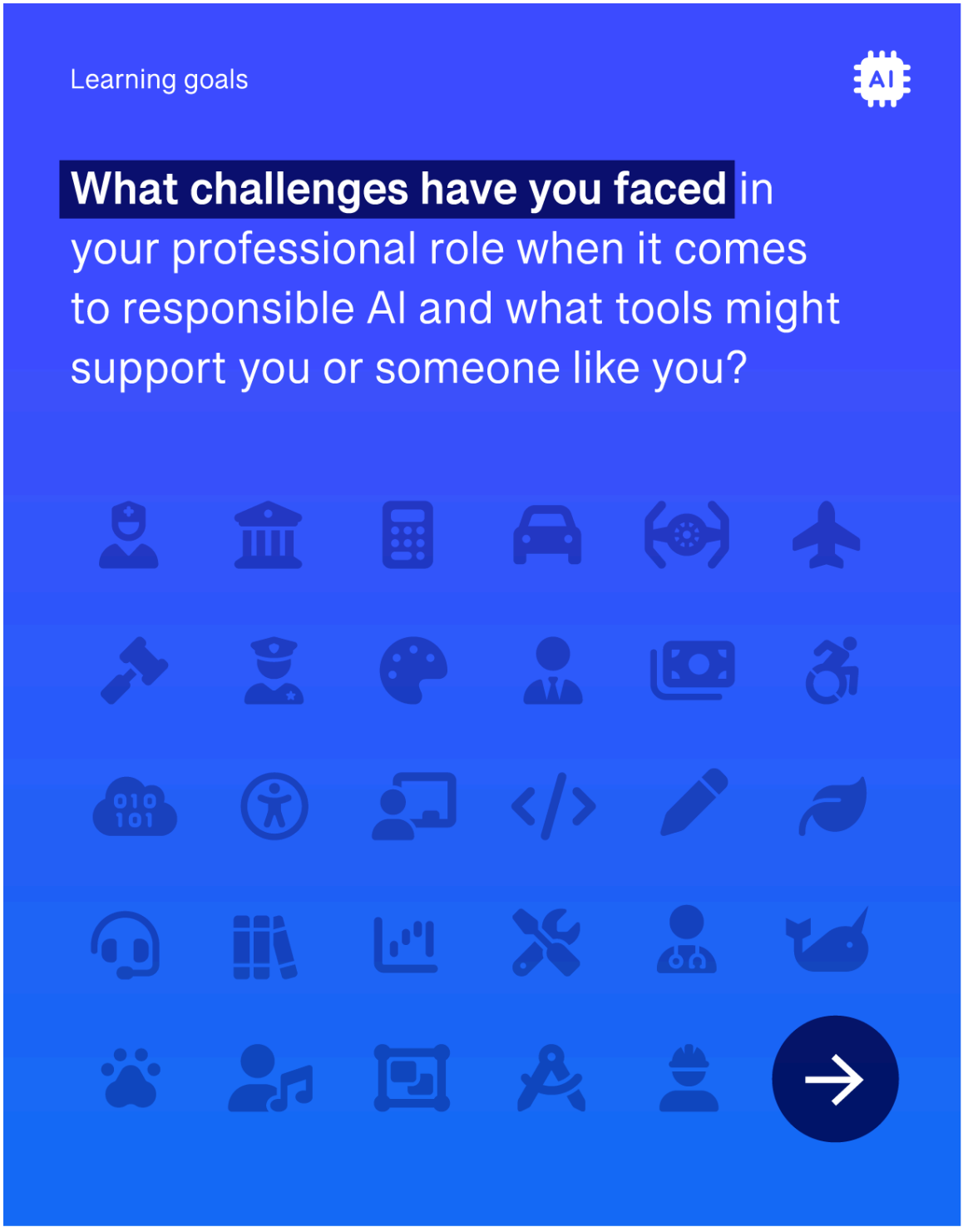
Appendix B: Recruitment Graphics



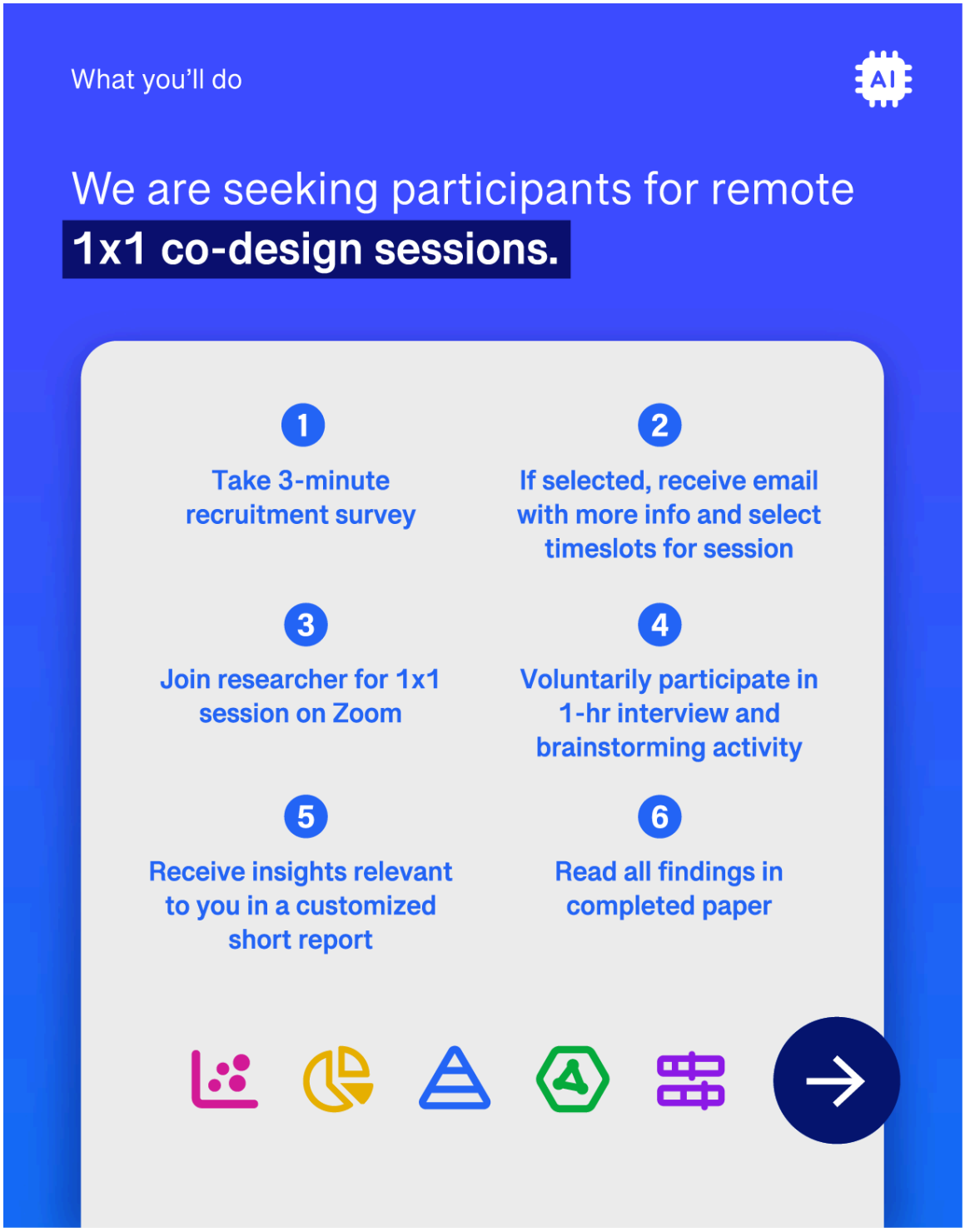
ALT TEXT: Call for research participants, 5 pages. Co-design toolkits for responsible AI operations relevant to professionals like you. The University of Texas at Austin School of Information. Learn more.



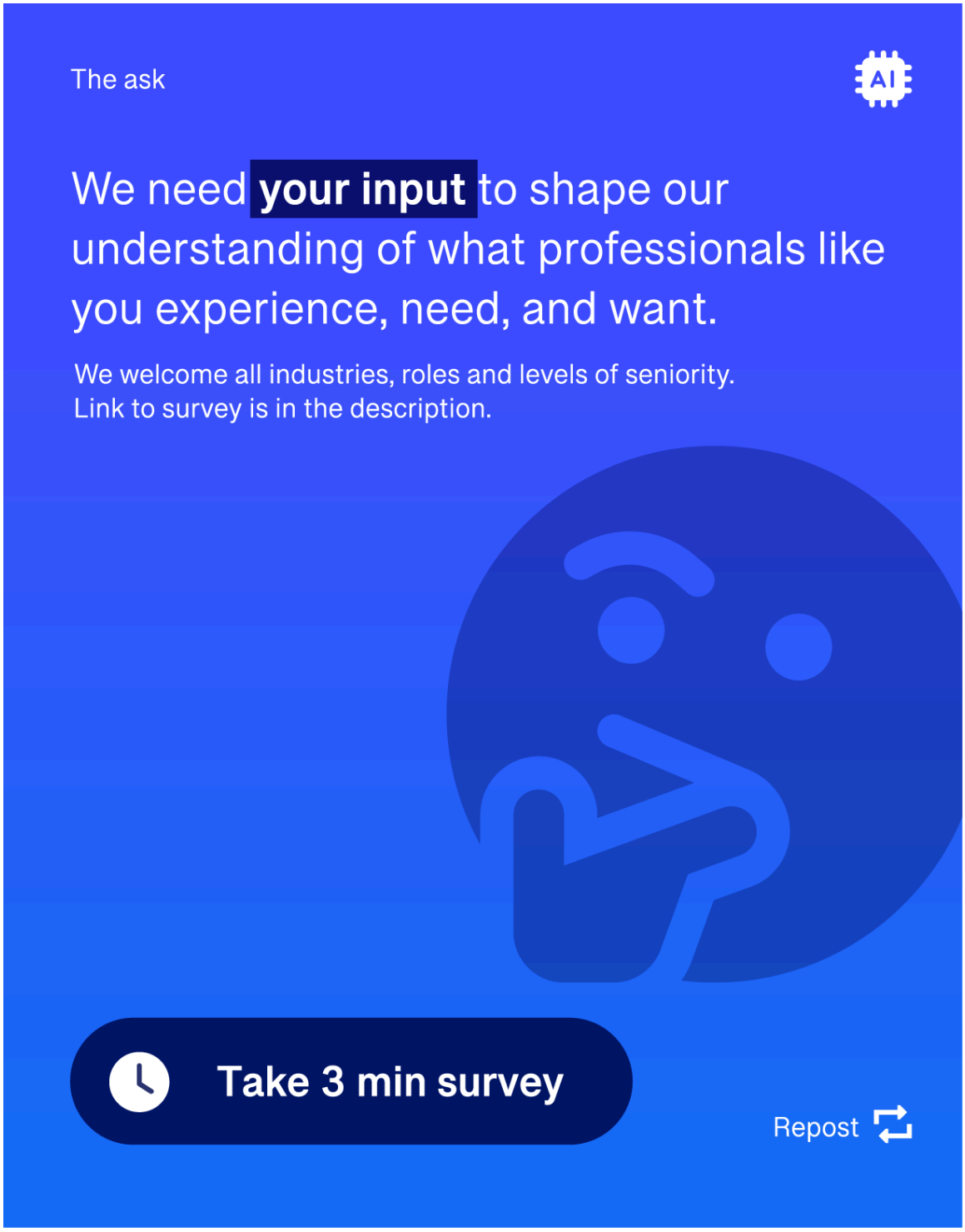
ALT TEXT: Problem: AI policies, principles, and guidelines are abundant and abstract. Professionals of all kinds need tangible context-specific tools to enable them.



ALT TEXT: Learning goals: What challenges have you faced in your professional role when it comes to responsible AI and what tools might support you or someone like you?



ALT TEXT: What you'll do: We are seeking participants for remote 1 on 1 co-design sessions. Step 1 - Take 3-minute recruitment survey. Step 2 - If selected, receive email with more information and select time slots for session. Step 3 - Join researcher for 1 on 1 session on Zoom. Step 4 - Voluntarily participate in 1-hour interview and brainstorming activity. Step 5 - Receive insights relevant to you in a customized short report. Step 6 - Read all findings in completed paper.



ALT TEXT: The ask: We need your input to shape our understanding of what professionals like you experience, need, and want. We welcome all industries, roles and levels of seniority. Link to survey is in the description. Take 3 minute survey. Repost.

Appendix C: Recruitment Survey

1) This survey will collect your responses and store that information in a secure UT cloud account and UT device only accessible by the research team. Personally identifiable information will NOT be revealed in any research findings. If you have questions you can contact the lead researcher at vanessa.sanchez@utexas.edu. Participation in this survey is voluntary. Do you consent and wish to proceed?	
Yes	**[CONTINUE]**
No	**[TERMINATE]**
2) Are you a U.S. citizen or living in the U.S.?	
Yes	**[CONTINUE]**
No	**[CONTINUE]**
3) Have you ever worked at a U.S. organization that collects, labels and trains data; develops ML models/algorithms; builds AI applications; or uses AI to provide services or enhance productivity?	
Yes	**[CONTINUE]**
No	**[TERMINATE]**
4) Have you directly had trouble in translating responsible AI policies, principles, and/or guidelines into practice/action within your profession or professional role?	
Regularly	**[CONTINUE]**
Frequently	**[CONTINUE]**
Occasionally	**[CONTINUE]**
Rarely	**[CONTINUE]**
Never	**[TERMINATE]**
5) Would you be willing and able to participate in a remote one-hour 1x1 interview and workshop via Zoom sometime in the next 2 weeks?	
Yes	**[CONTINUE]**
No	**[TERMINATE]**
6) What is your profession? (This is not necessarily your job title…feel free to use a general term)	
[SHORT ANSWER]	**[CONTINUE]**

7) What is your experience level within your profession?	
20+ years	**[CONTINUE]**
10+ years	**[CONTINUE]**
5+ years	**[CONTINUE]**
2+ years	**[CONTINUE]**
1+ years	**[CONTINUE]**
0-1 years	**[CONTINUE]**
8) What is your industry? (You can list more than one)	
[SHORT ANSWER]	**[CONTINUE]**
9) At which kind of U.S. organization(s) have you worked? (Select all that apply)	
Data collection, labeling, and/or training	**[CONTINUE]**
Development of ML model(s) / algorithm(s)	**[CONTINUE]**
Development of AI application(s)	**[CONTINUE]**
AI used to provide services	**[CONTINUE]**
AI used to enhance productivity	**[CONTINUE]**
10) What is the email where we can most easily reach you?	
[SHORT ANSWER]	**[CONTINUE]**
11) Please confirm your email.	
[SHORT ANSWER]	**[CONTINUE]**
12) What is your first name?	
[SHORT ANSWER]	**[CONTINUE]**

Appendix D: Scheduling Email Template

To: [name/email]

From: “Vanessa Sanchez” vanessa.sanchez@utexas.edu

CCd: “Fleischmann, Ken” kfleisch@ischool.utexas.edu, "Lassiter, Tina" <tl7257@my.utexas.edu>, "Collier, Chelsea" <chelsea@digi.city>

Subject Line: Responsible AI operations research project: You’ve been selected!

Dear <Name>,

Thank you for expressing interest in our research project. You’ve been selected to participate! Attached you will find a participant information sheet and a consent form. Below you will see a link to Calendly.

1. Please review the information sheet carefully to understand what the study entails.
2. Please read the attached consent document about having your session recorded to help us take better notes. We will ask you for your informed consent at the beginning of your session.
3. Select a time slot (up to 3) that works for you using this Calendly Link: [insert Calendly link]
4. To help you prepare for your session, please make sure you will have a quiet space available with a desktop computer and good internet connection.

We look forward to hearing from you.

Regards,
Vanessa

VANESSA SÁNCHEZ | M.S.I.S. Human-Computer Interaction and Responsible AI
The University of Texas School of Information
+1 (512) 919-9821
Email: vanessa.sanchez@utexas.edu
LinkedIn: linkedin

Appendix E: Interview Protocol

[Introduction (5 minutes)]

Hello _____, my name is Vanessa. I am a graduate student in the School of Information at The University of Texas at Austin. How are you today?

Thank you for volunteering to participate in today's study! You were emailed detailed information about the nature of this study and our intent in publishing the results. To reiterate the most important points of that document:

- Your personally identifiable information, including the name of your employer, will NOT be revealed in any of the research findings.
- All personally identifiable information collected from you is coded in our confidential research notes.
- All information collected from you is stored in a private location only accessible by the research team.
- This activity is entirely voluntary and you can stop at any time. If there is a question you do not wish to answer, you are free to skip it.
- With your permission, we'd like the session to be video/audio recorded to facilitate data analysis and reporting. All recordings will only be used for internal purposes and are completely confidential.

We **[received / did not receive]** your signed consent form allowing us to record the session.
[If received] So I'll start recording now.
[If not received] Do I have your permission to record?

[If no, move on without recording]
[If yes, start recording]

Do you have any questions before we begin?

[Section 01]

Ok, let's start with some background information. This should only take about 5 to 10 minutes.

1A. Can you tell me about your professional background?

1B. Can you describe your professional role and the responsibilities that have come with it?

1C. How familiar are you with AI and in what way?

[Section 02]

Now let's discuss ethical/responsible AI practices you've encountered in your professional role. This section shouldn't take more than 15 minutes.

2A. How has AI ethics shown up in your professional role? How has it impacted your activities?

2B. What kinds of questions do you personally have about the AI data, tools, services, or products that you have been involved with or impacted by in your professional role or that you expect to encounter?

2C. How do responsible AI questions or responsible AI operations show up in the typical teams or departments you interact with in your professional role?

2D. What are some things you've seen that work well in responsible AI practices that you've experienced or directly observed?

2E. Thinking about how those practices intersect with your role, how do you or would you measure the success of those practices? What does successful responsible AI operations look like to you in your role?

[Section 03]

Now let's discuss ideas on enabling ethical/responsible AI practices from a tangible tools, materials and assets perspective. This section shouldn't take more than 30 minutes.

[Paste Miro Board link in the chat]

Go ahead and click the link in the chat, which will take you to the Miro Board for the workshop activity. Feel free to take few seconds to look around the space. We'll start with the top board. You can edit a sticky note by clicking on it. You can also right click to duplicate it. You're free to interact with the board if you are comfortable doing so. Otherwise, I can help.

3A. Are there any tools, materials, or assets you currently find useful enablers of AI ethics operations in your role?

3B. Think about your role in terms of training, workflow, and evaluation. Now, considering challenges related to responsible AI operations you've encountered or expect to encounter—what might be some new tangible enablers (tools, materials, assets) you would find useful?

3C. Pretend you have a $1,000 budget to use on AI ethics operations development in your role and you need to spend it all. Based on your ideas from the last question, how would you allocate the cash across those ideas?

3D. To your knowledge, what roles would most likely interact with the idea you allocated the most money to?

3E. Any thoughts on what might make that idea more likely to be successful in service of your role, interacting roles, and your team's goals?

That concludes this interview. I'll stop the recording.

Would you like us to send you a report summary of your session or for your industry once it's completed?

Do you have any feedback or comments on the workshop experience?

Thank you for your time.

[End session]

Appendix F: Workshop Template

RESPONSIBLE AI OPERATIONS: TANGIBLE ENABLERS

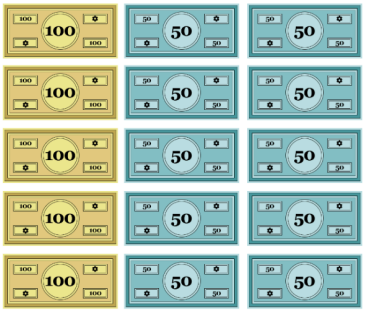
PARTICIPANT CODE: **INDUSTRY:** **PROFESSION:** **FACILITATOR:** Vanessa Sanchez

BRAINSTORM

What are existing tools or resources you find useful in supporting responsible AI operations?
Add sticky notes.

What are new tools or resources you would find useful in supporting responsible AI operations?
Add sticky notes.

Based on your ideas from the last question, how would you allocate the cash across those ideas?



miro

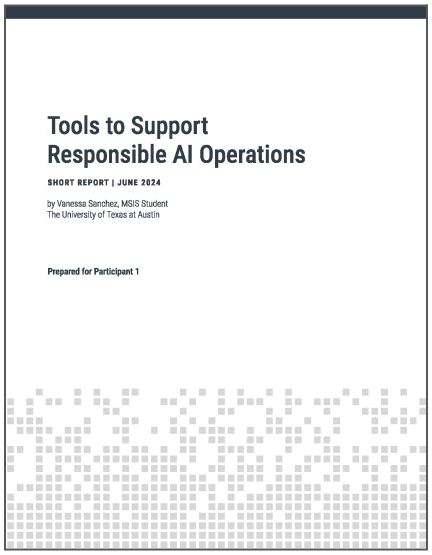
Appendix G: Cleaned Transcripts and Data

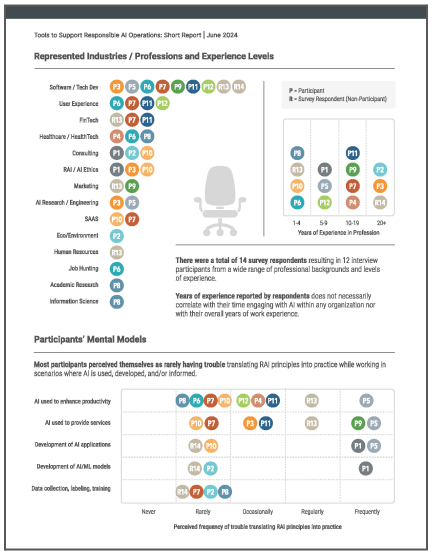
Cleaned and formatted transcripts are available at:
https://drive.google.com/drive/folders/1y4zNAsuLslacVIiQVwS6KI6IksLZd9Of?usp=drive_link

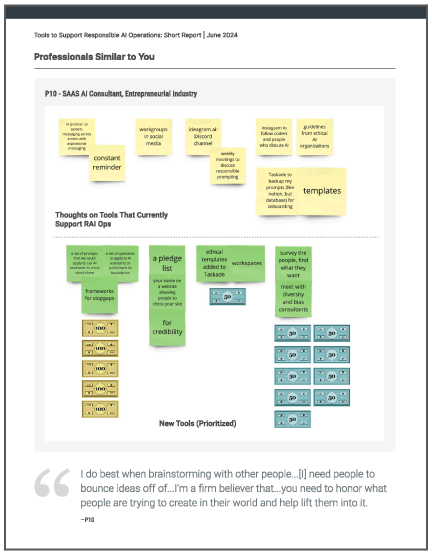
Cleaned data can be found at:
https://drive.google.com/drive/folders/1Q3lC8WSyuuvyW8r2NpSs6CNbpvNKzorM?usp=drive_link

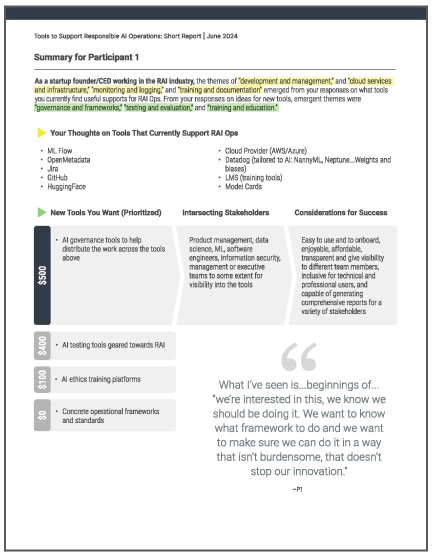
Appendix I: Short Reports

Short reports can be found at:
https://drive.google.com/drive/folders/1wIn7-JKIyVIRmyU-2_x1NW43peitllNy?usp=drive_link









(Thumbnails of a short report)

VITA

Vanessa Sanchez was born in Austin, Texas. After completing her work at Akins High School, Austin, Texas in 2004 within the top 10%, she entered Austin Community College in Austin, Texas. In Fall 2006, she attended The University of Texas at Austin to study Fine Arts. From 2007 to 2010, she attended Texas State University in San Marcos, Texas, to study Communication Design. Here, she was twice recognized on the Dean's List and achieved industry recognition for her graphic design work. She received the degree of Bachelor of Fine Arts in Communication Design, *magna cum laude,* with a minor in Anthropology from Texas State University in December 2010. During the following years, she freelanced as a visual designer and was employed by various companies, including GSD&M Idea City, NetSpend, and EY. In August 2021, she entered the Graduate School at the University of Texas at Austin.