Using Case Studies to Teach Responsible AI to Industry Practitioners

Julia Stoyanovich, Rodrigo Kreis de Paula, Armanda Lewis, Chloe Zheng

New York University, New York, NY, USA
{stoyanovich,rodrigo.kreis,al861,cz1300}@nyu.edu

Abstract

Responsible AI (RAI) encompasses the science and practice of ensuring that AI design, development, and use are socially sustainable——maximizing the benefits of technology while mitigating its risks. Industry practitioners play a crucial role in achieving the objectives of RAI, yet there is a persistent a shortage of consolidated educational resources and effective methods for teaching RAI to practitioners.

In this paper, we present a stakeholder-first educational approach using interactive case studies to foster organizational and practitioner-level engagement and enhance learning about RAI. We detail our partnership with Meta, a global technology company, to co-develop and deliver RAI workshops to a diverse company audience. Assessment results show that participants found the workshops engaging and reported an improved understanding of RAI principles, along with increased motivation to apply them in their work.

Introduction

Responsible AI (RAI) encompasses the science and practice of ensuring that AI design, development, and use are socially sustainable——maximizing the benefits of this technology while mitigating its risks. Widespread recognition of the importance of RAI has led to recent legislative and regulatory decisions, and high-level directives. Notable examples include the recent U.S. Executive Order on Safe, Secure, and Trustworthy Artificial Intelligence (EO 14110) (Biden Jr. 2023), which commits to "addressing safe, responsible, and non-discriminatory uses of AI," and the European Union's adoption of the Artificial Intelligence Act (Union 2023).

Stakeholders, including technical developers, designers, end-users, others impacted by AI, and society at large, have distinct priorities. For this reason, deep engagement with the tensions between differing stakeholder perspectives is necessary to build and deploy AI systems responsibly. Industry practitioners play a decisive role in our collective ability to achieve the goals of RAI, making it essential for them to collaborate with academic institutions to integrate RAI advances into applications and services. We describe such a collaboration in this paper.

We are members of the Center for Responsible AI at New York University (NYU R/AI), an academic institution dedicated to advancing RAI principles through research, education, and collaboration. We partnered with Meta, a global technology company, which had initiated internal advocacy efforts to promote RAI and recognized the importance of leadership-level support to embed these principles into its organizational culture. With input from Meta staff, we developed RAI training materials. We then offered workshops to equip legal, managerial, and technical professionals within the company with the knowledge and skills needed to integrate RAI principles into the design, development, and deployment of AI systems.

The driving questions behind our work are: "How can we engage organizational partners to better integrate RAI as a core component of their institutional goals?" and "What strategies facilitate diverse practitioners—across design, legal, and technical teams—to learn about and incorporate RAI principles in their daily work?" To answer these questions, we followed a stakeholder-first approach to training (Dominguez and Stoyanovich 2023), using *interactive case studies* to achieve organizational and practitioner-level engagement, promote inter-role collaboration, encourage a multi-perspective approach, and advance RAI learning in a practical environment. By grounding learning in practical, real-world scenarios, this approach not only deepens understanding of RAI principles but also fosters inter-role collaboration, making it more likely that these principles are embedded into organizational processes and decision-making.

This collaboration posed several challenges, including lack of access to information about the company's proprietary algorithms and systems during case study development. Additionally, organizational changes at Meta during the collaboration led to team restructuring, complicating efforts to achieve collective buy-in. Nonetheless, our qualitative and quantitative analysis of the workshops indicate that participants— many with limited prior knowledge—found the training engaging and reported a positive shift in their understanding and motivation to apply RAI concepts to their work. The use of case studies relevant to participants' day-to-day activities was particularly well-received.

Related Work

The concept of the stakeholder—"any group or individual who can affect or is affected by the achievement of the organization's objectives" (Freeman 2010) —has long been a

focus in organizational research and is increasingly relevant in AI contexts. Scholars have noted the importance of incorporating multi-stakeholder perspectives, including decision-makers, companies, shareholders, government and regulatory bodies, technologists, end-users, and society at large, at all stages of algorithmic system development (Güngör 2020; Lima and Cha 2020; Nabavi and Browne 2023). In examining stakeholder influence within the AI lifecycle, Miller (2022) extends the standard classification of Mitchell, Agle, and Wood (1997)—power, legitimacy, and urgency—by adding harm as another attribute, enabling the identification of additional groups potentially affected during the development and operating stages of an AI project.

Additionally, research has looked at integrating multiple stakeholders into the technical aspects of design and deployment processes, particularly targeting AI system designers, engineers, and researchers. For example, Bell, Nov, and Stoyanovich (2023) argue that technologists, equipped with both expertise and organizational influence, are well-positioned to drive meaningful change. They propose a four-tier hierarchy to guide the design and deployment of transparent automated decision systems. Within this framework, practitioners consider diverse stakeholders—such as affected individuals, policymakers, and other technologists—and use these insights to formulate appropriate goals, purposes, and methods. Further, Abdollahpouri and Burke (2019),Abdollahpouri et al. (2020), and Bell et al. (2023) demonstrate that fair machine learning models can be effectively developed through a stakeholder-driven approach, which integrates the priorities of practitioners with the needs of affected individuals, policymakers, and other technologists, all while maintaining minimal trade-offs in accuracy.

While we recognize the importance of individual stakeholders, we also emphasize the role of organizational representatives in advancing AI. These representative can drive adoption of RAI by elevating responsibility within the organization and signaling its importance to both internal and external stakeholders.

Our process includes two essential levels of engagement to advance practitioners' understanding of RAI principles. The first involves engaging with *organizational partners* and higher-level managerial representatives, leveraging their active participation and institutional buy-in to shape the practitioner experience, which centers on instructive RAI case studies (Freeman et al. 2017; Shah and Guild 2022). The second focuses on engagement with *practitioners* responsible for technical and non-technical deliverables. This approach involves direct instruction in RAI principles and aligns with educational policy literature, which reports direct learning interventions as essential for sustainable institutional change (Goldsmith and Burton 2017; Henry et al. 2013). We integrate pedagogical best practices by designing active learning experiences that enable practitioners to learn by doing and construct knowledge based on their roles, perspectives, and interests, fostering bottom-up engagement around integrating RAI principles (Lewis and Stoyanovich 2022; Sawyer 2022; Wenger 1998). *Our primary method for fostering active learning among practitioners is the implementation of thoughtfully crafted case studies.*

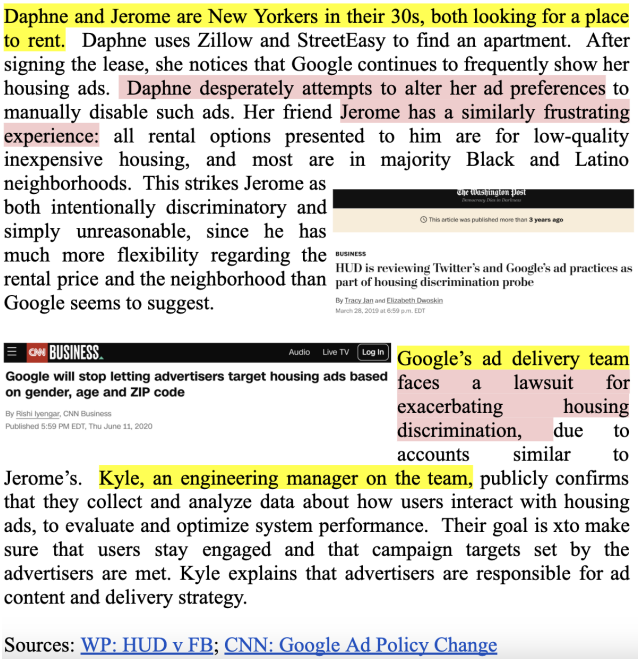

Figure 1: Housing ads case study: negative sentiment handout. Yellow highlights emphasize stakeholders, and red highlights emphasize the negative sentiment. Highlights were not part of the handout used during the workshops. See Appendix in the extended version of the paper for a complete set of handouts (Stoyanovich et al. 2024).

Engaging the Organizational Partner

We are a team of academic researchers and educators from NYU R/AI[1], a U.S.-based academic center, referred to here as the *academic partner*. We conducted this work in close collaboration with Meta, a major tech corporation, referred to as the *organizational partner*. Like many other technology companies today, Meta has a RAI team that implements and disseminates RAI practices within the organization. This team initiated our collaboration to develop and iteratively refine training materials on RAI for the organization's staff and to offer workshops using these materials. Our collaboration began in Fall 2021 (when the company still operated as Facebook) and concluded in Spring 2023. During this time, the *organizational partner* underwent structural changes, which led to several shifts in the project's "ownership" within the company.

In Fall 2021, we collaboratively designed a *pilot workshop* titled "Demystifying Responsible AI," offered over four 60-minute sessions in Winter 2022. While several pilot participants were part of the company's RAI team or worked closely with it, they did not represent the majority. The pilot workshop covered four main themes: (1) What is responsible AI? (2) Transparency and interpretability; (3) Algorithmic fairness; and (4) A lifecycle view of responsible AI. Each session followed a consistent structure, starting with a brief, discussion-based warm-up activity, followed by a two-

[1]https://r-ai.co

part presentation from the lead instructor. Between the presentation segments, there was a 10-minute discussion, with an additional discussion at the end of the session. Instructional materials included a rich set of examples and featured several case studies, albeit briefly presented. Participants also received optional reading, consisting of general-audience and scientific articles on RAI.

We collected informal feedback from pilot workshop participants to inform the next round of iterative development. Overall, the feedback was positive, while listing areas for improvement. *First*, several participants noted that they enjoyed the examples and case studies presented during each session. They suggested—including agreement from the organizational partners—that additional case studies should be incorporated into future iterations. Further, participants emphasized the importance of aligning case studies with their day-to-day professional activities. *Second*, several participants found discussions and Q&A to be the most engaging aspect of the workshop. They suggested making the workshop more interactive by reducing instructor-led presentation time and extending group discussion time. *Third*, although about 20 participants joined the first session, there was substantial attrition by the fourth session, possibly due to work-related demands that may have reduced participants' availability to attend all workshop sessions. To address this and reduce the human resources cost on the company, the *organizational partner* suggested condensing the workshop into fewer sessions.

We incorporated these suggestions and immediately commenced the difficult work of identifying and developing case studies for the next iteration of the workshop, as detailed in next section. We also redesigned the workshop's structure in line with suggestions from the *organizational partner* and pilot workshop participants. The second iteration, offered in early Spring 2023, was titled "What is Responsible AI and how does it apply to your work at Meta?" to emphasize the connection between the content and participants' day-to-day activities. We will describe the structure and content of the workshop under Workshop Implementation.

Case Studies for Teaching RAI

Creating and deploying effective case studies is pivotal in the pedagogical pursuit of raising RAI awareness and knowledge among practitioners. Case studies offer a pragmatic and investigative approach to understanding the complexities, challenges, and ethical dimensions of AI architecting, building, and implementation (Turner and Danks 2014). More generally, Kreber (2001) argues that case studies provide essential experiential learning opportunities, prompting reflection, reconceptualization, and applied use of learned materials when implemented effectively. The literature on teaching AI and AI ethics (Hishiyama and Shao 2022; Khan et al. 2022; Laine, Minkkinen, and Mäntymäki 2024) highlights that case studies help learners ground ethical principles in realistic scenarios and reveal tensions.

We follow this approach and use high-quality case studies based on current AI systems, platforms, and scenarios to prompt practitioners to *identify stakeholder benefits and harms, evaluate trade-offs, and reconcile tensions*. Case studies are also effective at fostering organizational improvement. Exploring real-world scenarios provides practitioners with a sandbox to refine processes, policies, and overlooked ethical issues, including competing interests and complex trade-offs (Garrett, Beard, and Fiesler 2020; Kazim and Koshiyama 2021; Dominguez and Stoyanovich 2023). This section explores our most relevant findings on designing case studies for teaching RAI and presents the case studies developed for the 2023 workshops.

Case study selection. Creating a case study that resonates with the audience begins with carefully *considering the individuals* who will be studying it. Practitioners come from diverse backgrounds with varying skills and professional goals, so understanding their knowledge and objectives is vital to ensuring the case study meets their educational needs. Alongside focusing on the audience, we prioritized *choosing cases rich in detail and complexity*. These cases present a range of challenges and ethical issues related to AI, fostering immersive learning and providing practitioners with a broad view of its landscape. We also emphasized the value of *cases with diverse stakeholders, whose differing perspectives can reveal conflicting interests and tensions*.

Finally, the organizational partner encouraged us to *choose cases that were directly relevant to their organization and participants' day-to-day activities*. Satisfying this requirement proved difficult, as we lacked access to internal information about Meta's products and services. Additionally, we aimed to develop case studies that could be shared publicly to support broader RAI training efforts. For these reasons, we selected case studies with sufficient publicly available information that were also broadly relevant to Meta's application domain and industry but that do not center around any of their strategic products or services.

Having selected the case studies, we proceeded to document each one comprehensively.

Case study documentation. We comprehensively documented each case study using the following format.

Overview: A succinct, informative introduction of the AI system, setting the stage for subsequent exploration, offering context, and framing key aspects to be examined.

Background and context of use: Insights into the broader background and context of the AI system, including its implementation timeline and challenges, to help practitioners grasp its real-world significance.

Technical details: A deep dive into the system's inner workings, such as its architecture, data sources, goals, performance metrics, validation, and improvement history.

Legal and ethical considerations: Identification of legal, ethical, and other RAI-related concerns and ascertaining whether and how they have been addressed.

Stakeholder analysis: A critical dimension of the case study involves identifying and comprehensively analyzing diverse stakeholders associated with the AI system. This includes surfacing stakeholder perspectives and goals, examining their level of participation in the design, development, evaluation, and oversight of the system (if any), and assessing the benefits and harms for each stakeholder.

Transparency and explainability: A closer investigation of these two specific RAI principles, focusing on their relevance to the goals and perspectives of different stakeholders.

Stakeholder matrices. Next, we adopted a structured assessment approach to create comprehensive matrices that systematically outlined the benefits, harms (and their possible origins), tensions, and strategies for tension mitigation or resolution between stakeholders. We filled out these matrices for each case study and used them to guide the design of interactive workshop activities with facilitators, using tailored questions to prompt and structure discussions in breakout groups. In these matrices and all other training materials, we intentionally avoided formal definitions of RAI concepts like fairness, agency, and safety. This approach shifted the focus from terminology to encouraging participants to interpret these concepts within specific case studies.

To encourage discourse, we adopted an interactive learning model centered on matrix structures as frameworks for workshop participants (see Figure 2 in the extended version of the paper (Stoyanovich et al. 2024)). These matrices featured placeholders for learners to complete, with facilitators guiding discussions using handouts. By providing the matrix structures without pre-filled content, we empowered practitioners to become co-creators, filling in the matrices with insights, ethical considerations, and real-world examples from the case studies. This hands-on approach ensured practitioners were active participants in developing their understanding of RAI, rather than passive recipients of information. By placing learners at the center of knowledge construction, this approach transcended traditional teaching methods, empowered them to actively shape their understanding of AI ethics and responsible decision-making through experiential and interactive learning.

Handouts. Next, based on case study documentation and matrices, we crafted practitioner-friendly handouts for the workshop. These handouts distilled the key elements of each case study into a clear, concise, and jargon-free format, providing participants with a structured guide to navigate the complexities of the cases. Practical case studies were presented as short stories, including excerpts from newspapers or magazines. To encourage active engagement and critical thinking, the handouts included discussion prompts and provocative or controversial statements.

To enhance the training approach, we created two versions of each handout: one presenting the technical system in a positive light and the other in a negative light, particularly in terms of its repercussions for affected individuals (see Figure 1 for an example). This strategy fostered critical thinking and promoted a comprehensive understanding of RAI.

Workshop Implementation

Learning outcomes The learning outcomes of the workshop are grouped into three categories: (1) RAI concepts, (2) stakeholders, benefits, and risks, and (3) risk mitigation strategies. Upon completion, learners should be able to: Identify and define basic RAI concepts; Recognize RAI concepts relevant to a specific system, product, or service ("system" for short); Explain RAI concepts related to a specific system to their team members; Identify key *stakeholders* of a specific system; Identify the *benefits* and *risks of harm* associated with a system for each stakeholder, and relate these to *RAI concepts*; Identify tensions between the benefits and risks of a system across stakeholders and apply techniques to reconcile these tensions; Explain stakeholders, benefits, risks, and tensions to members of their team; Analyze how specific harms may arise in relation to a system and how its data, technical properties, and context of use may increase these risks; Propose and describe *mitigation strategies* to address specific risks of a system; and Describe potential harms and mitigation strategies to their team.

Workshop schedule. The workshop was iteratively designed based on suggestions from pilot participants and conducted in two 120-minute sessions to minimize participant attrition. To enhance interactivity, participants spent 50% of the time (60 minutes per session) actively working through the case studies in small moderated groups. The workshop was offered twice in quick succession to accommodate all interested participants while keeping group sizes small enough to ensure active engagement. Session 1 was structured as a sequence of six activities:

- Introduction and welcome (10 min);
- Lecture (20 min);
- Moderated case study discussion in break-outs (30 min);
- Report-back to the full group (5 min);
- Moderated case study discussion in break-outs (30 min);
- Closing reflections (10 min).

Participants in different break-out rooms worked through different versions (positive/negative) of the same case study. Discussion during the first 30-minute activity revolved around stakeholder identification, goals and priorities, and benefits and harms to a selected set of stakeholders. The second activity continued with the same case study and focused on using RAI concepts to identify and reconcile tensions. Session 2 followed a similar structure. During Sessions 1 and 2, we presented concrete examples to illustrate RAI concepts, providing essential context. This initial exposure enabled all participants, including those with limited prior experience in RAI, to meaningfully engage with these topics in group discussions. The first case study was predefined and examined reasons for harm and potential sociotechnical mitigation strategies. The second case study was open-ended, inviting participants to compose and describe their own scenarios, identify stakeholders, and analyze and reconcile the benefits and harms for different stakeholders.

Selected case studies. We selected two case studies for the workshops. We used the *housing ads delivery* case study to support Session 1 activities, see Figure 2 in the extended version of the paper for a snapshot.[2] Participants explored the benefits of personalized ad delivery for vendors, advertisers, advertising platforms, and customers. They juxtaposed these benefits against the harms of bias and discrimination

[2]Additional details about the case studies are available in the full version of the paper (Stoyanovich et al. 2024).

against historically disadvantaged groups, as well as the loss of agency and control experienced by individuals targeted by the ads.

We used the *toxic content moderation* case study during Session 2. Participants considered automated moderation systems that flag and remove comments identified as toxic using supervised machine learning methods. These systems may be prone to bias, including pre-existing bias (e.g., if training data contains a disproportionate number of toxic comments from a specific demographic group, leading the system to flag comments from this group more frequently) and technical bias (e.g., if the system prioritizes certain types of comments during training or is tuned to favor specific performance metrics). Participants discussed how toxic comment moderation can help users reclaim their virtual space after being victims of online bullying. They also discussed scenarios where social media platforms fail to flag toxic comments deemed "not harmful enough" by their guidelines, or erroneously flag and remove valid posts and supportive comments.

The dual (positive vs. negative sentiment) prompts for both case studies stimulated participants to think critically about the presented situations, boosting deeper discussion by adopting disparate perspectives.

Practitioners as case study creators and assessors. At the end of the educational process, practitioners become creators and assessors of case studies informed by their personal and professional experience. This activity involved designing real-world AI scenarios, describing systems with controversial effects on diverse stakeholders, creating assessment matrices, identifying benefits, harms, tensions, and mitigation strategies, and conducting peer reviews. This method empowered practitioners to apply RAI principles in practice and simulate the assessment of real-world case studies they may design, develop, or deploy in their daily work.

Analysis of Engagement and Learning

Successes and challenges. Workshop participants stayed interested and engaged throughout the sessions, which we attribute to the relevance of the case studies and the ample opportunities to actively engage with the material. The quality of the interactions was remarkable, with participants contributing diverse perspectives from different stakeholder viewpoints and proposing effective strategies to mitigate potential tensions. This engagement was most evident when participants created thorough case studies from scratch, demonstrating both active participation and satisfactory level of understanding. However, attrition was a notable challenge, with just over 50% of participants from the first session returning for the second.

In the remainder of this section, we analyze engagement and learning based on workshop observation and pre-/post-survey results. We note that the cohorts across the two workshop offerings represent a range of technical and non-technical roles (see Table 1).

Measuring engagement

Participants reported feeling "curious," "excited," and "hopeful" about workshop content. They cited rationales for participating such as having a limited background on AI and RAI topics, recognizing the positive and negative transformative potentials of AI, and wanting to examine case studies and strategies for deploying AI responsibly. These adjectives align with the majority of participants (71%) lacking specific expectations for the workshop, indicating they were receptive to learning more without preconceived agendas. Over 44% of participants had never previously worked with RAI concepts, while the 29% who had familiarity were primarily interested in gaining a general understanding about RAI.

We compare our sample with responses from the 2019 Mozilla Foundation survey on public perception of AI (Foundation 2019). Among the 67,000 respondents in that survey, the most common adjectives describing AI were curious (30%), hopeful (27%), and concerned (32%). In contrast, responses from our workshop participants were more positive and less concerned. This difference is likely because our sample consists of technology company employees, who may be inclined to view AI in a more positive light.

Among the participants across the two workshops, 29 (76%) viewed RAI as important and 5 (13%) viewed it as somewhat important to their day-to-day work. Among the remaining participants, 2 (5%) were unsure about the connection of RAI to their work, and the remaining 2 (5%) responded that RAI is not important to their work.

On average, 61% of participants believed AI would definitely improve their lives, 37% thought it might, and only one participant felt AI would not improve their life. Notably, there was a 16% gap between the perceived work-related relevance of RAI and the belief that AI would improve life. For context, the 2019 Mozilla Foundation survey found 24% of respondents believed AI would improve society, 10% thought it would worsen society, and 41% held a nuanced view, acknowledging both positive and negative impacts. This difference likely reflects the AI-optimistic outlook of our sample, composed of technology company practitioners, compared to the general public surveyed by Mozilla.

A key aspect of the workshop was engagement with the provided case studies. In the post-workshop survey, respondents ($n = 16$) reported that the most important stakeholders to consider are current or potential users and society at large, indicating that participants were examining RAI beyond their immediate work-specific contexts. Most respondents (77%) reported being engaged by the participant-to-participant interaction provided through small case study discussion groups and hands-on activities. Additionally, 100% of respondents agreed that the case study format was an effective for understanding RAI concepts.

Measuring learning about RAI

The case studies formed the core of workshop interaction. We used Jamboard, a free visual brainstorming tool by Google, to facilitate simultaneous group activities.

Session 1. Figure 2 in the extended version of the paper (Stoyanovich et al. 2024) gives an example Jamboard,

role	number of participants
content / media specialist	3
designer	10
engineer / research scientist	12
legal / policy adviser	4
project manager	9
total	**38**

Table 1: Workshop participant roles.

where one group—similar to others—successfully identified stakeholders in the *housing ads delivery* case study who might be impacted by AI. They also outlined a range of positive and negative effects of the system on stakeholders. Across groups, practitioners extensively explored the system's implications for individual users and the broader public, addressing issues such as fairness, privacy, and safety. Survey results show that participants responded positively to the case study. The generated ideas and discussions demonstrated that the case study effectively encouraged critical thinking about RAI complexities, including competing stakeholder interests and strategic interventions.

Part of Session 1 involved ranking the benefits and harms as primary and secondary impacts, linking them to RAI concepts, and recommending mitigation strategies. Interestingly, some groups dedicated time to creating a taxonomy to distinguish primary vs. secondary concerns. While certain groups focused primarily on the benefits or harms to specific stakeholders (e.g., technology vendors or users), others approached the task by considering societal impact as the basis for the primary vs. secondary distinctions.

We observed that, across groups, users elicited the most primary harms, while the platform elicited the most secondary harms. Participants indicated no difficulty in adopting the user's perspective and recognized the platform's ultimate goal as a user-facing system. Participants discussed how to interpret RAI concepts like privacy, robustness, fairness, inclusion, or accountability with respect to stakeholders, and, further, how to map RAI concepts to benefits and harms. Mitigation strategies proposed by participants included technical, policy, and design interventions, though they consistently reported feeling they had insufficient time to explore these actions in depth.

Session 2. This session focused on applying RAI principles using the *toxic comment moderation* case study. Participants found it realistic and positively challenging, highlighting the complexities of addressing RAI concepts. Many endorsed the use of authentic case studies, with one participant noting, "the more applied the case studies are, the more valuable the workshop would be in the future."

The *toxic comment moderation* case study was intentionally designed to present more complex considerations than the *housing ads delivery* case. When focusing on users as stakeholders, several groups observed that users are not a monolithic group, making harm mitigation strategies under RAI principles inherently non-uniform. The case studies illustrated diverse perspectives on technology, offering, as one

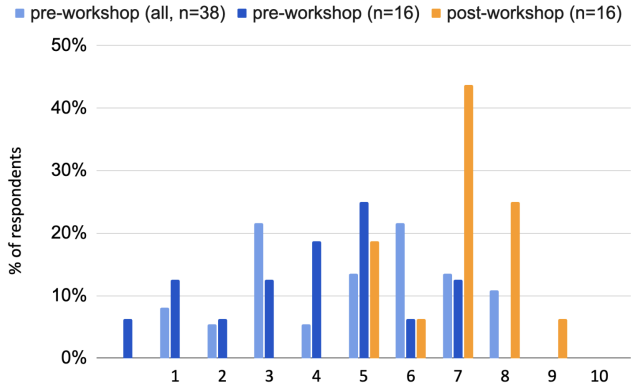

Figure 2: Self-assessment of RAI knowledge on a 10-point scale, from 1 ("What is RAI?") to 10 ("I'm an expert on RAI"). Pre-workshop responses from all participants ($n = 38$) are shown in light blue, pre-workshop responses from participants who also completed post-workshop surveys ($n = 16$) are in dark blue, and post-workshop responses from these participants are in orange.

participant noted, "real-life positive and negative examples of AI (e.g., Amazon's AI recruiting model biased against women)" (referencing (Dastin 2018)).

The final interactive activity challenged participants to create an original case study based on their personal or professional experiences. Topics included insurance claim classification, autonomous vehicles, job advertisement targeting systems, social media beauty influencers, and virtual reality characters. Participants demonstrated increased use of RAI concepts and terminology compared to the start of the workshop, as reflected in dialogue and content analysis. They also identified and discussed tensions more quickly than during the initial breakout session.

Assessment of implementation. We conducted pre- and post-workshop self-assessments of RAI knowledge, summarized in Figure 2. Among the participants, 38 completed the pre-workshop survey, and 16 completed both pre- and post-workshop surveys. Participants reported a significant increase in their understanding of RAI concepts and techniques after the workshop, with a weighted average score of 4.88 (all participants, $n = 38$) or 4.84 (participants who submitted both surveys, $n = 16$) before the workshop, rising to 6.94 post-workshop. All participants indicated at least some improvement in their RAI knowledge, with 77% reporting substantial to very substantial improvement and 100% expressing motivation to apply workshop content to their work. Participants appreciated the workshop's compact format but expressed interest in continuing their study of RAI.

Conclusion

In this paper, we shared our experience teaching Responsible AI (RAI) to industry practitioners at Meta. Assessment results show that participants, many with limited prior knowledge, found the workshops engaging and reported improved understanding of RAI, along with increased motivation to

apply it in their work. The use of case studies tailored to participants' work activities was especially well-received. Below, we outline best practices for creating effective case studies to teach RAI to practitioners.

Case study selection. Start by considering your practitioner audience's background, knowledge, interests, and goals. Tailor case studies to align with their specific needs and objectives, focusing on scenarios closely tied to their daily activities or organizational context. Choose case studies with depth and complexity, incorporating diverse challenges and ethical dimensions to foster a comprehensive exploration of RAI. Prioritize scenarios involving multiple stakeholders with differing perspectives, as these tensions encourage learners to engage with the complexities of real-world AI decision-making. For breakout activities, create heterogeneous groups by mixing participants with varied backgrounds, roles, and perspectives on AI and RAI.

Case study materials. Create matrices that systematically outline benefits, harms, tensions, and strategies for resolving or mitigating tensions across stakeholders. These matrices serve as interactive tools to deepen participants' understanding and promote the adoption of RAI practices. Additionally, develop multiple versions of handouts—for example, one with a positive and another with a negative portrayal of the AI system's impact. This approach encourages nuanced exploration of framing effects, fosters diverse perspectives in discussions, and prompts participants to critically reflect on their own viewpoints.

Practitioners as case study creators. Conclude the learning journey by empowering practitioners to create case studies from scratch. This step fosters critical thinking, practical application, and a deeper understanding of RAI principles. It also builds participants' confidence, enabling them to apply the process to real-world AI systems, effectively communicate RAI concepts, navigate tensions, and lead future educational activities.

Limitations and future directions. This study is an experience report based on workshops conducted for practitioners at a single—albeit highly significant—industry partner. Although the audience was limited, the high level of engagement, direct feedback, and participants' demonstrated understanding of RAI concepts through discussions validate the effectiveness of using case studies to deepen comprehension of complex topics. A key area for future exploration is the longitudinal impact of case studies. For instance, it would be valuable to examine how practitioners continue to engage with RAI concepts—both implicitly and explicitly—months after the workshop. Another important avenue for study is the capacity to mitigate RAI threats, which several participants identified as a natural progression from the workshop content.

Acknowledgements

The authors express their gratitude to Meta's Responsible AI team and colleagues for their invaluable support in developing educational materials, case studies, and workshops. Special thanks go to Parisa Assar, Emily McReynolds, Jacquieline Pan, Hunter Goldman, Eleonora Presani, and Miranda Bogen. The authors also thank Lucius Bynum, Venetia Pliatsika, and Lucas Rosenblatt from the NYU Center for Responsible AI for co-facilitating the workshops.

Extended version — https://arxiv.org/abs/2407.14686

References

Abdollahpouri, H.; Adomavicius, G.; Burke, R.; Guy, I.; Jannach, D.; Kamishima, T.; Krasnodebski, J.; and Pizzato, L. 2020. Multistakeholder recommendation: Survey and research directions. *User Modeling and User-Adapted Interaction*, 30(1): 127–158.

Abdollahpouri, H.; and Burke, R. 2019. Multi-stakeholder Recommendation and its Connection to Multi-sided Fairness. ArXiv:1907.13158 [cs].

Bell, A.; Bynum, L.; Drushchak, N.; Zakharchenko, T.; Rosenblatt, L.; and Stoyanovich, J. 2023. The Possibility of Fairness: Revisiting the Impossibility Theorem in Practice. In *Proceedings of the 2023 ACM Conference on Fairness, Accountability, and Transparency*, FAccT '23, 400–422. New York, NY, USA: Association for Computing Machinery. ISBN 9798400701924.

Bell, A.; Nov, O.; and Stoyanovich, J. 2023. Think about the stakeholders first! Toward an algorithmic transparency playbook for regulatory compliance. *Data & Policy*, 5: e12.

Biden Jr., J. R. 2023. Executive Order # 14410: Safe, Secure, and Trustworthy Artificial Intelligence.

Dastin, J. 2018. Amazon scraps secret AI recruiting tool that showed bias against women. *Reuters*.

Dominguez, D.; and Stoyanovich, J. 2023. Responsible AI literacy: A stakeholder-first approach. *Big Data and Society*.

Foundation, M. 2019. We Asked People Around the World How They Feel About Artificial Intelligence. Here's What We Learned. *Mozilla Foundation Blog*.

Freeman, R. E. 2010. *Strategic Management: A Stakeholder Approach*. Cambridge University Press.

Freeman, R. E.; Kujala, J.; Sachs, S.; and Stutz, C. 2017. Stakeholder Engagement: Practicing the Ideas of Stakeholder Theory. In Freeman, R. E.; Kujala, J.; and Sachs, S., eds., *Stakeholder Engagement: Clinical Research Cases*, volume 46, 1–12. Cham: Springer International Publishing. ISBN 9783319627847 9783319627854.

Garrett, N.; Beard, N.; and Fiesler, C. 2020. More Than "If Time Allows": The Role of Ethics in AI Education. In *Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society*, 272–278. New York NY USA: ACM. ISBN 9781450371100.

Goldsmith, J.; and Burton, E. 2017. Why Teaching Ethics to AI Practitioners Is Important. *Proceedings of the AAAI Conference on Artificial Intelligence*, 31(1).

Güngör, H. 2020. Creating Value with Artificial Intelligence: A Multi-stakeholder Perspective. *Journal of Creating Value*, 6(1): 72–85.

Henry, M.; Lingard, B.; Rizvi, F.; and Taylor, S. 2013. *Educational Policy and the Politics of Change*. Routledge, 0 edition. ISBN 9780203349533.

Hishiyama, R.; and Shao, T. 2022. Educational Effects of the Case Method in Teaching AI Ethics. In Rocha, A.; Adeli, H.; Dzemyda, G.; and Moreira, F., eds., *Information Systems and Technologies*, 226–236. Cham: Springer International Publishing. ISBN 9783031048265.

Kazim, E.; and Koshiyama, A. S. 2021. A high-level overview of AI ethics. *Patterns*, 2(9): 100314.

Khan, A. A.; Badshah, S.; Liang, P.; Waseem, M.; Khan, B.; Ahmad, A.; Fahmideh, M.; Niazi, M.; and Akbar, M. A. 2022. Ethics of AI: A Systematic Literature Review of Principles and Challenges. In *The International Conference on Evaluation and Assessment in Software Engineering 2022*, 383–392. Gothenburg Sweden: ACM. ISBN 9781450396134.

Kreber, C. 2001. Learning Experientially through Case Studies? A Conceptual Analysis. *Teaching in Higher Education*, 6(2): 217–228.

Laine, J.; Minkkinen, M.; and Mäntymäki, M. 2024. Ethics-based AI auditing: A systematic literature review on conceptualizations of ethical principles and knowledge contributions to stakeholders. *Information & Management*, 61(5): 103969.

Lewis, A.; and Stoyanovich, J. 2022. Teaching Responsible Data Science: Charting New Pedagogical Territory. *International Journal of Artificial Intelligence in Education*, 32(3): 783–807.

Lima, G.; and Cha, M. 2020. Responsible AI and Its Stakeholders.

Miller, G. J. 2022. Stakeholder roles in artificial intelligence projects. *Project Leadership and Society*, 3: 100068.

Mitchell, R. K.; Agle, B. R.; and Wood, D. J. 1997. Toward a Theory of Stakeholder Identification and Salience: Defining the Principle of Who and What Really Counts. *The Academy of Management Review*, 22(4): 853.

Nabavi, E.; and Browne, C. 2023. Leverage zones in Responsible AI: towards a systems thinking conceptualization. *Humanities and Social Sciences Communications*, 10(1): 1–9.

Sawyer, R. K., ed. 2022. *The Cambridge Handbook of the Learning Sciences*. Cambridge University Press, 3 edition. ISBN 9781108888295 9781108840989 9781108744669.

Shah, M. U.; and Guild, P. D. 2022. Stakeholder engagement strategy of technology firms: A review and applied view of stakeholder theory. *Technovation*, 114: 102460.

Stoyanovich, J.; de Paula, R. K.; Lewis, A.; and Zheng, C. 2024. Using Case Studies to Teach Responsible AI to Industry Practitioners. arXiv:2407.14686.

Turner, J. R.; and Danks, S. 2014. Case Study Research: A Valuable Learning Tool for Performance Improvement Professionals. *Performance Improvement*, 53(4): 24–31.

Union, E. 2023. The Artificial Intelligence Act.

Wenger, E. 1998. *Communities of practice*. Cambridge, England: Cambridge University Press.