Why (Not) Use AI? Analyzing People's Reasoning and Conditions for AI Acceptability

Jimin Mun[1], **Wei Bin Au Yeong**[1], **Wesley Hanwen Deng**[1], **Jana Schaich Borg**[2], **Maarten Sap**[1]
[1]Carnegie Mellon University
[2]Duke University
jmun@andrew.cmu.edu, wauyeong@andrew.cmu.edu

Abstract

In recent years, there has been a growing recognition of the need to incorporate lay-people's input into the governance and acceptability assessment of AI usage. However, how and why people judge acceptability of different AI use cases remains under-explored, despite it being crucial towards understanding and addressing potential sources of disagreement. In this work, we investigate the demographic and reasoning factors that influence people's judgments about AI's development via a survey administered to demographically diverse participants (N=197). As a way to probe into these decision factors as well as inherent variations of perceptions across use cases, we consider ten distinct labor-replacement (e.g., Lawyer AI) and personal health (e.g., Digital Medical Advice AI) AI use cases. We explore the relationships between participants' judgments and their rationales such as reasoning approaches (cost-benefit reasoning vs. rule-based). Our empirical findings reveal a number of factors that influence acceptance. We find lower acceptance of labor-replacement usage over personal health, significant influence of demographics factors such as gender, employment, education, and AI literacy level, and prevalence of rule-based reasoning for unacceptable use cases. Moreover, we observe unified reasoning type (e.g., cost-benefit reasoning) leading to higher agreement. Based on these findings, we discuss the key implications towards understanding and mitigating disagreements on the acceptability of AI use cases to collaboratively build consensus.

Datasets — https://github.com/JiMinMun/why-not-use-ai
Extended version — https://arxiv.org/abs/2502.07287

1 Introduction

There is a growing call from the public and experts alike to regulate the development and integration of AI into society (Pistilli et al. 2023; McClain 2025). These efforts, as reflected in the EU AI Act (Parliament 2023), NIST AI Risk Management framework (of Standards and Technology 2023), and recent U.S. Executive Order (exe 2023), have resulted in discussions about whether certain AI use cases should be pursued at all. As these high-stakes decisions shape the future of AI, it is essential to equitably determine which AI use cases warrant pursuit despite potential harms, requiring diverse public and expert perspectives. Furthermore, to mitigate potential disagreements, we must understand how individuals make such decisions.

One significant challenge when evaluating the acceptability and impact of AI use cases is that their effects can be simultaneously positive and negative, depending on the context of use, functionality, and broader societal implications (Mun et al. 2024). For instance, while educational AI can provide affordable and accessible personal tutor, it can, at the same time, lead to over-reliance of students and diminish the goal of education (Times 2024; Zhai, Wibowo, and Li 2024). Thus, as AI is applied across increasingly diverse domains, understanding how people make decisions about its use—especially when benefits and harms conflict—becomes critical for anticipating and addressing disagreements about specific use cases.

To tackle this challenge, we investigate the factors and reasoning that shape judgments of AI acceptability. First, we assess how judgments about the *acceptability of development and use* vary across different AI use cases,[1] and how they relate to scenario characteristics (**RQ1**). Second, we explore *personal factors influencing these judgments*, especially as they relate to demographic differences (Kingsley et al. 2024) (**RQ2**). Third, we analyze *reasoning strategies participants use when making judgments* about AI use cases, and how those strategies do or do not relate to the judgments that are ultimately made (**RQ3**).

To answer these questions, we develop a survey to collect judgments and reasoning processes of 197 demographically diverse participants with varying levels of experience with AI. We ask participants to report whether a certain AI use case should be developed or not, whether they would use such a system, and ask them to provide rationales for their judgment and conditions that would cause them to change their judgments (Figure 1). We perform a focused investigation of acceptability using ten AI use cases[2] that we systematically select for different risk levels, spanning two

[1]By use cases, we mean specific real-world scenarios or problems that an AI system is designed to address.

[2]We focus on text-based, non-embodied, digital systems, and while we do not specifically discuss the AI user and subject, in our use case description, we follow three of the five concepts used in EU AI Act to describe high risk use cases (Golpayegani, Pandit, and Lewis 2023): the domain, purpose, and capabilities.

highly-discussed domains with ongoing efforts to develop such use cases: personal health and labor replacement (McClain 2025; Kelly 2025; Kolata 2024; Pierson et al. 2025; Rajpurkar et al. 2022; Lee 2024). To understand characteristics of AI use cases that might affect perceptions beyond category, we vary them by required entry-level education and EU AI risk level (Table 1).

We perform a multi-pronged analyses of people's rationales. Drawing from moral philosophy, we examine participants' answers using two reasoning patterns: cost-benefit reasoning, which assesses expected outcomes (e.g., "using AI for this task would save time"; akin to utilitarian reasoning), and rule-based reasoning, which evaluates the intrinsic values of the action itself (e.g., "having humans/AI perform this task would be inherently wrong"; akin to deontological reasoning) (Cushman 2013; Cheung, Maier, and Lieder 2024). We then analyze the moral frameworks participants apply, drawing on moral foundations theory (Graham et al. 2011, 2008), to identify the dimensions they prioritize in decision making. Finally, to understand conditions under which participants might flip their decisions, we employ three dimensions based on prior studies (Solaiman et al. 2023; Mun et al. 2024): functionality (system capabilities like performance, bias, and privacy), usage (context of system integration, such as supervision, misuse, or unintended use), and societal impact (effects on individuals, communities, and society, such as job loss and over-reliance).

Our empirical results show general higher acceptance of personal health use cases over labor-replacement. While participants' acceptability judgments decreased with increased entry-level education and risk for each category respectively, professional use cases display more variability and disagreements across judgments (RQ1). Acceptability significantly varied among demographic groups and levels of AI literacy, with lower acceptability observed particularly among non-male participants and those familiar with AI ethics (RQ2). Finally, our results show varying distribution of reasoning types across acceptability decisions, with rule-based reasoning being associated with negative acceptance and unified reasoning types showing higher agreement. Further qualitative analysis revealed participants' normative assumptions about AI, humanness, and society—for example, viewing empathy as essential to humanness but lacking in AI (RQ3).

Our findings shed novel light onto the diversity of people's acceptability and reasoning of AI uses in distinct domains and risk levels. We conclude with a discussion highlighting three key implications for future researchers, practitioners, and policymakers working on advancing ethical and responsible AI development: first, diverse methodologies are needed to effectively analyze use cases and their characteristics; second, involving diverse stakeholders is crucial for assessing the acceptability of AI applications, particularly in workplaces; and third, further investigation into human reasoning processes about AI, notably rule-based reasoning, is needed to inform consensus-building in policy making.

2 Related Works

While there were many efforts towards ethical AI development and deployment by academics (Kieslich, Diakopoulos, and Helberger 2023; Bernstein et al. 2021; Lin et al. 2020), industry (OpenAI 2022; Deng, Barocas, and Vaughan 2024), and government (The White House 2023), they have largely lacked diverse public inputs. To address this gap, many works from both academia and civil society have sought to meaningfully engage lay people in assessing the impact of specific AI use cases. Prior works have focused on anticipating harms (Buçinca et al. 2023) and impacts (Kieslich, Diakopoulos, and Helberger 2023; ada 2023) through participatory foresight, uncovering diverse, sometimes diverging, viewpoints about AI biases and values (Kingsley et al. 2024; Jakesch et al. 2022; Kapania et al. 2022), and governance efforts of AI (Zhang and Dafoe 2020). However, to the best of our knowledge, only Mun et al. considered development decisions by diverse lay-users with an option of not developing a use cases. Among other findings, these prior works from AIES and broader Responsible AI venues revealed a substantial amount of variations in perceptions primarily among demographic lines (e.g., gender, race, political leaning) regarding the desired behavior of AI.

However, little attention has been given to identifying reasoning of participants over AI use cases. While some works have identified decision variations under ambiguous ethical implications of decisions made by AI for certain tasks (e.g., self driving cars (Awad et al. 2018), medical AI (Chen et al. 2023), predictive analysis (Barocas and Selbst 2016)) and inherent value conflicts (Jakesch et al. 2022), these works have not focused on self-reported reasoning. Our work addresses gap by closely examining the **detailed, self-reported reasoning processes of lay people** regarding the acceptability of AI use cases without explicitly guiding towards outcome-based (i.e., utilitarian) or value-based (i.e., deontological) reasoning, allowing participants to freely choose and express their deliberation process. Please refer to Appendix A for extended discussion of related works.

3 Study Design and Data Collection

To answer our research questions on how and why people judge AI use cases as acceptable, we conducted a survey-based study with demographically diverse participants. In this section, we discuss the selection of use cases (§ 3.1), survey design (§ 3.2), and data collection details and participant demographics (§ 3.3).

3.1 Use Cases

To answer RQ1 which examines the impact of different characteristics AI use cases on judgments and decision making processes, we carefully crafted ten different AI use cases as vignettes. We first chose two broad application categories frequently mentioned by lay-users in previous works (Kieslich, Helberger, and Diakopoulos 2024; Mun et al. 2024): **AI in labor-replacement** where AI takes on a role in society thus far done by a human as a profession (e.g., Lawyer AI), and **AI in personal health**, where participants could uniformly consider themselves as AI users. We systematically developed five use cases for each category, varying by required education level for labor-replacement applications and by EU AI Act-assigned risk level for personal health applications.

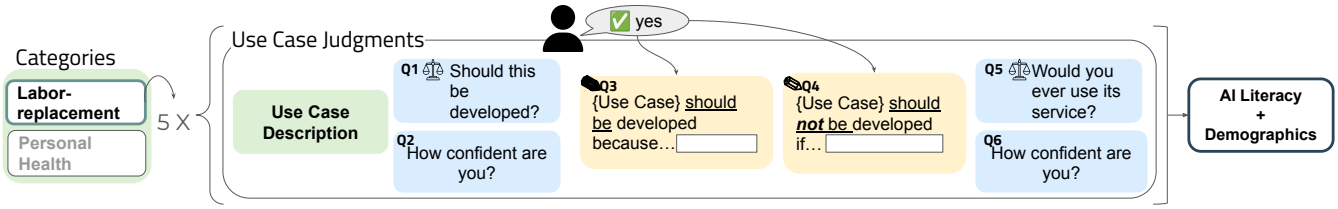

Figure 1: Five professional or personal use cases are presented in a random order. For each use case, we ask multiple-choice questions about its development and confidence levels (Q1, Q2), free-text questions on rationale and decision-switching conditions (Q3, Q4), and multiple-choice questions on usage and confidence (Q5, Q6). These are followed by questions on AI literacy and demographics.

Use Case	Factor	Description
Labor-replacement Use		
Lawyer	Doctoral/Prof	Digital legal advice
Elementary Teacher	Bachelor	Teaches elementary students
IT Support	Some college	Maintains networks; tech support
Eligibility Interviewer	High school	Determines benefit eligibility
Telemarketer	None	Calls to sell/solicit
Peronal Health Use		
Digital Medical Advice	High Risk	Medical advice pre-consultation
Lifestyle Coach	High/Limited	Personalized wellness advice
Health Research	Limited	Summarizes personal health info
Nutrition Optimizer	Lim./Low	Personalized meal/nutrition tips
Flavorful Swaps	Low	Suggests healthy food alternatives

Table 1: Study use cases by category. Descriptions are abbreviated. See Appendix B.2 for full descriptions.

Labor-replacement Use Case Scenarios For the first area of focus, AI in labor replacement, we collected jobs listed in the U.S. census bureau[3] and sorted them according to entry level education required as stated in the census. We chose education level as it has been tightly linked to socioeconomic and occupational status (Svensson 2006; Evetts 2006). We selected jobs that have a large portion of digital or intellectual components with minimal requirement for embodiment resulting in following five professional roles: Lawyer, Elementary school teacher, IT support specialist, Government support eligibility interviewer, and Telemarketer. See Table 1 for further details.

Personal Health Use Case Scenarios To understand the acceptability of different health applications in personal life, we drew from use cases written by participants from prior works (Mun et al. 2024; Kieslich, Helberger, and Diakopoulos 2024) to systematically craft use cases which varied by risk levels according to EU AI Act. We ensured accurate reflection of the risk levels through iterative refinement of descriptions and agreement with categories assigned by GPT-4, following Herdel et al.. See Table 1 for further details.

3.2 Survey Design

Our survey presents participants with five use case descriptions in random order, all from randomly assigned category, labor-replacement or personal health (see § 3.1 for details). After each description, participants answer: "Do you think a technology like this should be developed?" (Q1) and then, "How confident are you in your above answer?" (Q2). To allow for examining of their reasoning, participants then provide open-text rationales by finishing the sentence, "[Use Case] should [not] be developed because..." (Q3), adjusted dynamically depending on their answer to Q1. Participants also described that they would switch their opinion on acceptability of development of the use case: however, "[Use Case] should [not] be developed if..." (Q4; also dynamically rephrased based on Q1 answer). Subsequently, they answer, "If [Use Case] existed, would you use its service?" (Q5) and express confidence with, "How confident are you in your above answer?" (Q6). Refer to Table 6 in the Appendix for the exact wording of the questions.

Collecting Participant Characteristics Following the main survey, we asked participants questions about their AI literacy level and demographics to explore various factors affecting perception of AI acceptance (RQ2). We adopted a shortened version of AI literacy questionnaires from previous works (Wang, Rau, and Yuan 2023; Mun et al. 2024) with four AI literacy aspects, *AI awareness, usage, evaluation, and ethics*, and two additional questions for *generative AI, usage frequency and familiarity with limitations*. We collected demographic information of the participants such as *race, gender, age, sexual orientation, religion, employment status, income, and level of education*; see Appendix 7 for detailed list of questions. Additionally, we collected information about *discrimination chronicity*, i.e., prolonged experiences of everyday discrimination, of their discrimination experiences (if any) following Kingsley et al..

3.3 Data Collection and Participant Demographics

We used Prolific[4] to recruit participants. To represent diverse sample, we stratified our recruitment by the ethnicity categories (White, Mixed, Asian, Black, and Other) and age (18-48, 49-100) as provided by Prolific. We also added criteria for quality such as survey approval rating and number of previous surveys completed. Our study was approved by IRB at

[3]https://www.bls.gov/ooh/occupation-finder.htm

[4]https://www.prolific.com

our institutions, and we paid 12 USD/hour. Our final sample consisted of 197 participants across two categories, with professional usage assigned to 100 participants and personal to 97. See Appendix B.3 for further details on participants.

4 Acceptability & Reasoning Analysis Methods

Our surveys consisted of both multiple choice (numerical) and open-text questions designed to answer our research questions. In this section, we detail our process for numerical (§ 4.1) and open-text (§ 4.2) analysis.

4.1 Multiple Choice Analysis

We analyzed the judgment and confidence ratings by mapping judgment (Q1, Q5) to 1 ("Should be developed", "Would use") or -1 ("Should not be developed", "Would not use") and confidence (Q2, Q6) to a scale from 1 to 5. We used numerically converted judgment, confidence, and combined (judgment×confidence; -5 to 5) values as dependent variables in our analysis. We used repeated-measures ANOVAs to understand the differences in mean responses between conditions/groups and linear mixed effects regression models (lmer) to better understand the effects of specific factors. We included a subject-specific random effect when using ANOVA and regression models and added a use-case-specific random effect when applicable. We factorized demographic responses for analysis with the exception of discrimination chronicity, which we aggregated to a numerical value (Kingsley et al. 2024; Michaels et al. 2019). We also converted responses to AI literacy questions to numerical values for analysis.

4.2 Open-response Analysis

Background: Moral Decision Making To understand decision-making in AI use cases, we draw on moral psychology and dual system theory. We examine two decision-making systems: cost-benefit reasoning, which assesses outcomes and consequences, and rule-based reasoning, focusing on norms, rules, and virtues (Cushman 2013; Cheung, Maier, and Lieder 2024). These correspond to utilitarian reasoning (maximizing good) and deontological reasoning (adhering to moral duties and rights), respectively. Additionally, we apply moral foundations theory (Graham et al. 2008) to identify values and potential moral conflicts in AI development.

To assess the reasoning methods used by the participants, we analyzed the open-text responses on elaborations to their decisions (Q3) and circumstances in which their decisions would switch (Q4) along the following three dimensions: reasoning types (cost-benefit, rule-based, both, unclear), reference to moral foundations[5] (Care, Fairness, Purity, Authority, Loyalty), and switching conditions (Functionality, Usage, Societal Impact). By analyzing reasoning types and moral values reflected in the participants' justifications, we aim to characterize *how* participants made their decisions, and by analyzing various factors such as primary concerns in switching condition, we aim to discover *what* aspects were salient for the participants in their decisions.

Classification and Aggregation We classified participants' responses to Q3 (elaboration of judgment) and Q4 (conditions for switching decisions), totaling 985 samples for each question, using OpenAI's gpt-4o[6]. To validate the model's classification performance, results were compared with a reference set of 100 samples annotated by three annotators, comprised by members of the research team and a professional annotator. Initially, each annotator independently assessed the data, and then consensus was reached through discussion to establish a gold standard set. The inter-rater agreement between the gold standard and gpt-4o's annotations was evaluated using Gwet's AC1 metric, chosen for its robustness with infrequent labels (Wongpakaran et al. 2013). While the agreement levels varied, ranging from almost perfect to moderate (0.98–0.57), all dimensions had above substantial agreement except Societal Impact. Annotations for cost-benefit reasoning, rule-based reasoning, and authority reached near-perfect agreement. Due to minimal occurrences in both human and LLM annotations, the moral foundation dimension Loyalty was excluded from further analysis. The annotations were conducted based on presence or absence of the values and were converted into binary format for statistical analysis. See Appendix D for details on agreement, annotation settings, and their limitations.

5 Findings

Our work aims to uncover variations in acceptability of AI use cases and factors and reasoning processes that underlie these judgments. In this section, we discuss our findings about the judgments of the AI use cases (§5.1), personal factors that may influence the decision such as demographics and AI literacy (§5.2), and factors in rationales that could uncover reasoning processes that lead to judgments (§5.3).

5.1 RQ1. Use Case Perceptions & Disagreements

In our analysis, we investigated the effects of use cases on participants' judgments using our ten use case vignettes. Overall, acceptability statistically differed among the two categories ($t_{\text{DEV}}(983) = -9.05, p < .001$; $t_{\text{USAGE}}(983) = -5.50, p < .001$). Notably, personal health use cases had higher acceptability ($M_{\text{DEV}} = 0.68, SD_{\text{DEV}} = 0.74$; $M_{\text{USAGE}} = 0.51, SD_{\text{USAGE}} = 0.86$) than labor-replacement use cases ($M_{\text{DEV}} = 0.18, SD_{\text{DEV}} = 0.99$; $M_{\text{USAGE}} = 0.18, SD_{\text{USAGE}} = 0.98$). See Figure 7 in Appendix C for additional category comparison results.

Labor-replacement Use Cases Exploring specific use cases within the labor-replacement category (Figure 2), we observed that Elementary School Teacher AI ($M_{\text{DEV}} = -0.24, SD_{\text{DEV}} = 0.98$;

[5]We used the five foundational dimensions: Care, Fairness, Loyalty, Authority, and Purity. Although these dimensions have been updated to encompass a broader range of values beyond WEIRD populations (Atari et al. 2023), we selected this version for survey brevity.

[6]`gpt-4o-2024-11-20`

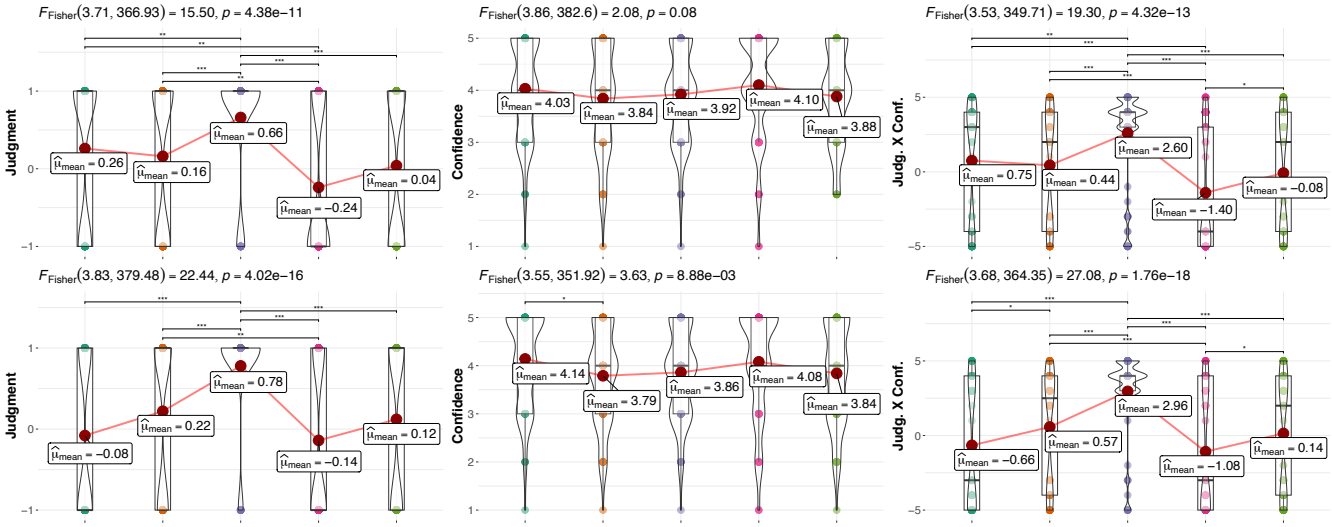

Figure 2: Labor-replacement use case means and distributions of numerically converted Judgment, Confidence, and Judgment×Confidence. First row shows results for development decisions (Q1, Q2) and second row shows results for usage decisions (Q5, Q6). ANOVA results for use cases are shown above each panel. Within subject test was performed using Student's t-test with Holm correction. * denotes following significant p-values: $^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$. Use cases from left to right (n=100): Telemarketer, Gov. Elig. Interviewer, IT Support Specialist, Elm. School Teacher, and Lawyer.

$M_{\text{USAGE}} = -0.14, SD_{\text{USAGE}} = 1.00$) had the lowest acceptability for both types of judgments followed by Lawyer AI ($M_{\text{DEV}} = 0.04, SD_{\text{DEV}} = 1.00$) and Telemarketer AI ($M_{\text{USAGE}} = -0.08, SD_{\text{USAGE}} = 1.00$) where their near zero mean suggest disagreement within judgments. Interestingly, IT Support Specialist AI had the highest acceptability ($M_{\text{DEV}} = 0.66, SD_{\text{DEV}} = 0.76$; $M_{\text{USAGE}} = 0.78, SD_{\text{USAGE}} = 0.63$) despite human replacement potential and median education level.

Personal Health Use Cases In personal health use scenarios, Digital Medical Advice AI ($M_{\text{DEV}} = 0.34, SD_{\text{DEV}} = 0.95$; $M_{\text{USAGE}} = 0.24, SD_{\text{USAGE}} = 0.98$), reflecting high risk usage, consistently had lower acceptance across judgment types, compared to all other use cases. Nutrition Optimizer ($M_{\text{DEV}} = 0.86, SD_{\text{DEV}} = 0.92$; $M_{\text{USAGE}} = 0.69, SD_{\text{USAGE}} = 0.73$) had the highest mean acceptance across both acceptability judgments. Interestingly, unlike the labor-replacement use cases which had slightly higher acceptance for usage, personal use cases had lower acceptance for usage in general compared to development.

Use Case Variations When selecting use cases, we used two underlying variations: entry level of education required for labor-replacement use cases and EU AI risk levels for personal health. As risk levels and required education increased, we observe consistent negative effects on judgments, with personal health use cases showing stronger effects ($\beta_{\text{DEV}} = -0.11, p < .001$; $\beta_{\text{USAGE}} = -0.10, p < .001$) compared to labor-replacement scenarios ($\beta_{\text{DEV}} = -0.08, p < .01$), where only development judgments were significantly associated. Confidence ratings showed a small but significant decrease with increasing risk levels in personal health use cases ($\beta_{\text{DEV}} = -0.08, p < .001$; $\beta_{\text{USAGE}} = -0.09, p < .001$), while labor-replacement use cases showed no significant impact on confidence.

	DEV($\beta(SE)$)		USAGE($\beta(SE)$)	
	Judg.	Conf.	Judg.	Conf.
Labor-replacement				
Coeff.	$\mathbf{-0.08}^{**}(0.03)$	$-0.00(0.02)$	$0.00(0.03)$	$-0.03(0.03)$
β_0	$\mathbf{0.43}^{***}(0.10)$	$\mathbf{3.97}^{***}(0.10)$	$0.17(0.10)$	$\mathbf{4.04}^{***}(0.10)$
Personal Health				
Coeff.	$\mathbf{-0.11}^{***}(0.02)$	$\mathbf{-0.08}^{***}(0.02)$	$\mathbf{-0.10}^{***}(0.02)$	$\mathbf{-0.09}^{***}(0.02)$
β_0	$\mathbf{1.02}^{***}(0.08)$	$\mathbf{4.20}^{***}(0.10)$	$\mathbf{0.80}^{***}(0.09)$	$\mathbf{4.05}^{***}(0.11)$

$^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$

Table 2: Mixed-effects models: use case factor effects (estimate; SE, β_0 denotes intercept) for judgment and confidence, split by labor-replacement/personal health. Use case variations are numerically coded from 1 (lowest risk/education) to 5 (highest risk/education). Bold indicates $p < 0.05$.

Disagreements We compare the standard deviation of judgments weighted by confidence to understand possible disagreements and their strength among use cases. Interestingly, the use cases with four highest disagreements in both judgments were all labor-replacement uses in order of Telemarketer ($SD_{\text{DEV}} = 4.08$; $SD_{\text{USAGE}} = 4.21$), Elementary School Teacher ($SD_{\text{DEV}} = 3.99$; $SD_{\text{USAGE}} = 4.09$), Lawyer ($SD_{\text{DEV}} = 4.00$; $SD_{\text{USAGE}} = 3.99$), and Government Eligibility Interviewer AI ($SD_{\text{DEV}} = 3.96$; $SD_{\text{USAGE}} = 3.89$). These four use cases were followed by Digital Medical Advice AI ($SD_{\text{DEV}} = 3.80$, $SD_{\text{USAGE}} = 3.83$). The use cases with the lowest disagreements were surprisingly Nutrition Optimizer ($SD_{\text{DEV}} =$

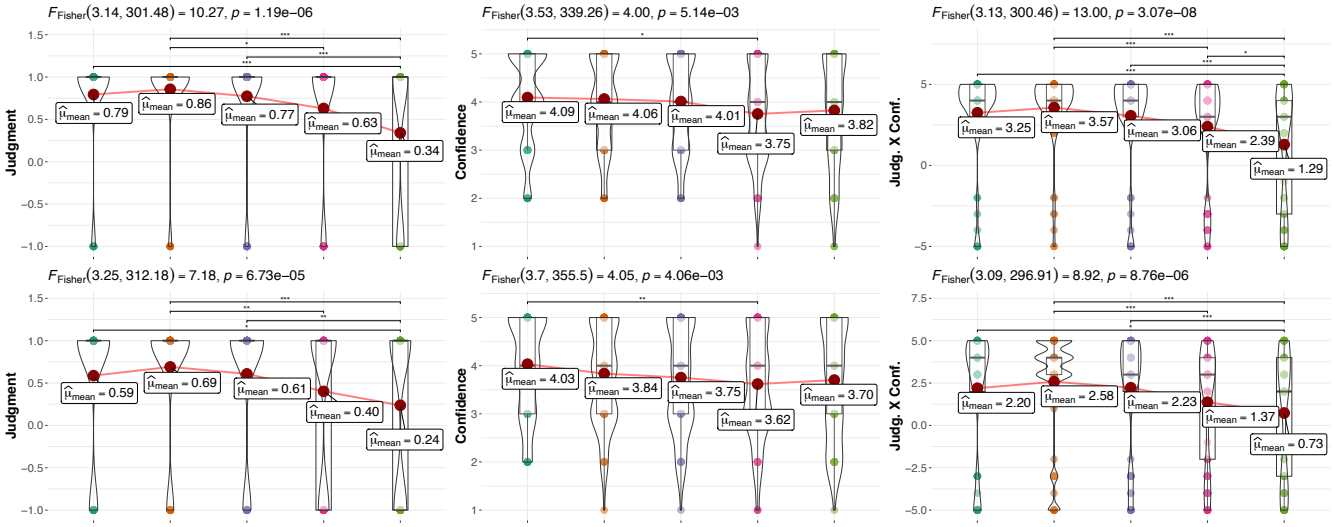

Figure 3: Personal health use case means and distributions of numerically converted Judgment, Confidence, and Judgment×Confidence. First row shows results for development decisions (Q1, Q2) and second row shows results for usage decisions (Q5, Q6). ANOVA results for use cases are shown above each panel. Within subject test was performed using Student's t-test with Holm correction. * denotes following significant p-values: $^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$. Use cases from left to right (n=97): Flavorful Swaps, Nutrition Optimizer, Health Research, Lifestyle Coach, and Medical Advice.

2.16; $SD_{\text{USAGE}} = 3.03$) followed by IT Support Specialist AI ($SD_{\text{DEV}} = 3.08$; $SD_{\text{USAGE}} = 2.65$).

RQ1 Takeaways Our results show that there are significant variation in judgments based on use case characteristics such as category, EU-defined risk level, and, to some extent, required education level for labor-replacement cases. The uniquely negative response to the Elementary School Teacher AI highlights potential concerns specific to care work. High disagreement in labor-replacement scenarios underscores the need for cautious integration of AI into existing roles. In contrast, the consistent positive judgments for IT Support Specialist AI suggest that not all labor-replacement use cases are viewed equally, indicating the need for nuanced understandings of acceptability. We further examine this variability in § 5.3.

5.2 RQ2. Impact of Personal Factors on Acceptability Judgment

Demographic Factors As shown in Table 3, several demographic factors significantly influenced use case judgments.[7] Across both categories of use cases, certain age groups were positively associated with confidence in usage: 25-34 ($\beta_{\text{USAGE}} = 0.41$, $p < .05$) and 55-64 ($\beta_{\text{USAGE}} = 0.48$, $p < .05$). Race also had notable influences; specifically, Asian participants exhibited significantly lower confidence in both development and usage judgments ($\beta_{\text{DEV}} = -0.37$, $p < .01$; $\beta_{\text{USAGE}} = -0.33$, $p < .05$), particularly in labor-replacement contexts.

[7]Prior to analysis, we observed that the independent variables in the model had less than 0.5 correlation, except for age 65+ and Retired employment status.

Gender emerged as a crucial determinant, with non-male participants consistently showing negative judgments ($\beta_{\text{DEV}} = -0.29$, $p < .001$; $\beta_{\text{USAGE}} = -0.33$, $p < .001$), indicating potential discrepancies in perception or experience with AI applications. Liberal views, especially among those identifying as strongly liberal, were associated with negative judgments across both categories of use cases ($\beta_{\text{DEV}} = -1.16$, $p < .05$; $\beta_{\text{USAGE}} = -1.51$, $p < .01$), suggesting a skeptical stance towards AI's prevalence and role. Employment hours also contributed, with individuals working 40+ hours per week displaying a positive association with development ($\beta_{\text{DEV}} = 0.25$, $p < .05$), suggesting more exposure or reliance on AI use cases. High experience of discrimination chronicity was significantly related to lower acceptance of development ($\beta_{\text{DEV}} = -0.36$, $p < .05$). See Table 25 in the Appendix for ANOVA results.

AI Literacy We identified a correlation greater than 0.5 among three AI literacy aspects: awareness, usage, and evaluation. To avoid multicollinearity, we aggregated them into a single factor, AI Skills. As shown in Table 4, understanding of AI Ethics was associated with lower acceptability for both personal health ($\beta_{\text{DEV}} = -0.05$, $p < .001$; $\beta_{\text{USAGE}} = -0.23$, $p < .001$) and labor-replacement ($\beta_{\text{DEV}} = -0.04, p < .05$). However, across both categories, high Generative AI Usage Frequency resulted in higher acceptance (Labor-replacement – $\beta_{\text{DEV}} = 0.14$, $p < .01$; $\beta_{\text{USAGE}} = 0.18$, $p < .001$, Personal Health – $\beta_{\text{DEV}} = 0.15$, $p < .001$; $\beta_{\text{USAGE}} = 0.19$, $p < .001$). Notably, for personal health use cases, AI Skills was positively associated with confidence of judgments ($\beta_{\text{DEV, USAGE}} = 0.06$, $p < .05$), while Generative AI Limitation Familiarity was positively associated with confidence for labor-replacement

	DEV (β (SE))			USAGE (β (SE))		
Demographics	Judg.	Conf.	Judg.×Conf.	Judg.	Conf.	Judg.×Conf.
(Intercept)	$\mathbf{0.50}^{*}$ (0.25)	$\mathbf{4.08}^{***}$ (0.32)	1.88 (1.09)	0.51 (0.28)	$\mathbf{3.09}^{***}$ (0.34)	1.34 (1.23)
(Intercept)$_{\texttt{Labor}}$	0.24 (.38)	$\mathbf{4.59}^{***}$ (.49)	1.16 (1.66)	0.71 (.41)	$\mathbf{3.74}^{***}$ (.45)	3.09 (1.74)
(Intercept)$_{\texttt{Pers}}$	0.66 (.30)	$\mathbf{3.76}^{***}$ (.47)	2.33 (1.34)	0.25 (.43)	$\mathbf{2.61}^{***}$ (.55)	−0.16 (1.92)
Age						
25-34	−0.13 (0.13)	0.04 (0.19)	−0.59 (0.58)	−0.22 (0.16)	$\mathbf{0.41}^{*}$ (0.20)	−0.99 (0.69)
55-64	−0.03 (0.16)	0.32 (0.23)	−0.14 (0.69)	0.12 (0.19)	$\mathbf{0.48}^{*}$ (0.24)	0.40 (0.82)
25-34$_{\texttt{Pers}}$	−0.06 (.16)	0.36 (.26)	−0.01 (.70)	−0.14 (.23)	$\mathbf{0.76}^{*}$ (.30)	−0.10 (.103)
Race						
Asian	0.17 (0.10)	$\mathbf{-0.37}^{**}$ (0.14)	0.73 (0.44)	0.10 (0.12)	$\mathbf{-0.33}^{*}$ (0.15)	0.42 (0.52)
Black	0.07 (0.10)	0.24 (0.14)	0.49 (0.43)	0.07 (0.12)	$\mathbf{0.32}^{*}$ (0.15)	0.53 (0.51)
Mixed	0.19 (0.13)	0.16 (0.18)	0.85 (0.56)	−0.13 (0.15)	$\mathbf{0.43}^{*}$ (0.19)	−0.12 (0.66)
Asian$_{\texttt{Labor}}$	$\mathbf{0.40}^{**}$ (.15)	$\mathbf{-0.41}^{*}$ (.20)	$\mathbf{1.67}^{*}$ (.65)	0.20 (.16)	$\mathbf{-0.44}^{*}$ (.19)	0.59 (.68)
Asian$_{\texttt{Pers}}$	0.01 (.12)	$\mathbf{-0.54}^{**}$ (.20)	−0.05 (.56)	0.00 (.18)	−0.37 (.24)	0.07 (.82)
Black$_{\texttt{Pers}}$	0.12 (.12)	0.35 (.20)	0.94 (.55)	0.02 (.18)	$\mathbf{0.59}^{*}$ (.23)	0.65 (.80)
Gender						
Non-male	$\mathbf{-0.29}^{***}$ (0.07)	−0.05 (0.10)	$\mathbf{-1.29}^{***}$ (0.32)	$\mathbf{-0.33}^{***}$ (0.09)	0.10 (0.11)	$\mathbf{-1.36}^{***}$ (0.38)
Non-male$_{\texttt{Labor}}$	$\mathbf{-0.48}^{***}$ (.11)	0.03 (.15)	$\mathbf{-2.11}^{***}$ (.48)	$\mathbf{-0.52}^{***}$ (.12)	0.06 (.14)	$\mathbf{-2.25}^{***}$ (.50)
Political View						
Str. liberal	−0.19 (0.11)	−0.28 (0.16)	$\mathbf{-1.16}^{*}$ (0.49)	$\mathbf{-0.34}^{*}$ (0.13)	0.14 (0.17)	$\mathbf{-1.51}^{**}$ (0.59)
Str. Liberal$_{\texttt{Labor}}$	−0.25 (.18)	0.02 (.24)	−1.08 (.76)	$\mathbf{-0.42}^{*}$ (.18)	$\mathbf{0.58}^{**}$ (.22)	$\mathbf{-1.63}^{*}$ (.79)
Str. Liberal$_{\texttt{Pers}}$	−0.15 (.16)	$\mathbf{-0.53}^{*}$ (.25)	−1.14 (.70)	−0.18 (.23)	−0.36 (.30)	−1.01 (.102)
Liberal$_{\texttt{Pers}}$	−0.08 (.11)	$\mathbf{-0.52}^{**}$ (.18)	−0.55 (.50)	−0.07 (.17)	−0.39 (.21)	−0.38 (.74)
Education						
Advanced$_{\texttt{Labor}}$	$\mathbf{0.43}^{*}$ (.19)	−0.48 (.26)	1.39 (0.83)	0.31 (0.20)	−0.40 (0.24)	1.03 (0.87)
Employment						
40+ hrs	$\mathbf{0.25}^{*}$ (0.11)	0.07 (0.16)	$\mathbf{1.13}^{*}$ (0.51)	0.09 (0.14)	0.01 (0.17)	0.73 (0.60)
Discrimination						
High$_{\texttt{Labor}}$	$\mathbf{-0.36}^{*}$ (.18)	0.14 (.25)	$\mathbf{-1.79}^{*}$ (.79)	−0.13 (.19)	0.27 (.23)	−0.39 (.82)

$^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$

Table 3: Coefficients, Standard Errors, and Significance of Demographic Factors. Models regress decision metrics on demographic factors, with random effects for subjects and use cases. Only significant factors are reported. The intercept represents the dominant demographic group (White, Christian, Male), the lowest natural ordering (18-24, Not Employed), and median values (Moderate, Associate's degree). Subscripts `Labor` and `Pers` indicate labor-replacement and personal health use cases.

	DEV (β (SE))			USAGE (β (SE))		
AI Literacy	Judg.	Conf.	Judg.×Conf.	Judg.	Conf.	Judg.×Conf.
Labor-replacement						
(Intercept)	0.07 (0.30)	$\mathbf{3.13}^{***}$ (0.33)	−0.49 (1.27)	−0.50 (0.32)	$\mathbf{3.46}^{***}$ (0.34)	−2.44 (1.35)
AI Ethics	$\mathbf{-0.04}^{*}$ (0.02)	0.02 (0.02)	−0.15 (0.08)	−0.00 (0.02)	−0.01 (0.02)	−0.06 (0.08)
Gen AI Usage Freq.	$\mathbf{0.14}^{**}$ (0.04)	−0.00 (0.05)	$\mathbf{0.59}^{**}$ (0.19)	$\mathbf{0.18}^{***}$ (0.05)	−0.07 (0.05)	$\mathbf{0.80}^{***}$ (0.20)
Gen AI Limit. Familiarity	−0.09 (0.07)	$\mathbf{0.20}^{*}$ (0.08)	−0.38 (0.28)	−0.06 (0.07)	$\mathbf{0.18}^{*}$ (0.08)	−0.38 (0.30)
Personal Health						
(Intercept)	$\mathbf{0.87}^{***}$ (0.22)	$\mathbf{2.72}^{***}$ (0.40)	$\mathbf{2.81}^{**}$ (1.00)	0.49 (0.30)	$\mathbf{2.44}^{***}$ (0.45)	1.17 (1.30)
AI Skills	0.00 (0.01)	$\mathbf{0.06}^{*}$ (0.02)	0.07 (0.06)	0.02 (0.02)	$\mathbf{0.06}^{*}$ (0.03)	$\mathbf{0.15}^{*}$ (0.08)
AI Ethics	$\mathbf{-0.05}^{***}$ (0.01)	0.01 (0.03)	$\mathbf{-0.23}^{***}$ (0.07)	$\mathbf{-0.06}^{**}$ (0.02)	$\mathbf{0.07}^{*}$ (0.03)	$\mathbf{-0.26}^{**}$ (0.09)
Gen AI Usage Freq.	$\mathbf{0.15}^{***}$ (0.03)	0.06 (0.06)	$\mathbf{0.62}^{***}$ (0.15)	$\mathbf{0.19}^{***}$ (0.05)	−0.01 (0.07)	$\mathbf{0.79}^{***}$ (0.20)
Gen AI Limit. Familiarity	−0.06 (0.05)	0.00 (0.09)	−0.25 (0.21)	−0.10 (0.06)	−0.10 (0.10)	−0.50 (0.28)

$^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$

Table 4: Coefficients, Standard Errors, and Significance of AI Literacy Factors. Models regress decision metrics on AI literacy factors, with random effects for subjects and use cases. Only significant factors are reported.

usage ($\beta_{\texttt{DEV}} = 0.20, p < .05$; $\beta_{\texttt{USAGE}} = 0.18, p < .05$).

RQ2 Takeaways Our findings highlight the significant role of lived experiences and backgrounds, such as age, race, gender, political view, employment, and discrimination experience, in shaping attitudes toward AI usage. Moreover, our results indicate that different understandings of and experiences with AI can impact judgments of acceptability, corroborating previous findings (Kramer et al. 2018).

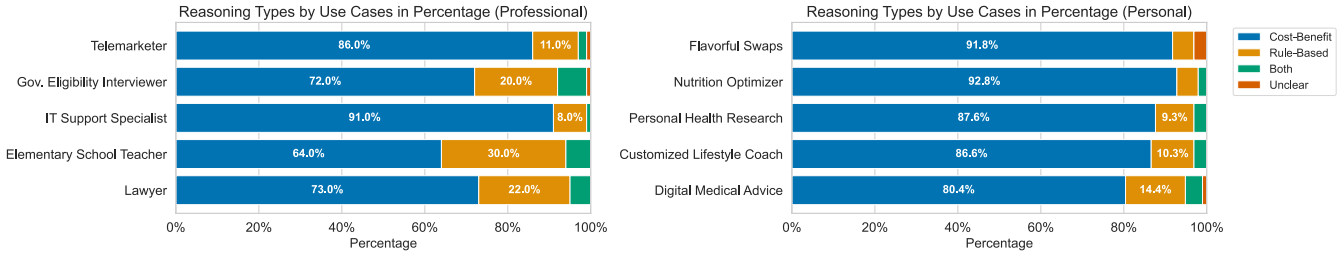

Figure 4: Percentage of reasoning types (cost-benefit and rule-based) by use cases in participant provided rationales (Q3, Q4).

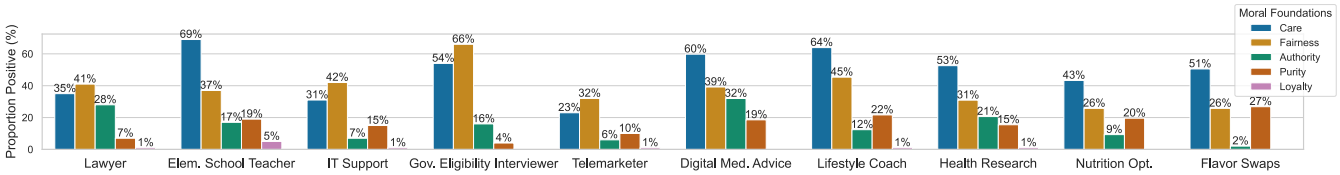

Figure 5: Proportion (%) of presence of moral foundations in participant's rationale responses (Q3, Q4) aggregated by use case.

	DEV (β (SE))			USAGE (β (SE))		
	Judg.	Conf.	Judg.×Conf.	Judg.	Conf.	Judg.×Conf.
Reasoning Type						
Cost-benefit	$\mathbf{0.32}^{**}$ $(\mathbf{0.10})$	0.15 (0.15)	$\mathbf{1.31}^{***}$ $(\mathbf{0.39})$	$\mathbf{-0.10}^{*}$ $(\mathbf{0.04})$	$\mathbf{0.32}^{*}$ $(\mathbf{0.16})$	$\mathbf{2.40}^{***}$ $(\mathbf{0.52})$
Rule-based	$\mathbf{-0.46}^{***}$ $(\mathbf{0.09})$	$\mathbf{0.43}^{**}$ $(\mathbf{0.14})$	$\mathbf{-1.61}^{***}$ $(\mathbf{0.35})$	-0.06 (0.04)	0.25 (0.14)	-0.88 (0.48)
Moral Value						
Care	0.05 (0.04)	-0.06 (0.06)	0.15 (0.15)	-0.00 (0.02)	$\mathbf{-0.15}^{*}$ $(\mathbf{0.06})$	-0.37 (0.21)
Fairness	$\mathbf{0.14}^{***}$ $(\mathbf{0.04})$	$\mathbf{-0.17}^{**}$ $(\mathbf{0.06})$	$\mathbf{0.53}^{***}$ $(\mathbf{0.16})$	0.00 (0.02)	-0.08 (0.07)	$\mathbf{0.71}^{**}$ $(\mathbf{0.22})$
Authority	$\mathbf{-0.20}^{***}$ $(\mathbf{0.05})$	-0.06 (0.08)	$\mathbf{-0.77}^{***}$ $(\mathbf{0.20})$	-0.00 (0.02)	-0.08 (0.08)	-0.54 (0.28)
Switching Condition						
Usage	$\mathbf{0.12}^{***}$ $(\mathbf{0.01})$	0.02 (0.01)	$\mathbf{0.56}^{***}$ $(\mathbf{0.02})$	$\mathbf{0.24}^{***}$ $(\mathbf{0.00})$	$\mathbf{-0.04}^{***}$ $(\mathbf{0.01})$	
Societal Impact	-0.08 (0.04)	0.00 (0.07)	$\mathbf{-0.39}^{*}$ $(\mathbf{0.17})$	0.03 (0.02)	-0.12 (0.07)	$\mathbf{-0.74}^{**}$ $(\mathbf{0.23})$
(Intercept)	0.09 (0.11)	$\mathbf{3.88}^{***}$ $(\mathbf{0.16})$	0.18 (0.44)	0.16^{***} (0.04)	$\mathbf{3.81}^{***}$ $(\mathbf{0.17})$	-0.53 (0.64)

$^{***}p<0.001$; $^{**}p<0.01$; $^{*}p<0.05$

Table 5: Coefficients (SE) and significance of rationale factors. All factors coded as binary. Only showing significant results.

5.3 RQ3. Factors in Participant Rationale

To deepen our analysis of use case acceptability, we examined open-text rationales (Q3) and decision-switching conditions (Q4) related to development judgments (Q1).

Decision-making Types As we defined in § 4.2, we focus on two distinct reasoning types for decision-making: cost-benefit reasoning, which emphasizes outcomes (e.g., *"it gives more people access to medical advice and treatment"*, P365), and rule-based reasoning, which reflects values inherent in the action itself (e.g., should not be developed because *"human interaction is better"*, P249). As shown in Figure 4, generally, participants used more cost-benefit reasoning, especially for the IT Support Specialist (91.0%) and Nutrition Optimizer (92.8%) use cases. This result is particularly interesting as we observed these two use cases to have the lowest disagreement (see § 5.1), suggesting unified reasoning type may lead to more consistent judgments. On the other hand, rationales for Elementary School Teacher AI contained most percentage of rule-based reasoning (30.0%) followed by Lawyer AI (22.0%), both of which were the two use cases with lowest development acceptability.

Through further analysis of the influence of rationale factors on judgments using a mixed-effects model (Table 5), we found that use of cost-benefit reasoning in rationale is positively associated with development acceptability ($\beta_{\text{DEV}} = 0.32, p < .01$), while rule-based reasoning is negatively associated ($\beta_{\text{DEV}} = -0.46, p < .001$). Interestingly, for usage, cost-benefit reasoning was negatively associated with acceptability ($\beta_{\text{USAGE}} = -0.10, p < .05$) but positively for judgment weighted by confidence ($\beta_{\text{USAGE}} = 2.40, p < .001$). This suggests that, although cost-benefit considerations may reduce usage acceptance, they boost confidence when judgments are favorable.

The significant negative association of rule-based reasoning with judgments indicate that for certain contexts, the inherent action of using AI is viewed negatively. Participants cited diverse concerns related to AI itself (e.g., *"because it would not have human sympathy"*, P16), human needs (e.g., *"Humans need human interactions in order to learn properly"*, P44), societal impact (e.g., *"Overreliance on AI ..."*, P132), and morality (e.g., *"Having artificial intelligence try*

and sell you things is immoral", P13). These results suggest that AI's acceptability heavily lies in its positive outcomes but can be outweighed by established rules.

Moral Foundations Beyond reasoning types, we explored moral foundations to provide insights into what values are relevant for AI use case decisions (Figure 5). For example, P12 responded that Elementary School Teacher AI use case should be developed because it *"could give elementary schooling to children who are bed ridden..."*, which was annotated with both values of Care (focusing on the well-being of bed-ridden children) and Fairness (focusing on fair access to education). Upon analysis, Care (i.e., dislike of pain of others, feelings of empathy and compassion toward others) was the most prevalent moral foundation in participants' rationales across the use cases in both categories (48%). Interestingly, Care could be invoked in both positive and negative regards for AI, as conveyed by P385 who noted that Customized Lifestyle Coach AI should be developed because *"it may help improve some people's health"* but would change their decision if *"it caused harm to even one person."*

Although Care was the most dominant moral foundation overall, Fairness emerged more prominently in context-specific evaluations such as Lawyer AI (41%) and Government Eligibility Interviewer AI (66%). These results could be due to use case attributes, such as their main purpose and function; as noted by P88, Government Eligibility Interviewer should be developed because *"it might be less biased and therefore more fair in it decisions (sic)"*. Authority was most apparent in participant rationales for Lawyer AI (28%) and Purity for Flavorful Swaps (27%). Moreover, Fairness in rationales had positive associations with acceptance ($\beta_{\texttt{DEV}} = 0.53, p < .001$, $\beta_{\texttt{USAGE}} = 0.71, p < .01$; judgment weighted by confidence; see Table 5).

Switching Conditions We further explored the flexibility of participants' judgments to understand possible mitigation of disagreements through criteria for switching their decisions (Figure 6). Functionality (53%; e.g., Medical Adivce AI should not be developed if it *"consistently or had a high percentage of failure to diagnose correctly."*, P373) was the most commonly noted condition for switching decisions in both directions (positive to negative and vice versa). This was followed by Usage (40%; e.g., Government Eligibility Interviewer should be developed if *"it was only used to read and screen applications but not for making decisions"*), Societal Impact (36%; e.g., Lawyer AI should not be developed if *"it puts too many human lawyers out of work"*, P97), and Not Applicable (7%; e.g., will not change decision).

Interestingly, for labor-replacement use cases Societal Impact (45%) was more frequently mentioned as switching conditions followed by Functionality (43%), whereas Functionality (53%) was more frequently mentioned than Societal Impact (37%) for personal use cases. Frequency of Societal Impact in labor-replacement use cases could be closely linked to concerns of labor replacement: as described by P296, if Elementary School Teacher AI *"was to replace teachers with the ai to save money"*, they would switch their decision from positive to negative. Moreover, for all use cases except Elementary School Teacher AI, participants tended to switch from positive to negative decisions for lack of functionality reasons. In contrast, for Elementary School Teacher AI, the most common shift towards acceptability when the use case showed a positive societal impact (38%). As shown in Table 5, mentions of Societal Impact as conditions to switch decisions were more negatively associated with judgments ($\beta_{\texttt{DEV}} = -0.39, p < .05$, $\beta_{\texttt{USAGE}} = -0.74, p < .05$; judgment $\times$ confidence). However, emphasis on Usage ($\beta_{\texttt{DEV}} = 0.12, p < .001$) as a condition to reverse their decisions was positively associated.

RQ3 Takeaways Reasoning types varied by context, with rule-based reasoning more common in contested use cases and negatively associated with acceptability, while cost-benefit reasoning showed a positive association. The moral foundation of Care was especially salient, highlighting its importance in AI judgments. When explaining what might change their decisions, participants most frequently cited Functionality for personal health use cases and Societal Impact for labor-replacement, underscoring the context-dependent nature of these concerns, especially to be considered when mitigating disagreements.

6 Conclusion and Discussion

We conducted a study to understand how and why laypeople perceive various AI use cases as acceptable or not. To achieve this, we developed a survey that gathered judgments and reasoning processes from 197 participants who were demographically diverse and had varying levels of experience with AI. Participants were asked to provide their judgments on the acceptability of AI use cases, along with rationales for their decisions (e.g., "Should / Should not be developed, because...") and conditions that might change their decisions (e.g., "I would switch my decision if..."). The survey covered ten different AI use cases, spanning both personal and professional domains, and included varying levels of risk. Our findings revealed significant variation in the acceptability judgments and reasoning factors based on the domain, risk level, and participants' attributes, such as AI literacy and gender. We discuss the implications of these findings below.

Use Case Perceptions and Disagreements In our study, we explored the varying acceptability of AI across different use cases. Generally, acceptance was lower in scenarios with higher educational requirements and greater EU AI risk levels. Professional use cases displayed more variability, notably with Elementary School Teacher AI, which was uniquely unacceptable. This underscores the necessity for further research into how AI should be developed and integrated, as well as what skills it should have, particularly in fields where empathy and care are crucial (Wu et al. 2024; Kawakami et al. 2024; Borg and Read 2024). In addition, prior research have also highlighted how AI practitioners desire understanding lay people's perception on AI fairness in specific use cases (Sonboli et al. 2021; Deng et al. 2022; Smith, Beattie, and Cramer 2023; Deng et al. 2023). Drawing from prior HCI and AI research (Deng et al. 2025; Lee et al. 2019; Cheng et al. 2019), future researchers and practitioners should explore how to meaningfully connect lay people's use case perceptions with AI developers' workflows.

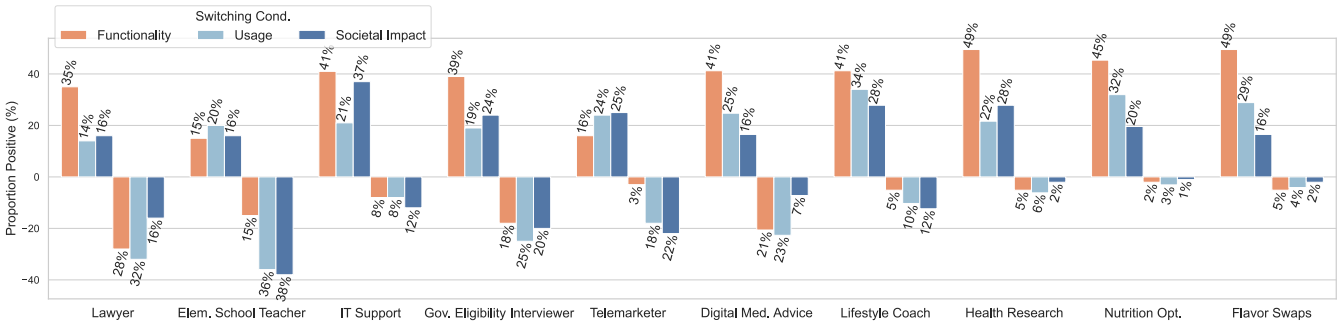

Figure 6: Proportion (%) of presence of switching conditions (Functionality, Usage, Societal Impact) mentioned in participant's switching conditions (Q4) aggregated by use case and divided into positive and negative development acceptability. The total proportion of the switching condition by use case is the sum of both positive and negative bars.

While prior research has emphasized understanding AI consequences (Kieslich, Helberger, and Diakopoulos 2024) and providing tools and processes to uncover impact (Wang et al. 2024b; Buçinca et al. 2023; Deng, Barocas, and Vaughan 2024), our findings reveal a greater presence of rule-based reasoning in contentious use cases, suggesting a need for diverse approaches to understanding AI beyond mere consequence anticipation. Moreover, while care was generally predominant, we observed that fairness gained prominence in Lawyer AI and Government Eligibility Interviewer AI. This variability underscores the importance of considering values in AI evaluation and training (Barocas et al. 2021; Bhardwaj et al. 2024), rather than solely emphasizing functionality, which is the current trend in AI research (Birhane et al. 2022). Additionally, societal impact considerations were more evident in unacceptable use cases, emphasizing the necessity for implementing safety guardrails when deploying AI with significant social implications (Solaiman 2023).

Demographics and AI Literacy In line with prior work (Kingsley et al. 2024; Mun et al. 2024), our results highlighted significant differences among demographic groups and perceived acceptance of use cases, especially for professional use (§5.2). Non-majority demographic groups, especially non-male gender groups, found both personal and professional use cases less acceptable. Those experiencing high discrimination chronicity also found professional use less acceptable. Our findings offer empirical insights for future research on AI integration in workplaces, where marginalized workers' agency, income, and well-being are disproportionately impacted (Ming et al. 2024; Alcover et al. 2021).

Furthermore, our work highlighted a potential polarization on perceptions of AI among workers; those with 40+ hours employment and with advanced degrees were more positive towards AI use cases, suggesting that the relationship stakeholders have to AI and jobs might influence acceptability. This concern was expressed by a participant who opposed the development of Telemarketer AI, stating that it *"overlaps with my industry, and hence serves as a threat to my job security"* (P35). Thus, our results corroborate the need to further explore methods to include diverse workers and various stakeholders into the discussion of workplace AI integration and development (Fox et al. 2020; Cheon 2023). We also found that frequent AI usage increased acceptance, while understanding AI ethics and limitations decreased acceptance. This suggests that balanced AI awareness and education, encompassing usage, skills, and ethics, could guide and improve decision-making (Raji, Scheuerman, and Amironesei 2021), e.g., through educational interventions targeting AI skills and ethical implication literacy (e.g., Wong and Nguyen 2021; Shen et al. 2021).

Rationales Through analyzing participants' rationales, we observed an interesting pattern: use cases with less disagreement tended to elicit more cost-benefit (utilitarian) reasoning, while those with greater disagreement showed more rule-based (deontological) reasoning (§5.3). This suggests participants may apply different valuation frameworks, leading to diverging judgments, and highlights that some use cases raise concerns beyond simple utilitarian considerations. Our results thus underscore crucial elements of participants decision making pattern when only assessing impact as many prior works have done. Building upon our empirical findings, future work could develop diverse open-ended analysis for eliciting deliberations as well as tools and interventions using specific types of acceptability reasoning such as rule and value based (Sorensen et al. 2024) or cost-benefit analyses (Li et al. 2024).

However, as our study was limited to the two reasoning type categories, expanding this analysis would be essential for future work including finding ways to classify what features people are considering in their decisions, how the weights on those features impact what kind of decision-making strategy they will use, and whether there are other ways to understand their decision strategies beyond our current classification. Future research can build upon these further understandings to guide policy making and consensus building. For example, future work could explore how group discussions, beyond individual surveys, shape communities' collective understanding of AI impacts (e.g., Kuo et al. 2024; Lee et al. 2019; DeVos et al. 2022; Gordon et al. 2022; Zhang et al. 2023).

Acknowledgements

We thank the reviewers for their valuable feedback. We are also grateful to Tzu-Sheng Kuo, Jing-Jing Li, and many others for their insightful discussions. This work was supported in part by funding from Google's Society-Centered AI program and the Block Center at Carnegie Mellon University.

References

2023. Executive Order on the Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence. https://www.whitehouse.gov/briefing-room/presidential-actions/2023/10/30/executive-order-on-the-safe-secure-and-trustworthy-development-and-use-of-artificial-intelligence/. Accessed: 2024-01-02.

2023. How Do People Feel About AI? A Nationally Representative Survey of Public Attitudes to Artificial Intelligence in Britain.

ACL, Ethic Policy. 2023. ACL 2023 Responsible NLP Research and Ethics Policy.

Ada Lovelace Institute. 2022. Looking before we leap: Expanding ethical review processes for AI and data science research.

Alcover, C.-M.; Guglielmi, D.; Depolo, M.; and Mazzetti, G. 2021. "Aging-and-Tech Job Vulnerability": A proposed framework on the dual impact of aging and AI, robotics, and automation among older workers. *Organizational Psychology Review*, 11(2): 175–201.

Ashurst, C.; Barocas, S.; Campbell, R.; Raji, D.; and Russell, S. 2020. Navigating the Broader Impacts of AI Research. NeurIPS workshop.

Atari, M.; Haidt, J.; Graham, J.; Koleva, S.; Stevens, S. T.; and Dehghani, M. 2023. Morality beyond the WEIRD: How the nomological network of morality varies across cultures. *Journal of Personality and Social Psychology*.

Awad, E.; Dsouza, S.; Kim, R.; Schulz, J.; Henrich, J.; Shariff, A.; Bonnefon, J.-F.; and Rahwan, I. 2018. The moral machine experiment. *Nature*, 563(7729): 59–64.

Barocas, S.; Guo, A.; Kamar, E.; Krones, J.; Morris, M. R.; Vaughan, J. W.; Wadsworth, W. D.; and Wallach, H. 2021. Designing Disaggregated Evaluations of AI Systems: Choices, Considerations, and Tradeoffs. In *Proceedings of the 2021 AAAI/ACM Conference on AI, Ethics, and Society*, AIES '21, 368–378. New York, NY, USA: Association for Computing Machinery. ISBN 9781450384735.

Barocas, S.; and Selbst, A. D. 2016. Big data's disparate impact. *Calif. L. Rev.*, 104: 671.

Bernstein, M. S.; Levi, M.; Magnus, D.; Rajala, B. A.; Satz, D.; and Waeiss, Q. 2021. Ethics and society review: Ethics reflection as a precondition to research funding. *Proceedings of the National Academy of Sciences*, 118(52): e2117261118.

Bhardwaj, E.; Gujral, H.; Wu, S.; Zogheib, C.; Maharaj, T.; and Becker, C. 2024. Machine learning data practices through a data curation lens: An evaluation framework. In *The 2024 ACM Conference on Fairness, Accountability, and Transparency*, 1055–1067.

Birhane, A.; Kalluri, P.; Card, D.; Agnew, W.; Dotan, R.; and Bao, M. 2022. The values encoded in machine learning research. In *Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency*, 173–184.

Borg, J. S.; and Read, H. 2024. What Is Required for Empathic AI? It Depends, and Why That Matters for AI Developers and Users. In *Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society*, volume 7, 1306–1318.

Brailsford, J.; Vetere, F.; and Velloso, E. 2024. Exploring the Association between Moral Foundations and Judgements of AI Behaviour. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*, 1–15.

Buçinca, Z.; Pham, C. M.; Jakesch, M.; Ribeiro, M. T.; Olteanu, A.; and Amershi, S. 2023. AHA!: Facilitating AI Impact Assessment by Generating Examples of Harms. *arXiv preprint arXiv:2306.03280*.

Center for Advanced Study in the Behavioral Sciences. 2020. Ethics & Society Review · Stanford University.

Chen, R. J.; Wang, J. J.; Williamson, D. F.; Chen, T. Y.; Lipkova, J.; Lu, M. Y.; Sahai, S.; and Mahmood, F. 2023. Algorithmic fairness in artificial intelligence for medicine and healthcare. *Nature biomedical engineering*, 7(6): 719–742.

Cheng, H.-F.; Wang, R.; Zhang, Z.; O'connell, F.; Gray, T.; Harper, F. M.; and Zhu, H. 2019. Explaining decision-making algorithms through UI: Strategies to help non-expert stakeholders. In *Proceedings of the 2019 chi conference on human factors in computing systems*, 1–12.

Cheon, E. 2023. Powerful Futures: How a Big Tech Company Envisions Humans and Technologies in the Workplace of the Future. *Proc. ACM Hum.-Comput. Interact.*, 7(CSCW2).

Cheung, V.; Maier, M.; and Lieder, F. 2024. Measuring the decision process in (moral) dilemmas: Self-report measures of reliance on rules, cost-benefit reasoning, intuition, & deliberation.

Cushman, F. 2013. Action, outcome, and value: A dual-system framework for morality. *Personality and social psychology review*, 17(3): 273–292.

CVPR, Ethics Guidelines. 2023. CVPR 2024 Ethics Guidelines for Authors.

Deng, W. H.; Barocas, S.; and Vaughan, J. W. 2024. Supporting Industry Computing Researchers in Assessing, Articulating, and Addressing the Potential Negative Societal Impact of Their Work. *arXiv preprint arXiv:2408.01057*.

Deng, W. H.; Guo, B.; Devrio, A.; Shen, H.; Eslami, M.; and Holstein, K. 2023. Understanding Practices, Challenges, and Opportunities for User-Engaged Algorithm Auditing in Industry Practice. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, 1–18.

Deng, W. H.; Nagireddy, M.; Lee, M. S. A.; Singh, J.; Wu, Z. S.; Holstein, K.; and Zhu, H. 2022. Exploring how machine learning practitioners (try to) use fairness toolkits. In *Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency*, 473–484.

Deng, W. H.; Wang, C.; Han, H. Z.; Hong, J. I.; Holstein, K.; and Eslami, M. 2025. WeAudit: Scaffolding User Auditors and AI Practitioners in Auditing Generative AI. *arXiv preprint arXiv:2501.01397*.

DeVos, A.; Dhabalia, A.; Shen, H.; Holstein, K.; and Eslami, M. 2022. Toward User-Driven Algorithm Auditing: Investigating users' strategies for uncovering harmful algorithmic behavior. In *Proceedings of the 2022 CHI conference on human factors in computing systems*, 1–19.

Eslami, M.; Fox, S.; Shen, H.; Fan, B.; Lin, Y.-R.; Farzan, R.; and Schwanke, B. 2025. From Margins to the Table: Charting the Potential for Public Participatory Governance of Algorithmic Decision Making. In *Proceedings of the 2025 ACM Conference on Fairness, Accountability, and Transparency*, FAccT '25, 2657–2670. New York, NY, USA: Association for Computing Machinery. ISBN 9798400714825.

Evetts, J. 2006. Introduction: Trust and professionalism: Challenges and occupational changes.

Fox, S. E.; Khovanskaya, V.; Crivellaro, C.; Salehi, N.; Dombrowski, L.; Kulkarni, C.; Irani, L.; and Forlizzi, J. 2020. Worker-Centered Design: Expanding HCI Methods for Supporting Labor. In *Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems*, CHI EA '20, 1–8. New York, NY, USA: Association for Computing Machinery. ISBN 9781450368193.

Ge, X.; Xu, C.; Misaki, D.; Markus, H. R.; and Tsai, J. L. 2024. How culture shapes what people want from AI. In *Proceedings of the 2024 CHI conference on human factors in computing systems*, 1–15.

Golpayegani, D.; Pandit, H. J.; and Lewis, D. 2023. To Be High-Risk, or Not To Be—Semantic Specifications and Implications of the AI Act's High-Risk AI Applications and Harmonised Standards. In *Proceedings of the 2023 ACM Conference on Fairness, Accountability, and Transparency*, FAccT '23, 905–915. New York, NY, USA: Association for Computing Machinery. ISBN 9798400701924.

Gordon, M. L.; Lam, M. S.; Park, J. S.; Patel, K.; Hancock, J.; Hashimoto, T.; and Bernstein, M. S. 2022. Jury learning: Integrating dissenting voices into machine learning models. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*, 1–19.

Graham, J.; Nosek, B. A.; Haidt, J.; Iyer, R.; Koleva, S.; and Ditto, P. H. 2011. Mapping the moral domain. *Journal of personality and social psychology*, 101(2): 366.

Graham, J.; Nosek, B. A.; Haidt, J.; Iyer, R.; Spassena, K.; and Ditto, P. H. 2008. Moral foundations questionnaire. *Journal of Personality and Social Psychology*.

Hecht, B.; Wilcox, L.; Bigham, J. P.; Schöning, J.; Hoque, E.; Ernst, J.; Bisk, Y.; De Russis, L.; Yarosh, L.; Anjum, B.; et al. 2021. It's time to do something: Mitigating the negative impacts of computing through a change to the peer review process. *arXiv preprint arXiv:2112.09544*.

Herdel, V.; Šćepanović, S.; Bogucka, E.; and Quercia, D. 2024. ExploreGen: Large language models for envisioning the uses and risks of AI technologies. In *Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society*, volume 7, 584–596.

ICML, Publication Ethics. 2023. International Conference on Machine Learning, Publication Ethics.

Islam, T.; and Goldwasser, D. 2025. Can LLMs Assist Annotators in Identifying Morality Frames?-Case Study on Vaccination Debate on Social Media. In *Proceedings of the 17th ACM Web Science Conference 2025*, 169–178.

Jakesch, M.; Buçinca, Z.; Amershi, S.; and Olteanu, A. 2022. How different groups prioritize ethical values for responsible AI. In *proceedings of the 2022 ACM conference on fairness, accountability, and transparency*, 310–323.

Kapania, S.; Siy, O.; Clapper, G.; Sp, A. M.; and Sambasivan, N. 2022. "Because AI is 100% right and safe": User attitudes and sources of AI authority in India. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*, 1–18.

Kawakami, A.; Taylor, J.; Fox, S.; Zhu, H.; and Holstein, K. 2024. AI Failure Loops in Feminized Labor: Understanding the Interplay of Workplace AI and Occupational Devaluation. In *Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society*, volume 7, 683–683.

Kelly, J. 2025. Jobs AI will replace first in the workplace shift.

Kieslich, K.; Diakopoulos, N.; and Helberger, N. 2023. Anticipating Impacts: Using Large-Scale Scenario Writing to Explore Diverse Implications of Generative AI in the News Environment. *arXiv preprint arXiv:2310.06361*.

Kieslich, K.; Helberger, N.; and Diakopoulos, N. 2024. My Future with My Chatbot: A Scenario-Driven, User-Centric Approach to Anticipating AI Impacts. In *Proceedings of the 2024 ACM Conference on Fairness, Accountability, and Transparency*, FAccT '24, 2071–2085. New York, NY, USA: Association for Computing Machinery. ISBN 9798400704505.

Kingsley, S.; Zhi, J.; Deng, W. H.; Lee, J.; Zhang, S.; Eslami, M.; Holstein, K.; Hong, J. I.; Li, T.; and Shen, H. 2024. Investigating What Factors Influence Users' Rating of Harmful Algorithmic Bias and Discrimination. In *Proceedings of the AAAI Conference on Human Computation and Crowdsourcing*, volume 12, 75–85.

Kolata, G. 2024. Chatgpt defeated doctors at diagnosing illness - The New York Times.

Kramer, M. F.; Schaich Borg, J.; Conitzer, V.; and Sinnott-Armstrong, W. 2018. When Do People Want AI to Make Decisions? In *Proceedings of the 2018 AAAI/ACM Conference on AI, Ethics, and Society*, AIES '18, 204–209. New York, NY, USA: Association for Computing Machinery. ISBN 9781450360128.

Kuo, T.-S.; Chen, Q. Z.; Zhang, A. X.; Hsieh, J.; Zhu, H.; and Holstein, K. 2024. PolicyCraft: Supporting Collaborative and Participatory Policy Design through Case-Grounded Deliberation. *arXiv preprint arXiv:2409.15644*.

Lee, H. R. 2024. Contrasting Perspectives of Workers: Exploring Labor Relations in Workplace Automation and Potential Interventions. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems*, 1–17.

Lee, M. K.; Kusbit, D.; Kahng, A.; Kim, J. T.; Yuan, X.; Chan, A.; See, D.; Noothigattu, R.; Lee, S.; Psomas, A.; and Procaccia, A. D. 2019. WeBuildAI: Participatory Framework for Algorithmic Governance. *Proc. ACM Hum.-Comput. Interact.*, 3(CSCW).

Li, J.-J.; Pyatkin, V.; Kleiman-Weiner, M.; Jiang, L.; Dziri, N.; Collins, A. G. E.; Schaich Borg, J.; Sap, M.; Choi, Y.; and Levine, S. 2024. SafetyAnalyst: Interpretable, transparent, and steerable LLM safety moderation. *arXiv*.

Lin, H.-T.; Balcan, M.-F.; Hadsell, R.; and Ranzato, M. 2020. Getting Started with NeurIPS 2020. NeurIPS blog.

McClain, C. 2025. How the U.S. public and AI experts view Artificial Intelligence.

Méndez-Suárez, M.; Monfort, A.; and Hervas-Oliver, J.-L. 2023. Are you adopting artificial intelligence products? Social-demographic factors to explain customer acceptance. *European Research on Management and Business Economics*, 29(3): 100223.

Metcalf, J.; Moss, E.; Watkins, E. A.; Singh, R.; and Elish, M. C. 2021. Algorithmic impact assessments and accountability: The co-construction of impacts. In *Proceedings of the 2021 ACM conference on fairness, accountability, and transparency*, 735–746.

Michaels, E.; Thomas, M.; Reeves, A.; Price, M.; Hasson, R.; Chae, D.; and Allen, A. 2019. Coding the Everyday Discrimination Scale: implications for exposure assessment and associations with hypertension and depression among a cross section of mid-life African American women. *J Epidemiol Community Health*, 73(6): 577–584.

Microsoft. 2022a. Microsoft Responsible AI Impact Assessment Guide.

Microsoft. 2022b. Microsoft Responsible AI Impact Assessment Template.

Ming, J.; Pei, L.; Varanasi, R. A.; Kawakami, A.; Verdezoto, N.; and Cheon, E. 2024. Labor, Visibility, and Technology: Weaving Together Academic Insights and On-Ground Realities. In *Companion Publication of the 2024 Conference on Computer-Supported Cooperative Work and Social Computing*, CSCW Companion '24, 708–711. New York, NY, USA: Association for Computing Machinery. ISBN 9798400711145.

Mun, J.; Jiang, L.; Liang, J.; Cheong, I.; DeCario, N.; Choi, Y.; Kohno, T.; and Sap, M. 2024. Particip-AI: A Democratic Surveying Framework for Anticipating Future AI Use Cases, Harms and Benefits. arXiv:2403.14791.

National Artificial Intelligence Research Resource Task Force. 2023. Strengthening and Democratizing the U.S. Artificial Intelligence Innovation Ecosystem.

of Standards, N. I.; and Technology. 2023. Artificial Intelligence Risk Management Framework (AI RMF 1.0).

Olteanu, A.; Ekstrand, M.; Castillo, C.; and Suh, J. 2023. Responsible AI Research Needs Impact Statements Too. *arXiv preprint arXiv:2311.11776*.

on AI, P. 2021. Managing the Risks of AI Research: Six Recommendations for Responsible Publication.

OpenAI. 2022. OpenAI: Our approach to alignment research.

Parliament, E. 2023. Artificial Intelligence Act: deal on comprehensive rules for trustworthy AI.

Pierson, E.; Shanmugam, D.; Movva, R.; Kleinberg, J.; Agrawal, M.; Dredze, M.; Ferryman, K.; Gichoya, J. W.; Jurafsky, D.; Koh, P. W.; et al. 2025. Using Large Language Models to Promote Health Equity.

Pistilli, G.; Muñoz Ferrandis, C.; Jernite, Y.; and Mitchell, M. 2023. Stronger together: on the articulation of ethical charters, legal tools, and technical documentation in ML. In *Proceedings of the 2023 ACM Conference on Fairness, Accountability, and Transparency*, 343–354.

Prabhakaran, V.; Mitchell, M.; Gebru, T.; and Gabriel, I. 2022. A human rights-based approach to responsible AI. *arXiv preprint arXiv:2210.02667*.

Raji, I. D.; Scheuerman, M. K.; and Amironesei, R. 2021. You Can't Sit With Us: Exclusionary Pedagogy in AI Ethics Education. In *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*, FAccT '21, 515–525. New York, NY, USA: Association for Computing Machinery. ISBN 9781450383097.

Rajpurkar, P.; Chen, E.; Banerjee, O.; and Topol, E. J. 2022. AI in health and medicine. *Nature medicine*, 28(1): 31–38.

Reisman, D.; Schultz, J.; Crawford, K.; and Whittaker, M. 2018. Algorithmic Impact Assessments: A Practical Framework for Public Agency. *AI Now*.

Shen, H.; Deng, W. H.; Chattopadhyay, A.; Wu, Z. S.; Wang, X.; and Zhu, H. 2021. Value cards: An educational toolkit for teaching social impacts of machine learning through deliberation. In *Proceedings of the 2021 ACM conference on fairness, accountability, and transparency*, 850–861.

Smith, J. J.; Beattie, L.; and Cramer, H. 2023. Scoping fairness objectives and identifying fairness metrics for recommender systems: The practitioners' perspective. In *Proceedings of the ACM Web Conference 2023*, 3648–3659.

Society, D. . 2023. Data & Society Announces the Launch of its Algorithmic Impact Methods Lab.

Solaiman, I. 2023. The Gradient of Generative AI Release: Methods and Considerations. In *Proceedings of the 2023 ACM Conference on Fairness, Accountability, and Transparency*, FAccT '23, 111–122. New York, NY, USA: Association for Computing Machinery. ISBN 9798400701924.

Solaiman, I.; Talat, Z.; Agnew, W.; Ahmad, L.; Baker, D.; Blodgett, S. L.; Chen, C.; Daumé III, H.; Dodge, J.; Duan, I.; et al. 2023. Evaluating the social impact of generative ai systems in systems and society. *arXiv preprint arXiv:2306.05949*.

Sonboli, N.; Smith, J. J.; Cabral Berenfus, F.; Burke, R.; and Fiesler, C. 2021. Fairness and transparency in recommendation: The users' perspective. In *Proceedings of the 29th ACM Conference on User Modeling, Adaptation and Personalization*, 274–279.

Sorensen, T.; Jiang, L.; Hwang, J.; Levine, S.; Pyatkin, V.; West, P.; Dziri, N.; Lu, X.; Rao, K.; Bhagavatula, C.; Sap, M.; Tasioulas, J.; and Choi, Y. 2024. Value Kaleidoscope:

Engaging AI with Pluralistic Human Values, Rights, and Duties. In *AAAI*.

Svensson, L. G. 2006. Professional occupations and status: A sociological study of professional occupations, status and trust.

Telkamp, J. B.; and Anderson, M. H. 2022. The implications of diverse human moral foundations for assessing the ethicality of artificial intelligence. *Journal of Business Ethics*, 178(4): 961–976.

The Ada Lovelace Insitute. 2022. Algorithmic impact assessment: a case study in healthcare.

The White House. 2023. National Artificial Intelligence Research Resource Task Force Releases Final Report.

Times, T. N. Y. 2024. Will Chatbots Teach Your Children?

Walker, K.; and Croak, M. 2021. An update on our progress in responsible AI innovation.

Wang, B.; Rau, P.-L. P.; and Yuan, T. 2023. Measuring user competence in using artificial intelligence: validity and reliability of artificial intelligence literacy scale. *Behaviour & information technology*, 42(9): 1324–1337.

Wang, X.; Kim, H.; Rahman, S.; Mitra, K.; and Miao, Z. 2024a. Human-llm collaborative annotation through effective verification of llm labels. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems*, 1–21.

Wang, Z. J.; Kulkarni, C.; Wilcox, L.; Terry, M.; and Madaio, M. 2024b. Farsight: Fostering Responsible AI Awareness During AI Application Prototyping. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*, 1–40.

Weidinger, L.; Uesato, J.; Rauh, M.; Griffin, C.; Huang, P.-S.; Mellor, J.; Glaese, A.; Cheng, M.; Balle, B.; Kasirzadeh, A.; et al. 2022. Taxonomy of risks posed by language models. In *Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency*, 214–229.

Wong, R. Y.; and Nguyen, T. 2021. Timelines: A worldbuilding activity for values advocacy. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, 1–15.

Wongpakaran, N.; Wongpakaran, T.; Wedding, D.; and Gwet, K. L. 2013. A comparison of Cohen's Kappa and Gwet's AC1 when calculating inter-rater reliability coefficients: a study conducted with personality disorder samples. *BMC medical research methodology*, 13: 1–7.

Wu, Y.; Lee, J.-J.; Pillai, A. G.; Cho, J.; Ahmadpour, N.; Roto, V.; Sachathep, T.; Liu, J.; Sawan, M.; Song, D.; Čaić, M.; Cheng, L.; Liu, R.; Kettley, S.; Soares, L.; Grace, K.; and Astell-Burt, T. 2024. Collective Imaginaries for the Futures of Care Work. In *Companion Publication of the 2024 Conference on Computer-Supported Cooperative Work and Social Computing*, CSCW Companion '24, 732–735. New York, NY, USA: Association for Computing Machinery. ISBN 9798400711145.

Zhai, C.; Wibowo, S.; and Li, L. D. 2024. The effects of over-reliance on AI dialogue systems on students' cognitive abilities: a systematic review. *Smart Learning Environments*, 11(1): 28.

Zhang, A.; Walker, O.; Nguyen, K.; Dai, J.; Chen, A.; and Lee, M. K. 2023. Deliberating with AI: Improving Decision-Making for the Future through Participatory AI Design and Stakeholder Deliberation. *Proc. ACM Hum.-Comput. Interact.*, 7(CSCW1).

Zhang, B.; and Dafoe, A. 2020. US public opinion on the governance of artificial intelligence. In *Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society*, 187–193.