

The Hall of AI Fears and Hopes: Comparing the Views of AI Influencers and those of Members of the U.S. Public Through an Interactive Platform

Gustavo Moreira
gmorei3@uic.edu
University of Illinois Chicago
USA

Marios Constantinides
marios.constantinides@cyens.org.cy
CYENS Centre of Excellence & Nokia Bell Labs
Cyprus & UK

Edyta Bogucka
edyta.bogucka@nokia-bell-labs.com
Nokia Bell Labs & University of Cambridge
UK

Daniele Quercia
quercia@cantab.net
Nokia Bell Labs & Politecnico di Torino
UK & Italy

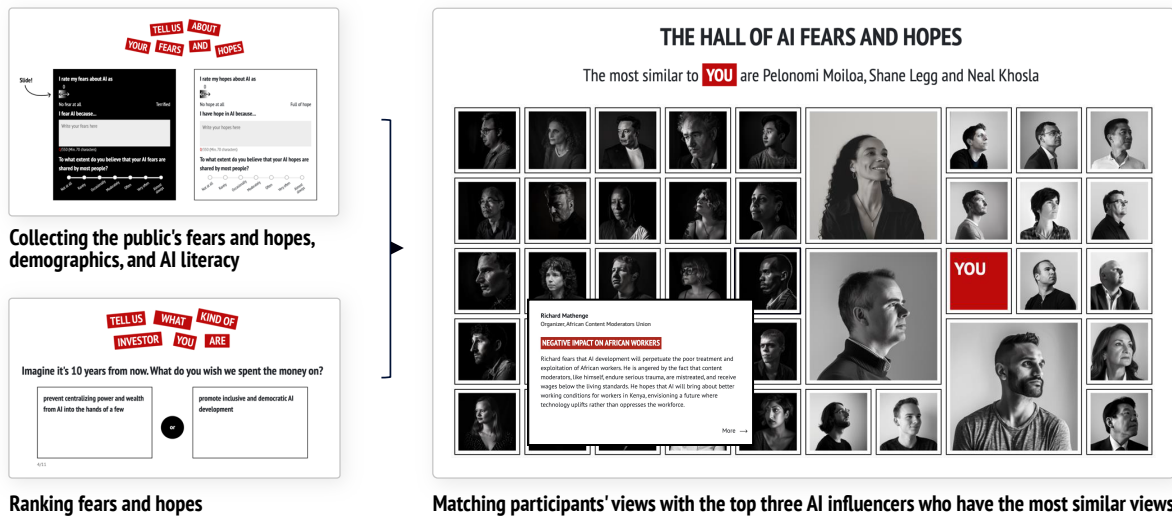


Figure 1: “The Hall of AI Fears and Hopes” is an interactive platform for collecting the public’s views on AI.

Abstract

AI development is shaped by academics and industry leaders—let us call them “influencers”—but it is unclear how their views align with those of the public. To address this gap, we developed an interactive platform that served as a data collection tool for exploring public views on AI, including their fears, hopes, and overall sense of hopefulness. We made the platform available to 330 participants representative of the U.S. population in terms of age, sex, ethnicity, and political leaning, and compared their views with those of 100 AI influencers identified by the Time magazine. The public fears AI getting out of control, while influencers emphasize regulation, seemingly to deflect attention from their alleged focus on

monetizing AI’s potential. Interestingly, AI influencers from underrepresented groups, such as women and people of color, often differ from the public within those same groups.

CCS Concepts

• **Human-centered computing** → **Empirical studies in HCI**; **Empirical studies in collaborative and social computing**; *HCI design and evaluation methods*; • **Computing methodologies** → *Artificial intelligence*.

Keywords

ethical AI, responsible AI, value alignment, hopes, fears

ACM Reference Format:

Gustavo Moreira, Edyta Bogucka, Marios Constantinides, and Daniele Quercia. 2025. The Hall of AI Fears and Hopes: Comparing the Views of AI Influencers and those of Members of the U.S. Public Through an Interactive Platform. In *CHI Conference on Human Factors in Computing Systems (CHI '25)*, April 26–May 1, 2025, Yokohama, Japan. ACM, New York, NY, USA, 27 pages. <https://doi.org/10.1145/3706598.3714117>



This work is licensed under a Creative Commons Attribution 4.0 International License. *CHI '25*, April 26–May 1, 2025, Yokohama, Japan
© 2025 Copyright held by the owner/author(s).
ACM ISBN 979-8-4007-1394-1/25/04
<https://doi.org/10.1145/3706598.3714117>

1 Introduction

The rapid advancement of Artificial Intelligence (AI) systems has been shaped by individuals commonly referred to as “AI influencers”—researchers, technologists, policymakers, and thought leaders—whose expertise, advocacy, or commercial interests shape public discourse, policy, and industry trends [17, 19]. Researchers, for example, push the boundaries of what is possible with AI by shaping the theoretical underpinnings that guide its development [96]. On the other hand, industry leaders and affluent individuals enable the commercialization and widespread adoption of these technologies through their financial and organizational resources [10]. By investing in AI startups, funding research initiatives, and shaping market trends, these individuals ensure that cutting-edge AI technologies transition from academic laboratories to real-world applications, making them accessible to the public [93]. Given the considerable power these individuals hold, they often emerge as key decision-makers in the AI landscape [22, 86]. The decisions they make are not isolated; they carry profound implications for society at large. These decisions can then determine the trajectory of AI development, influence regulatory frameworks [18], and shape public perception and trust in AI systems [51]. Consequently, the impact of these decisions is far-reaching, affecting not only the technological landscape but also societal structures, and the daily lives of the public.

However, the public is frequently excluded from these critical decision-making processes [8, 17]. Despite being the most affected by AI technologies (yet relegated to the role of “AI subjects” [28] or “human subjects” [72]), the public’s voice is often marginalized or entirely absent in discussions about AI development and governance [77]. When policies are developed without adequately reflecting public concerns, the public may be compelled to follow rules they find unfair or misaligned with their values. This issue is worsened by the fact that many key decisions around AI are made by unelected officials, corporate executives, and industry leaders rather than through democratic processes [92]. Similar observations were made on smart cities and civic innovation wherein citizens are either represented by professional and bureaucratic elites, or are entirely absent from key decision making [97, 114]. Therefore, the concentration of decision-making power outside democratic oversight risks alienating the public and diminishing trust in both AI and its governing institutions. To truly serve the public good, AI development must actively incorporate public perspectives [5, 60, 70].

In particular, work in the Human-Computer Interaction (HCI) community has demonstrated the importance of including diverse public viewpoints in AI development and governance [46, 50, 55]. A thematic analysis of public views has identified several key drivers of AI-phobia: AI substitutability, AI accountability, AI literacy, and AI fever, all of which point out public anxiety surrounding AI’s impact on employment and healthcare [61, 71]. These findings reveal that concerns over job displacement and the fairness of AI’s integration into healthcare are top priorities for the public [13, 36], reinforcing the need for a more inclusive AI governance [49].

A growing body of literature has focused on understanding public views toward AI in isolation, without adequately examining how these views compare to those of the influential figures who shape the AI discourse [13, 43, 51, 54, 90, 91]. This lack of comparative analysis creates a gap in our understanding: *we do not fully*

comprehend whether the decisions made by AI influencers reflect the desires and concerns of the public. This (mis)alignment could either exacerbate existing societal inequalities or foster more equitable and inclusive AI systems. Recent studies from the HCI community stress the need to bridge this gap by creating tools and processes that ensure public concerns are not only heard but actively inform AI policy and development [31, 49, 50]. Therefore, understanding this dynamic is essential for developing AI technologies that are not only innovative but also socially responsible and aligned with the public’s best interests. To investigate this dynamic, we ask:

(RQ₁) What are the views of a sample of the public on AI?

(RQ₂) What are the views of AI influencers?

(RQ₃) Which subgroups of AI influencers have views that are most (or least) aligned with our sample of the public?

(RQ₄) What themes emerge from the most (or least) aligned views?

To address these questions, we conducted a three-phase study. First, we curated a dataset of views from 100 influential figures in AI, as identified by the Time magazine, including researchers, innovators, and policymakers. Next, we co-designed an interactive platform, “The Hall of AI Fears and Hopes”, with two designers, and tested it with 30 participants to refine the data collection process. We then deployed this platform on Prolific with 330 U.S. participants, representative in terms of age, sex, ethnicity, and political leaning, to collect their views. Finally, we conducted a comparative analysis to identify patterns of alignment and misalignment between public views and those of AI influencers. Our study makes the following contributions to the understanding of AI discourse:

- (1) We propose a structured approach to compare how AI influencers and the public perceive AI, including (1) a dataset of views from influential AI figures (§4.1); (2) a platform that allows people to explore these influencer views and share their own (§4.2); and (3) metrics and methods to analyze the differences between these views (§4.3).
- (2) We provide empirical insights into (mis)alignment patterns between AI influencers and the public. Our analysis reveals that older influencers shared fewer views with the public, while younger influencers, academics, and non-billionaires aligned more closely with public opinions. Interestingly, AI influencers from underrepresented groups (e.g., women, people of color) showed poor alignment with the public, indicating a disconnect even among historically underrepresented voices. At the same time, each group has different concerns: the public worries about losing control to AI and perceives only a few benefits, while influencers focus on supposedly controlling AI through calls for regulation and see many potential benefits.

Building on these contributions, we discuss how our results align with or challenge existing literature and highlight the need for follow-up studies across different contexts and a broader range of AI influencers (§6). While our analysis provides a valuable starting point with a subset of 100 influencers, it may not fully capture the diversity of views within the AI community. Therefore, we propose exploring alternative sampling approaches to better capture this diversity in future work. To support researchers in advancing this future work, we have publicly released the anonymized data and platform at <https://social-dynamics.net/fears-and-hopes>.

2 Related Work

We surveyed various lines of research that our work draws upon, and grouped them into two main areas for collecting views on AI (§2.1), and analyzing them (§2.2).

2.1 Collecting Views on AI

Historical power imbalances have resulted in some groups gaining greater access to the benefits of AI, while others are more likely to face its risks [8, 21, 99]. To ensure that AI development and regulation are equitable and inclusive, it is crucial to gather the perspectives of all these groups. Researchers achieve this through both direct and indirect methods [34, 71]. Direct methods explicitly ask for opinions through surveys, deliberative processes like community juries and assemblies, in-depth interviews, co-design studies and online narrative studies [27, 31, 48, 59, 71]. Indirect methods capture views in their natural settings or as they emerge organically in discussions or behaviors [41, 69].

The most common direct method for gathering views on AI is through surveys, which provide valuable population-level insights but often limit participant agency. These surveys range from large-scale studies across multiple countries [36, 51, 91] to more focused ones targeting representative groups within a single country [13] or specific subgroups, such as AI practitioners, employees and students [90]. They may employ customized questionnaires or standardized tools like the AI Attitude Scale (AIAS), the Perceptions on AI by the Citizens of Europe (PAICE) questionnaire [91], or the Technology Acceptance Model (TAM) questionnaires [52]. The focus of these surveys also varies, with some assessing general attitudes towards AI, while others exploring perceptions of specific AI technologies, such as remote biometric identification [54] or facial recognition [56]. However, surveys alone may not fully capture the depth of people's views or the reasoning behind them. To address this, they are often combined with additional direct methods, such as interviews or focus groups, to provide a richer, more nuanced understanding. Adus et al. [2] conducted virtual focus groups with patient participants to understand of how patients can and should be meaningfully engaged within the field of AI development in healthcare. Kapania's study [48] on AI perceptions in India supplemented survey data with in-depth interviews of adult Internet users. Similarly, Nader et al. [71] combined an online survey with focus groups of entertainment media creators to explore how media shapes public beliefs about AI. However, interviews and focus groups often structure participation in ways that limit spontaneous deliberation and collective sense-making. To foster more dynamic and inclusive discussions, alternative methods such as deliberative polling were explored. For example, the Polis system [106] helped identify distinct opinion groups by enabling participants to vote on or contribute preferences for AI chatbot behavior [44]. Similarly, the Stanford Online Deliberation Platform [15] facilitated Taiwan's national deliberation by allowing the public to discuss policy options for addressing deepfakes and forged media [107].

Indirect methods for gathering views on AI often involve analyzing secondary sources [81] or engaging participants in imaginative and interactive scenarios. For instance, Cave and Dihal [14] analyzed over 300 news articles, literary works, and films to identify eight fundamental hopes and fears about AI in Western societies.

Awad et al. [5] had participants imagine solutions to AI-related dilemmas using "Dilemma Vignettes," while Robb et al. [87] paired a public exhibition with a quiz to explore public perceptions of robots. Another approach is seen in the Collective Constitutional AI (CCAI) project [43], which gathers preferences for large language model behavior by allowing participants to contribute their preferences and vote for the preferences of others using options like "Agree," or "Disagree". Some studies combine direct and indirect methods; for example, Hohendanner et al. [41] surveyed computer science students and also conducted a speculative design workshop with designers to explore AI perceptions from multiple angles.

2.2 Analyzing Views on AI

Once views on AI are collected, they can be analyzed and compared to understand the different perspectives and priorities held by various subgroups. This diversity of views on AI is explored in Jakesch et al. [46], where the authors analyze and compare how the U.S. population, crowdworkers, and AI practitioners perceive and prioritize aspects of Responsible AI (RAI). Similarly, Kapania et al. [48] collected the views of Indian citizens on AI, highlighting how their optimism contrasts with Western skepticism. Recognizing this range of opinions, several studies propose methods to incorporate the perspectives of various stakeholders, including the public, into the AI design process. For example, Lee et al. [62] introduces We-BuildAI, a participatory design framework that balances differing views by allowing participants to vote on others' perspectives. Additionally, STELA provides a methodology for community-centered norm elicitation to ensure AI value alignment [6], while the AI Failure Cards help foster a deeper understanding of AI failures and gather preferred mitigation strategies [105]. Kawakami et al. [50] developed a toolkit to guide early-stage discussions on whether, and under what conditions, the development or deployment of an AI system should proceed in the public sector.

However, incorporating this range of views presents challenges, as noted by Varanasi and Goyal [112], including difficulties in engaging with and prioritizing unfamiliar perspectives. Furthermore, some cases can result in what is known as "the tyranny of the majority," where the interests of the majority are enforced above those of the minority, potentially leading to the systematic disadvantaging of minority groups [5].

Research Gap. Previous works have concentrated on collecting views from either the public or AI practitioners. However, it is typically individuals in positions of influence, such as industry leaders and academics, who drive decisions regarding AI. This leaves an important gap in understanding whether the views of the public are aligned with those of the AI influencers who shape the direction of AI development and policy.

3 Researcher Positionality Statement

Before outlining our methodology, we first position ourselves in relation to the research we present in this work. The team consists of individuals with expertise in Data Visualization, Computer Science, and AI. Our team consists of three men and one woman, bringing together diverse experiences from both industrial research labs and academic institutions. We have cultural and professional backgrounds spanning Europe, North America, and South America.

We also represent a range of religious affiliations. We acknowledge that our positionality may influence various aspects of our research, including but not limited to our user-study methods, design decisions, aesthetic choices for the platform, language selection, and topics emphasized in qualitative analyses. We recognize the importance of including a broader range of voices from academia and the public to bring unique perspectives to the work we develop.

4 Methods for Collecting and Comparing AI Views from AI Influencers and the Public

Next, we describe our three-step methodology (Figure 2) for collecting views from AI influencers (§4.1); gathering public's views on AI using an interactive platform (§4.2); and analyzing both public's and AI influencers' views (§4.3).

4.1 Methods for Collecting Views from the AI Influencers

We defined three main criteria for selecting a list of influencers to analyze views on AI: (1) demographic diversity, (2) sectoral coverage, and (3) professional influence. To evaluate demographic diversity, we measured the percentage of women, men, and nonbinary individuals; the percentage of white individuals and people of color; and the percentage of participants within each of six generational cohorts. For sectoral coverage, we examined the distribution of roles across academia, government, industry, and other sectors, and the percentage of participants professionally active in different geographic locations (i.e., based in Silicon Valley vs. other locations). Finally, for professional influence, we identified the percentage of billionaires and millionaires as well as the percentage of individuals holding C-suite positions.

We sourced three candidate lists for comparison: the list of members from the UNESCO High-level Advisory Body on Artificial Intelligence [109], which includes 39 individuals; the list of members from the OECD Working Party on Artificial Intelligence Governance [76], which consists of 14 individuals; and the Time100 AI list by the Time magazine [108], which features 100 of the most influential individuals in AI. While these lists differ in size and focus, there is some overlap: five individuals appear on both the Time100 AI list and the UNESCO list, and one individual appears on both the Time100 AI list and the OECD list. These lists performed as follows across our criteria:

- **Criterion 1: Demographic diversity:** The OECD list had the highest proportion of women (57%) but no representation of nonbinary individuals, with the remaining 43% being men. The UN list followed with 51% women and 49% men, also lacking nonbinary inclusion. The Time list balanced representation across genders with 39% women, 59% men, and 2% nonbinary individuals. In terms of ethnicity, the OECD list split evenly between white participants (50%) and people of color (50%) but excluded Black or African American individuals. The UN list had the highest proportion of people of color (69%) but included fewer white individuals (31%). The Time list struck a middle ground, with 52% white and 48% people of color. Time100 AI provided the broadest age representation, spanning from younger individuals aged 18–24 (1%) to

older individuals aged 65+ (6%), with the largest group being 35–44 (40%). In contrast, both the OECD and UN lists were heavily skewed toward professionals aged 35–44, accounting for 57% and 41% of their participants, respectively.

- **Criterion 2: Sectoral coverage:** The OECD list was dominated by government representatives (71%) with limited roles in industry (22%). The UN list leaned heavily toward academia (41%) and government (31%) but underrepresented industry leaders (7%). The Time list features 31% of influencers from the industry, and significant representation across academia (19%) and the arts (11%). For geographic distribution, 100% of OECD influencers, 95% of UN influencers, and 69% of Time100 AI influencers were active in other locations outside Silicon Valley.
- **Criterion 3: Professional influence:** The Time100 AI list had the highest proportion of financially influential figures, with 11% billionaires and 31% millionaires. The UN list included 3% billionaires and 8% millionaires, while no data was available for the OECD list. Additionally, 33% of the Time100 AI comprised C-suite executives, significantly more than the UN list (13%) and the OECD list (which had none).

Based on these comparisons, we selected the Time100 AI list for our analysis. It offered the most balanced demographic diversity, broadest sectoral coverage, and strongest representation of professionally influential individuals. To verify these statistics, we contacted the editorial team of the Time magazine who provided detailed information on their selection and interview process. The list was developed through an editorial process including more than five dozen reporters and editors specializing in AI and business reporting, with input from field experts. The process, spanning 4–5 months, began with research and nominations in April 2023 and culminated in the list's release in August 2023. Influencers were nominated based on notable achievements such as publishing influential research, developing impactful products, or driving societal change through regulatory actions. The team evaluated a few hundred candidates and narrowed them down to the final 100 through weekly reviews, ensuring thematic diversity between sectors, roles, and regions for inclusion. These nominees were interviewed via video or phone, with discussions tailored to their individual achievements and broader impact. The resulting interviews were published on a dedicated webpage, with quotes clearly labeled as direct or paraphrased. In cases where interviews incorporated statements from publicly available materials, such as YouTube videos, these sources were clearly attributed to ensure transparency.

The Time100 AI list features 43 CEOs, founders, and co-founders, and 41 women and nonbinary individuals, with ages ranging from 18 to 76. The nominees are grouped into four categories: leaders (e.g., founders), innovators (e.g., artists), shapers (e.g., political advisors), and scholars (e.g., researchers). The list includes short bios of the nominees and long interviews that capture their perspectives on AI's current state and future development. To collect their views, three authors read the interviews and independently marked excerpts that focused on AI's intended uses and potential impacts. They then jointly categorized these excerpts into hopes and fears and assessed the overall sentiment of each article.

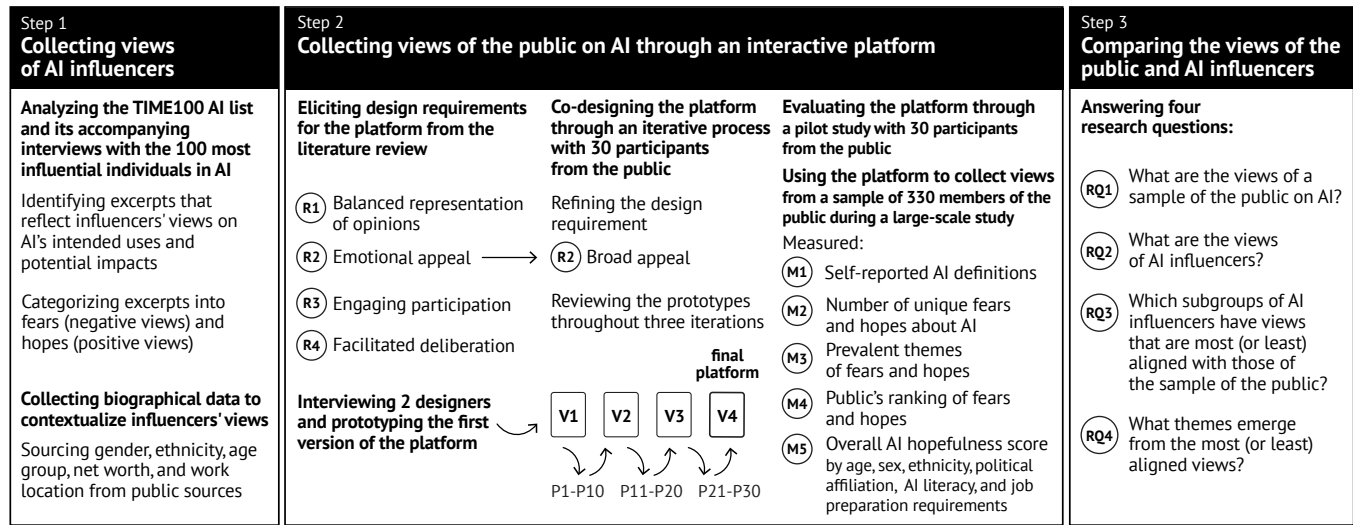


Figure 2: Overview of our three-step method for comparing the public's views and AI influencers' views on AI using an interactive platform. In the first step, we collected views from AI influencers—Time100 nominees [108]—by analyzing their interviews, focusing on key excerpts about AI's uses and impacts, and categorizing them into hopes and fears. In the second step, we built a platform to collect views from the broader public, starting with a literature review and interviews with two designers to elicit four design requirements and create the initial prototype (V1). Next, we engaged individuals from the public to iteratively update one design requirement and co-design three iterations of the platform (V2–V4). Next, we conducted a pilot study to evaluate the platform and identify improvements for the large-scale study. Finally, we used the platform to gather views from 330 U.S. individuals, representative of the population by age, sex, ethnicity, and political affiliation, capturing their fears and hopes about AI. In the third step, we compared the collected views.

To assess the sentiment, they searched for direct expressions of emotions (e.g., fear, anticipation) as per Plutchik's wheel of emotions [82] and previous works exploring human emotions at scale [47]. Plutchik's wheel of emotions was favored over alternatives (e.g., Ekman's six emotion model [26], or the positive-negative activation (PANAS) model of emotion by Watson and Tellegen [113]) because it provides a wider range of emotions and highlights relationships between them, offering flexibility in analyzing complex emotional expressions. However, we acknowledge it fails to account for contextual factors like personal history, social norms, or situational influences, as emotional expression and interpretation vary across cultures. If direct expressions were not present, the authors inferred the sentiment based on contextual clues and overall tone. In cases where the emotional content was unclear or there were disagreements, the authors discussed these instances and collaboratively determined the most dominant emotion. To enrich our dataset, we extended the Time's magazine bios by including additional biographic data such as gender, ethnicity, age group, net worth, and current working location (e.g., Silicon Valley or other locations). This biographic data was manually extracted from public company portals.

4.2 Methods for Collecting AI Views from the Public

4.2.1 Designing an Interactive Platform to Collect Views from the Public on AI. The platform served as a data collection tool to answer our four RQs. To design it, we followed three steps (Figure 2).

These are: 1) a literature review; 2) interviews with two designers employed at a large tech company and in academia; and 3) an iterative co-design process with 30 participants (representing the public) recruited from Prolific [83].

Step 1: Literature Review. To gather an initial set of requirements for the visualization, we conducted a targeted review of key papers focused on surveying public opinion on AI. This review was not intended to exhaustively cover all relevant literature but served as a starting point to inform our design, following similar co-designing practices [9, 16, 20]. Specifically, we formulated two questions to guide our literature search, that is: *What challenges do participants face when providing views about AI?* and *What features do participants prefer in tools for collecting and visualizing these views?*

Search Query

Subject: (ARTIFICIAL INTELLIGENCE)
What: (PUBLIC OPINION OR PUBLIC PERCEPTION OR PUBLIC ATTITUDE OR PUBLIC VIEW OR PUBLIC ACCEPTANCE)
How: (SURVEY OR TOOL OR VISUALIZATION)

We searched for papers in the ACM Digital Library [4] using a three-components query, involving: *Subject*, *What*, and *How* (as shown in the box above). This search identified a total of 120 papers.

We then screened these papers using inclusion criteria. Papers were included if they: 1) discussed public opinion, perception, views, or attitudes towards AI; 2) presented tools for gathering and visualizing public views; and 3) included user requirements or recommendations for future tools. This procedure left us with 8 papers [13, 38, 41, 51, 54, 87, 90, 118].

We then conducted a full-text analysis of these papers to document challenges and preferences for the platform. We highlighted key excerpts related to challenges, opportunities, and user requirements, and grouped these excerpts into common themes, such as public engagement, usability, and accuracy of view representation. Each theme included excerpts from at least two papers, signifying that data saturation was reached [40]. From these themes, we derived four design requirements for the platform:

- R1 *Balanced representation of views*: Ensure that both positive and negative views are presented in the platform.
- R2 *Emotional appeal*: Use visual elements that increase public's sense of contribution and agency.
- R3 *Engaging participation*: Make the interaction with the platform interactive and engaging.
- R4 *Facilitated deliberation*: Enable the public to form well-rounded opinions about AI.

Step 2: Interviews with Designers. Having the requirements at hand, we conducted 30-minute semi-structured interviews with two designers—one male and one female, aged 24 and 32, with 4 and 10 years of experience in data visualization, respectively, employed in academia and at a large tech company—to explore potential ways to visualize these requirements and prototype the first version of the platform (Appendix A.1). To recruit these designers, we used a responsible AI email list at a large technology company and adopted a purposive sampling approach [11]. This approach allowed us to intentionally target designers possessing the qualifications necessary for the study, particularly those with expertise in responsible AI, HCI, and data visualization, who: (1) have experience in analyzing large-scale public opinion data, including tools for collecting, interpreting, and visualizing public sentiment, and (2) publish their findings in peer-reviewed venues such as academic conferences. Out of five designers who expressed interest, two were available to participate within the project's timeline.

Step 3: Co-design Process. After designing the first version of the platform, we conducted three co-design sessions with 30 participants from the public who were recruited from Prolific [83]. These participants were selected to reflect the U.S. population demographics in terms of gender, [110, 111] (14 males and 16 females), with ages ranging from 18 to 64 years, and included 19 White, 3 Black, 1 Asian, and 8 individuals of Mixed or Other ethnicities. They were also digitally literate, exposed to visual media, and skilled in communication for providing feedback. Participants were asked to carefully review the prototype, consider how they would interact with the platform if fully functional, and then provide feedback on what worked well, what was confusing or needed improvement, and whether the design would encourage them to share their views on AI (Appendix A.2).

The study was approximately 30–45 minutes long and participants were paid on average about \$12 (USD) per hour. We then

conducted inductive thematic analysis [68], examining the recommendations for the next version of the platform. In total, we made three iterations, expanding our second design requirement R2 from *Emotional appeal* to *Broad appeal* to ensure the platform resonated with people of varying AI knowledge levels. We iteratively refined the design based on participants' recommendations (details in Appendix A.3), leading to its final version.

Final Visualization Platform. The platform consists of six sections and balances between collecting the public's views on AI and exposing them to the views held by AI influencers¹ (Figure 3).

In the first section of the platform (*Self-reported Definitions of AI*), we invited participants to share their interpretation of AI by asking, "How would you describe Artificial Intelligence to a friend?" [13]. In the second section of the platform (*Self-assessment of AI Fears and Hopes*), participants are given two cards: one black card on the left for AI fears, and one white card on the right for AI hopes (Figure 3A). On each card, they rate their level of fear or hope on a scale from 0 ("No fear / No hope") to 10 ("Terrified / Full of hope"). They also write down their specific fears and hopes and indicate how strongly they believe others share these views. In the third section (*Gallery of AI Influencers' Views*), participants are presented with black-and-white portraits of AI influencers alongside their views (Figure 3B; see §4.1 for how these views were collected from the interviews). The portraits are organized into two columns: on the left, influencers with more fearful views appear on a black background, while on the right, influencers with more hopeful views appear on a white background. When participants hover over a portrait, the influencer's name and surname appear (Figure 3B1). When they click on a portrait, the influencer's organizational role is displayed, along with their hopes and fears, and a link to the full interview (Figure 3B2). One portrait in the gallery stands out—it is enlarged with red labels, inviting participants to click and "Discover your place in the hall of AI thinkers" (Figure 3B3). Once clicked, participants are taken to the fourth section (*Demographics and AI Literacy Assessment*) as shown in Figure 3C. In this section, participants provide information on their gender, ethnicity, age, education, occupation, job training and education requirements, and AI literacy. The job training and education requirements are based on O*NET's job zones, which classify occupations by the level of education, training, and experience required, ranging from zone 1 (minimal education and training, e.g., housekeepers) to zone 5 (extensive education training, e.g., surgeons) [78]. For clarity, we rephrased O*NET's original "job zones" as "job training and education requirements." AI literacy is measured using the AI Literacy Scale (AILS) [7]. This information, along with the fears and hopes self-assessment from the first section, will be used to match and present the participants' views alongside those of AI influencers in the final, sixth section of the visualization. In the fifth section (*Pair-wise Comparison of AI Views*), participants are asked to imagine a future 10 years from now and answer "What do you wish we spent the money on?" (Figure 3D). They are then shown 10 randomly selected pairs from a set of 20 hopes and 20 fears expressed by AI influencers. Each pair of items may consist of two hopes, two fears, or one hope and one fear, enabling many varied comparisons.

¹"The Hall of AI Fears and Hopes" is available at <https://social-dynamics.net/fears-and-hopes>

Participants have the option to skip any pair if they do not wish to choose. To avoid any predictable patterns in the comparison process, each pair was presented in a randomized order. Additionally, the order of items within each pair (e.g., hope on the left, fear on the right) was randomized in each round to control for potential biases related to item positioning. We chose this pairwise comparison method over alternatives like Likert scales for two reasons [89]. First, it reduces participants' cognitive load, as comparing two items at a time is simpler than rating many at once. Second, it allows us to generate a ranked list of hopes and fears that reflects the preferences of the members of the public. After completing the pairwise comparison, participants move to the final, sixth section (*Gallery of AI Influencers' and a Participant's Views*) as shown in Figure 3E. This section displays both AI influencers' portraits and the participant's portrait. The participant's portrait is positioned based on their fear and hope ratings from the first section, the similarity of their fears and hopes to those of the influencers, and the alignment of their demographics (i.e., age, gender, ethnicity) from the fourth section with those of the influencers. Horizontally, portraits are placed on a spectrum from fearful to hopeful, while vertically, the placement is determined by the cosine similarity between the participant and their top three most similar influencers, considering both their written responses and demographic alignment. By positioning the participant among their most similar influencers, we apply the concept of homophily—the idea that people tend to connect more with those who are similar to them. Research shows that individuals often look to those who are alike when forming opinions and making decisions [25, 30]. Therefore, we used this principle to strengthen the participant's connection to the displayed opinions and encourage a more thoughtful and engaged reflection. Overall, the platform's design emphasizes the fear vs. hope framing through a black-and-white color palette, while the intentional use of realistic portraits humanizes the concept of AI by empowering participants to see AI as shaped by human perspectives.

4.2.2 Evaluating the Platform for Collecting Views from the Public about AI. We conducted a pilot study with 30 new participants, reflecting U.S. demographics in terms of sex and ethnicity, to evaluate the platform's design and simulate the upcoming large-scale study [79, 104]. The study assessed four design requirements: balanced representation of views (R1), broad appeal (R2), engaging participation (R3), and facilitated deliberation (R4). Participants rated these requirements on a 1–7 Likert scale, with average scores ranging from 4.94 to 5.88, indicating general agreement that the platform effectively met these design goals. Key areas for improvement included navigation and question phrasing. In response, we added arrow annotations, refined question wording, and updated survey settings, including adjusted monetary rewards for older participants, to enhance the platform's usability and response quality. These updates ensure the platform is ready for the large-scale study. Full demographics, methodological details, ratings, and a thematic analysis of participant feedback are provided in Appendix A.4.1.

4.3 Methods for Comparing the Views of the Public versus AI Influencers

We conducted a large-scale user study with 330 participants (representing the public) to collect their views on AI and expose them to

the views of AI influencers. Next, we describe our study's design: its setup (§4.3.1); research questions, metrics and analysis (§4.3.2); and execution (§4.3.3).

4.3.1 Setup. The study consisted of three steps. In the first step, participants received a brief introduction to the study and completed a warm-up task in which they described Artificial Intelligence to a friend. We chose this question based on a study of the public's fears and hopes in the UK [13] to ease participants into the survey and smoothly transition to more complex, personal questions about their fears, hopes and AI literacy. In the second step, participants interacted with our platform. We incorporated three attention checks directly into the platform to ensure that participants were engaged. In the third step, participants rated how well the platform met the four design requirements we identified with the members of the public and provided any additional feedback.

4.3.2 Research Questions, Metrics, and Analysis. Upon that procedure, we answered four research questions.

RQ1: What are the views of a sample of the public on AI?

We answered this question both quantitatively and qualitatively. Quantitatively, we defined two metrics:

- The number of unique fears and hopes. We computed this in four steps. First, we collected all sentences of hopes and fears provided by participants in the first section of the platform (Figure 3A) and generated their embeddings using the sentence transformer model 'all-mpnet-base-v2' [29]. We chose this model for its high performance on single sentences and short paragraphs [101]—the types of texts participants used to formulate their hopes and fears—and for its ability to produce high-quality embeddings. These embeddings were essential for the second step: clustering sentences to identify and remove duplicate fears and hopes. For this, we used HDBSCAN [12], a density-based clustering algorithm that does not require specifying the number of clusters, can manage clusters of different sizes and densities, and detects outliers (non-duplicates). We ran the algorithm to generate the initial set of clusters along with an outlier group. Two of the authors manually reviewed all sentences by visualizing them on the Figma platform [32], agreeing on where outliers should be placed and finding opportunities to merge clusters. Third, from each cluster, we randomly selected one fear or hope to keep as a representative, since the sentences in each cluster were very similar, and deleted the rest. For example, consider the two sentences: "AI will take away your own thinking and creativity" and "We will lose skills in writing and creativity". These sentences convey the same fear, so only one would be kept as the representative, while the other would be removed. Finally, we counted the resulting overall number of unique hopes and that of fears.
- The hopefulness score for particular subgroups of the public. We analyzed different subgroups of the public (e.g., males vs. females; younger participants, defined as those under 38 years old—about half the median life expectancy in the U.S.—vs. older participants, defined as those above 38 years old). Each participant i within a subgroup s provided two ratings in the second section of the platform (Figure 3B):

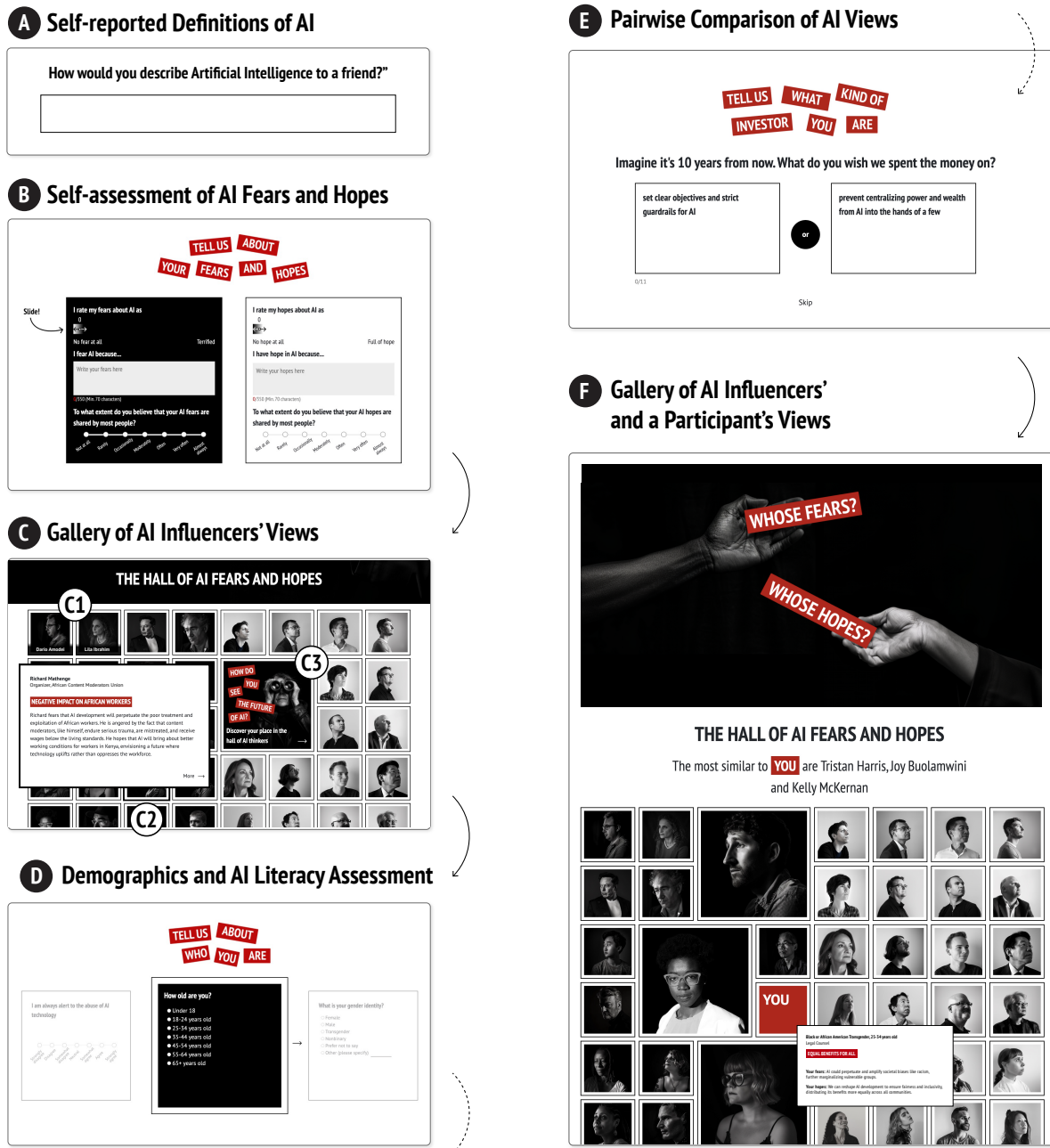


Figure 3: The “The Hall of AI Fears and Hopes” is an interactive platform consisting of six sections for collecting public views on AI and presenting the views of AI influencers. In the first section (A), participants share their definitions of AI. In the second section (B), they rate their fears and hopes, describe their specific fears and hopes, and indicate how widely they believe these views are shared. In the third section (C), participants can view portraits of AI influencers, organized by their more negative (left) and more positive (right) attitudes toward AI. By hovering (C1) and clicking on a portrait (C2), more information about the influencers’ fears and hopes appears in a pop-up box. Participants proceed to the next section (D) by clicking on (C3), where they provide data on gender, ethnicity, age, education level, occupation, and AI literacy. In the fifth section (E), participants imagine a future 10 years ahead and vote on how to allocate money based on 10 randomly selected pairs of AI influencers’ fears and hopes. In the final, sixth screen (F), participants see their symbolic portrait displayed alongside AI influencers, positioned based on their fear and hope ratings, as well as the similarity of their fears, hopes, and demographics (age, gender, ethnicity) to those of the influencers.

fear_{rating} and *hope_{rating}*. Then, for each participant in the subgroup, we computed an individual hopefulness score by subtracting their self-reported *fear_{rating}* from their *hope_{rating}*. Finally, the *hopefulness_{score_s}* was obtained by averaging the individual scores of all participants within the subgroup (Equation §1), with possible values ranging from $[-10, 10]$.

$$hopefulness_score_s = \frac{1}{N_s} \sum_{i=1}^{N_s} (fear_rating_i - hope_rating_i) \quad (1)$$

where N_s is the number of participants in subgroup s ; *fear_{rating_i}* and *hope_{rating_i}* are the fear and hope ratings provided by participant i .

Qualitatively, we thematically analyzed participants' self-reported definitions of AI, along with their associated fears and hopes. To analyze the definitions, we adopted two complementary approaches. First, we categorized them into clusters [55]. This involved pre-processing the definitions by removing repetitive phrases such as "Artificial Intelligence is..." and generating text embeddings using the 'sentence-transformers/all-mpnet-base-v2' model [29]. K-means clustering was then applied to these embeddings, with the optimal number of clusters ($k = 4$) determined using the elbow method. Second, we examined the linguistic components of the definitions [13, 51, 117]. This involved tokenizing the definitions and analyzing the relationships between key linguistic elements such as nouns, adjectives, and verbs. For fears and hopes, we focused on identifying the most prevalent themes, and examining differences among specific subgroups of the public. Prevalent themes were derived through inductive coding, where two authors systematically reviewed unique fears and hopes, grouping them into themes based on established qualitative methodologies [66, 68, 88]. We included themes with input from at least ten participants and reported their occurrences, with participant quotes marked with "P". Differences among specific public subgroups were identified by analyzing their hopefulness scores in two ways. First, we examined opposing perspectives within subgroups by dividing them into segments with contrasting average scores (positive vs. negative). Second, we analyzed outlier perspectives within high-variance subgroups, focusing on participants whose scores deviated by more than one standard deviation, representing extreme fearfulness or hopefulness. These differences were further explored through inductive coding [66, 68, 88].

RQ₂: What are the views of AI influencers?

In a way similar to *RQ₁*, we answered *RQ₂* in a quantitative and qualitative way. Quantitatively, we calculated the number of unique hopes and fears by following a four-step process. First, we collected all fears and hopes sentences from the Time magazine [108] using qualitative content analysis, a commonly used method in HCI for analyzing interview data [57, 80]. To do this, two authors independently analyzed the same subset of 25 interviews, identifying excerpts that reflected influencers' fears and hopes. They started by looking for explicit phrases, such as "Influencer fears/hopes that...". If these phrases were not present, they focused on direct quotes from the interviews. When neither explicit phrases nor direct quotes contained clear fears or hopes, they carefully examined the rest of the interview text for any contextual cues that implied

fears or hopes. Second, we conducted a collaborative coding session with the entire research team to compare the independently marked excerpts. During the session, we discussed and resolved discrepancies and iteratively refined a shared coding frame to classify direct quotes and contextual cue excerpts as either fears or hopes. This refined coding frame was subsequently applied to the remaining interviews, with the same two authors independently coding another set of 25 interviews. Throughout this process, the authors maintained ongoing dialogue to resolve uncertainties and ensure consistent application of the coding frame. Third, to improve clarity and comparability between fears and hopes, two authors employed ethical fabrication [64], making minor adjustments to unify the style of excerpts while preserving their original meaning. For example, they began each fear or hope with an active verb to emphasize the intended action or outcome behind the excerpt. To illustrate, consider this excerpt from an interview with a female CEO: she "hopes its data can facilitate the preservation of [...] forests and other natural environments." This excerpt was categorized as a "hope" and adjusted to read as "develop AI to help preserve natural environments." Finally, as the fourth step, we identified and removed duplicates and counted the resulting overall number of fears and hopes. Since we did not have any ratings from influencers regarding their individual levels of fears and hopes, we did not compute the hopefulness score for them.

Qualitatively, we thematically analyzed the fears and hopes of influencers through an inductive thematic analysis [66, 68, 88]. Starting with a subset of the data, two authors independently performed open coding and discussed their findings with the research team to establish a consensus on code format and granularity. Using this coding framework, the same two authors independently coded the remaining data, and resolved major disagreements through discussion. Finally, the entire research team grouped the codes into key themes. We report them alongside their occurrences and illustrative quotes from influencers, which are marked with "I" for clarity.

RQ₃: Which subgroups of AI influencers have views that are most (or least) aligned with our sample of the public?

We answered this question in a quantitative way by defining four metrics:

- The public's ranking of fears and hopes of AI influencers. In the fourth section of the platform, all participants were asked to vote on 10 randomly chosen pairs (Figure 3D), where the pairs could be either two hopes, two fears, or one of each; these pairs were based on a set of 20 representative fears and 20 representative hopes from AI influencers. For each fear or hope, we calculated a *Q*-score [89], a measure reflecting how frequently it was selected over other options in pairwise comparisons. This score was determined using the "win" and "loss" ratios from all participants comparisons, ensuring a fair and balanced evaluation. We then sorted the fears and hopes in descending order by their *Q*-scores, transforming the pairwise comparisons into an ordered list with the most important fears and hopes of the public at the top and the least important at the bottom.
- The ranking of fears and hopes of AI influencers by subgroups of the public. We analyzed the voting preferences of mutually exclusive public subgroups (i.e., younger vs. older

participants). To do this, we filtered the votes from all participants to isolate those from each subgroup. Using these subgroup-specific votes, we recalculated the Q -score [89] for each fear and hope and then sorted them in descending order to create a ranked list that reflects the preferences of each subgroup.

- The misalignment score between subgroups of AI influencers and the public as a whole. We analyzed different, mutually exclusive subgroups of AI influencers (i.e., younger influencers vs. older influencers, Silicon Valley workers vs. those in other locations, white individuals vs. people of color)². To compute the misalignment score, we compared the public’s ordered list of fears and hopes with an unordered list of fears and hopes from different AI influencer subgroups, and did so using the method described by Hu et al. [42] (Equation §2):

$$\text{misalignment_score}_s = \frac{\sum_i \text{importance}_i \cdot \text{rank}_i}{\sum_i \text{importance}_i} \quad (2)$$

The importance_i of each sentence i (whether it is a fear or a hope) was determined by the number of times that this fear or hope was mentioned by AI influencers from a subgroup s . The rank_i represented the percentile position of a sentence i on the public’s ordered list, with 0% indicating the top and 100% the bottom. Lower misalignment score values ($\text{misalignment_score}_s$) indicated that the public ranks sentences considered important by AI influencers closer to the top of their ranked list. In this context, lower values were preferable: a score of 0 signified complete alignment, 0.5 signified no alignment (a result that is essentially random), and 1 signified complete misalignment. When calculating the misalignment score, we accounted for the different number of sentences in the unordered lists of fears and hopes, which varied depending on the subgroup of AI influencers being analyzed (Appendix B.2, Figure 10).

- The misalignment score between subgroups of AI influencers and the matching subgroups of the public. We analyzed how subgroups of AI influencers aligned with their matching subgroups of the public (i.e., younger influencers with younger participants). To do this, we calculated the misalignment score for each subgroup pair by comparing the unordered list of fears and hopes from the AI influencer subgroup with the ordered list of fears and hopes from the corresponding public subgroup.

RQ4: What themes emerge from the most (or least) aligned views? We answered this question in a qualitative way by conducting a thematic analysis of the unique hopes and fears expressed by subgroups of AI influencers whose views were either most aligned or least aligned with those of the public. To do this, two authors first analyzed the misalignment scores between AI influencer subgroups and their corresponding public subgroups, calculated in the previous step (RQ3). This allowed them to identify which public subgroups were the most and least aligned with the influencers.

²These binarized forms, although problematic because they generalize complex identities into broad categories—thereby overlooking intersectionality and individual nuances—are required for the statistical significance of our statistical analyses and allow us to systematically compare subgroup differences in misalignment scores.

Next, they examined the hopes and fears in these aligned and misaligned subgroups using an inductive thematic analysis [66, 68, 88]. Each theme was supported by quotes from at least five influencers (marked with “I”).

4.3.3 Execution. We used Prolific’s built-in screeners to control for the participants’ geographic location, age, sex, ethnicity, and political affiliation. We used stratified random sampling to represent U.S. census demographics in terms of age (12% in range 18–24 years, 17% in range 25–34 years, 17% in range 35–45 years, 15% in range 45–54 years, 17% in range 55–64 years, 13% in range 65–74 years, and 9% over 75 years), sex (50% female, 50% male), ethnicity (63% White, 11% Black, 11% Mixed, 7% Asian, 8% Other), and political affiliation (30% Republicans, 30% Democrats, 40% Independent). We limited our participant pool to individuals residing in the U.S. All participants were paid on average about \$12 (USD) per hour.

We developed a web-based survey and administered it on Prolific (Appendix B, Figure 11). The survey comprised four pages, each corresponding to a setup step plus a final confirmation screen. To ensure response quality, we implemented four real-time content validation measures, following best practices for conducting crowdsourcing studies [24, 79, 104]. First, we set word ranges of 70–350 characters for open-ended questions not only to ensure participants provided concise, relevant answers but also to prevent survey fatigue. Second, we disabled the ability to paste content from external sources or edit previous responses, ensuring that all answers were original and thoughtfully composed. Third, we tracked clicks on platform’s visual elements (e.g., portraits of AI influencers) and monitored time spent on each survey step.

Fourth, we incorporated three attention checks throughout the survey to identify low-quality responses, in line with Prolific’s attention and comprehension check policy [84]. In the *Demographics and AI Literacy Assessment* (Figure 3C), participants encountered two instructional manipulation checks. In the first, they were explicitly instructed to “Select ‘Blue’ when asked for your favorite color”. In the second, they were told to “Strongly disagree or disagree with the idea that AI technologies are mainly developed by little squirrels”. In the *Pairwise Comparison of AI Views* (Figure 3D), participants were asked to vote on a pair of statements, one of which included a nonsensical fear: “[I wish I had spent the money to] prevent AI from turning all humans into pineapples overnight”. This statement should not be preferred over the second option, which was drawn from our dataset of the fears and hopes of AI influencers. To be included in the final analysis, participants had to pass all three attention checks and read the fears and hopes of at least five AI influencers.

5 Comparing Public and Influencer Views on AI

RQ1: What are the views of our sample of the U.S. public on AI?

Self-Reported Definitions of AI. We derived four clusters of AI definitions, ranked from most to least common. The largest cluster, predictive system ($n = 132$), describes AI as an analytical tool capable of processing vast amounts of information to identify patterns and generate predictions. The second most common cluster, human-like companion ($n = 81$), defines AI as a system designed to mimic human behavior such as conversation or emotional support.

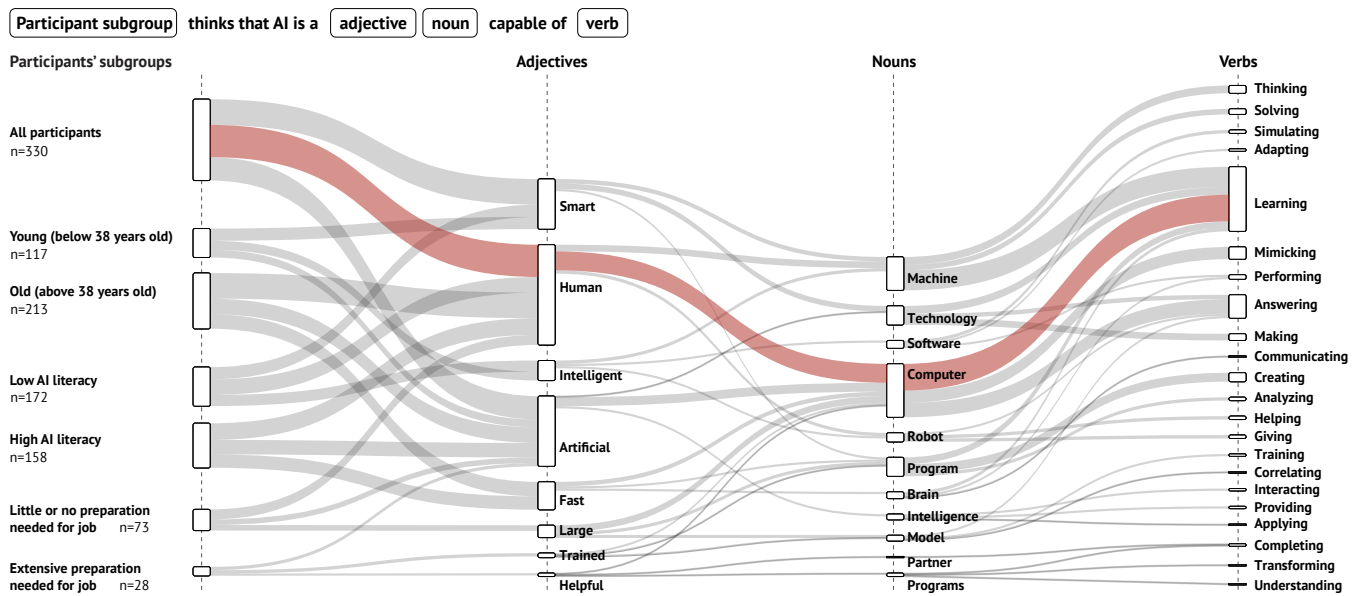


Figure 4: Self-reported AI definitions across subgroups in our sample of the U.S. public. The chart reconstructs these definitions from left to right, highlighting popular adjectives (“human”, “artificial”, “smart”), nouns (“computer”, “machine”, “technology”, “program”), and verbs (“learning”, “answering”, “creating”). The most common definition across participants is: “AI is a human-like computer capable of learning” (path marked in red).

The disruptive agent cluster ($n = 76$) characterizes AI as a transformative force capable of reshaping jobs and societal norms. Finally, the time-saving assistant cluster ($n = 41$) views AI as a tool designed to reduce effort in daily activities. Figure 4 shows how different subgroups of participants define AI using combinations of adjectives, nouns, and verbs. Subgroup definitions can be reconstructed by following the chart from left to right. The most common definition across participants is: “AI is a human-like computer capable of learning”. Distinctions emerge between subgroups: older and high AI literacy participants prefer adjectives like “artificial” and “fast”, while younger and low AI literacy participants favor “smart” and “intelligent”. Those with extensive preparation (e.g., over five years of working experience) use terms like “training”, “model”, and “program”, reflecting a more technical understanding of AI. The top three verb types used to describe AI capabilities [117] were “acting” ($n = 300$), “generating” ($n = 130$), and “identifying” ($n = 60$).

Fears and Hopes About AI. We identified 86 distinct fears and 75 distinct hopes among the public, which we grouped into five key themes (Appendix B.3). The most prevalent theme ($n = 144$) is reduced employment opportunities. Participants view AI as a tool that could drastically reshape the job market by making many roles obsolete, particularly those involving physical or creative tasks. Specific fears include AI being favored by companies over human employees due to lower costs, such as “not needing to provide benefits like health insurance or time off” (P6), which could trigger a chain reaction, leading to increased poverty, social unrest, and a deteriorating quality of life for many. The second theme centers on advances in social services ($n = 74$), with medicine mentioned most often. Participants are hopeful that AI can help cure diseases, improve diagnoses, and create personalized treatments. However,

they are concerned about over-reliance on technology and potential misdiagnoses: “[...] I am worried that they won’t be able to correctly diagnose humans. I am worried that us healthcare workers will be liable for their mistakes, if/when they make them” (P80). The third theme focuses on the irresponsible use of AI ($n = 66$), with participants expressing concerns about AI being used for criminal activities, unauthorized surveillance, and exploitation by corporations and governments. They also worry about the environmental impact due to AI’s energy consumption and the potential for AI to violate intellectual property rights. The overarching fear is that without proper regulation and ethical guidelines, AI could be misused in ways that cause significant harm. The fourth theme focuses on the potential of AI to enhance productivity by handling repetitive tasks ($n = 65$). Participants highlighted several specific examples where AI could streamline work tasks, such as automating data entry, writing emails, generating reports, scheduling, and summarizing meetings. Additionally, AI is seen as a tool to improve everyday tasks, such as creating grocery lists and assist with creative projects. For example, P77 used AI to “generate character ideas for an upcoming role-playing game campaign, including their backstories and character art”. The fifth theme focuses on the increased spread of misinformation through AI ($n = 55$), with most participants citing deepfakes, including “videos of public figures saying things they never did” (P55), and “AI impersonating the voices of loved ones” (P304).

Hopefulness in Particular Subgroups of the Public. Overall, the results show that typical public views depend on three key factors: knowledge (i.e., AI definition articulation, AI literacy, and training and education requirements), gender, and generation. Our participants reported an average hopefulness score of 0.8 on a scale ranging from -10 to 10, indicating a sense of optimism (Figure 5).

The most notable differences were associated with knowledge (as shown by the diverging black bars in Figure 5, representing subgroups based on AI literacy and job training and education requirements). Participants who perceived AI as merely a tool for predictions were less hopeful than those who attributed transformative capabilities to it. Individuals with low AI literacy reported lower hopefulness compared to those with high literacy. Those in occupations requiring extensive training and education were less hopeful than those in roles requiring little to moderate training. Males were less hopeful than females, while older participants were less hopeful compared to their younger counterparts. Participants with diverging levels of knowledge shared similar fears, such as reduced job opportunities ($n = 22$ each) and AI bias ($n = 18$ each). Their hopes were also similar, focusing on advancements in social services ($n = 52$ for less knowledgeable, $n = 45$ for more knowledgeable) and technical innovation ($n = 45$ for less knowledgeable, $n = 35$ for more knowledgeable). However, their perspectives differed in complexity. Those with less knowledge tended to focus on personal or immediate effects, offering simpler examples (e.g., “reducing tiredness”). In contrast, participants with more knowledge highlighted broader impacts, providing more nuanced perspectives that weighed both the benefits and drawbacks (e.g., “promoting equality while increasing energy consumption”).

Subgroup differences are further influenced by factors such as job training and education requirements, ethnicity, and political affiliation. The most differing views were found among participants whose jobs required little to no training or education, white participants, and independent supporters, as indicated by the highest standard deviations (Figure 5). Among them, the most fearful participants with minimal training or education expressed concerns about a “decline in inventive thinking” ($n = 3$), while participants with moderate views prioritized “the spread of misinformation” ($n = 5$). Highly optimistic participants highlighted innovations like “new AI programs for arts” ($n = 3$), while those with moderate views saw AI as a way to “improved social connectedness” ($n = 5$). White participants also showed distinct patterns. The most fearful highlighted integration and control issues, such as “AI becoming uncontrollable” ($n = 15$). In contrast, average white participants expressed a broader range of fears, including “personal creativity losses” ($n = 10$) and “harmful military uses” ($n = 15$). On the hopeful side, highly optimistic white participants focused on tackling major societal challenges like “advancing social services” ($n = 20$) and “creating technological innovation” ($n = 10$), while average white participants highlighted more practical applications like “handling repetitive tasks” ($n = 30$). The most concerned independent supporters feared the “decline in inventive thinking” ($n = 3$) and “overreliance on AI” ($n = 7$), while the average highlighted “AI gaining consciousness” ($n = 10$). On the hopeful side, highly optimistic supporters shared similar hopes with the average, but expressed them more passionately, such as P105, who hoped that “digital mind could find a way to help cure cancer and survive in space”.

RQ₂: What are the views of AI influencers?

Unlike the members of the U.S. public, whose hopes and fears of AI were gathered via the platform, our analysis of AI influencers relied on data from Time magazine [108]. We identified 56 unique

fears and 71 unique hopes, which we grouped into five key themes. The most common theme was reduced employment opportunities ($n = 13$), focusing on job replacement and work precarization. As I23 highlighted, “[...] we should avoid poor working conditions for African workers in AI development.” The second theme focused on the loss of personal privacy ($n = 12$), with concerns centered on data consent and tracking, emphasizing the need to prevent AI misuse in surveillance. The third theme revolved around irresponsible use of AI ($n = 11$), with a strong emphasis on the need for effective regulation and oversight of AI technologies, such as “[...] a global AI governing body to oversee research and ensure fair use” (I30). The fourth theme focused on increased innovation ($n = 10$) across various industries, such as using AI to “cut waste in the fashion industry” (I40) and to “assess wildfire risks for insurance and finance sectors” (I70). The fifth theme centered on AI increasing global conflicts ($n = 8$), with influencers discussing how AI could shift global power dynamics and potentially lead to arms races, widening divide between “democratic and authoritarian countries” (I67). Less prevalent themes include advancing social services, such as healthcare ($n = 7$), through “improved medical diagnosis and treatment of tuberculosis patients” (I69), and enhancing access to culture ($n = 6$) by “developing legal actions to protect artists’ rights and ensure fair compensation for their work used to train AI” (I56) or “enabling everyone to make their own movies with AI” (I33).

RQ₃: Which subgroups of AI influencers have views that are most (or least) aligned with those of our sample of the U.S. public?

The results show that the views of the influencers depend on three key factors: generation, professional influence, and geography.

Subgroups of AI Influencers Most Aligned With the Members of the U.S. Public. Figure 6 shows the AI influencers’ alignment scores per 14 subgroups based on demographics. The young influencers (under 38) subgroup aligns most closely with the members of the U.S. public, followed by academics, non-billionaires, white influencers and those working in locations outside Silicon Valley. This indicates that age, wealth, and ethnicity might be important components for the alignment, and that younger generations are better at representing public opinions.

Subgroups of AI Influencers Least Aligned With the Members of the U.S. Public. Mirroring the most aligned subgroups, the older influencers (over 38 years old) have the worst alignment with the public. Interestingly, subgroups with monetary influence (e.g., billionaires and Silicon Valley workers) tend to rank lower, highlighting a potential disconnect between groups with very high decision-making power and those impacted by the technology [8, 17, 53]. Figure 7 shows the misalignment scores between the AI influencers subgroups and their corresponding public subgroups. Young influencers are best aligned with young public representatives. In contrast, typically underrepresented AI influencer groups (e.g., females and people of color) show poor alignment with their public counterparts, suggesting their voices are still insufficiently represented even among their own representatives.

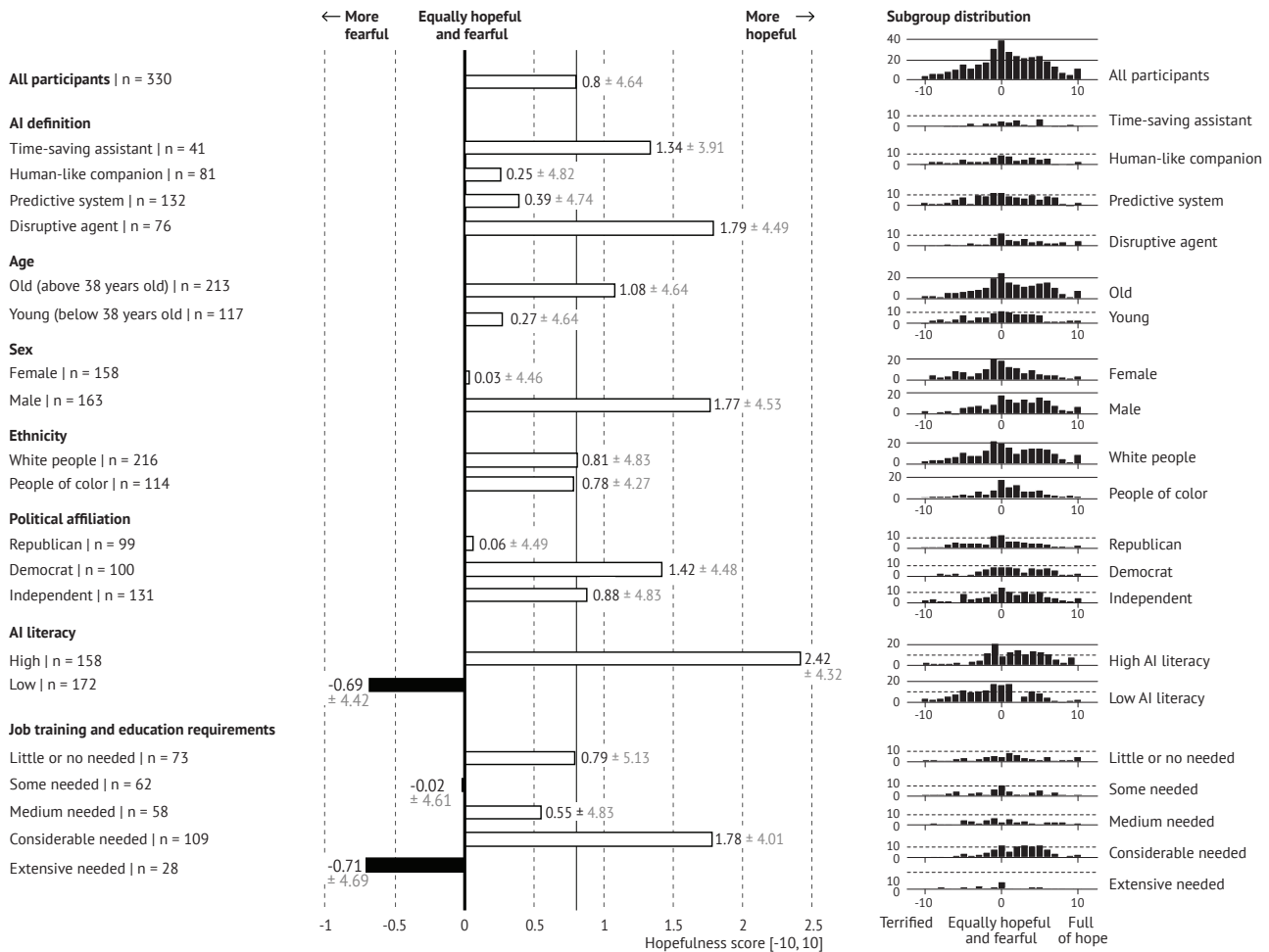


Figure 5: Average hopefulness scores across 20 subgroups based on age, sex, ethnicity, political affiliation, AI literacy, and job training and education requirements. The value “0” indicates being equally hopeful and fearful. Negative values indicate being more fearful, while positive values suggest being more hopeful. Standard deviations are shown as \pm values next to the subgroup scores. The most notable differences in hopefulness are between older vs. younger participants, males vs. females, Democrats vs. Republicans, individuals with low vs. high AI literacy, and those in occupations requiring extensive training and education vs. those requiring less.

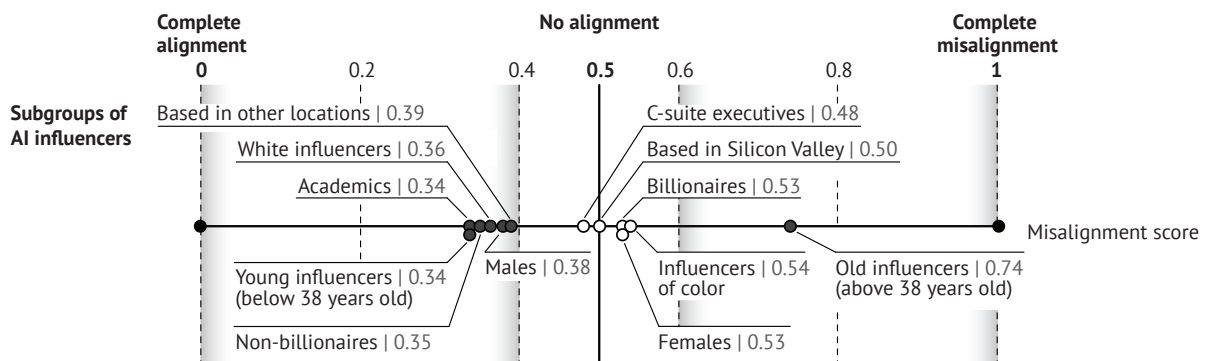


Figure 6: Misalignment scores between subgroups of AI influencers and participants representative of the U.S. population. Young influencers’ views are most closely aligned with those of our participants, followed by academics and non-billionaires.

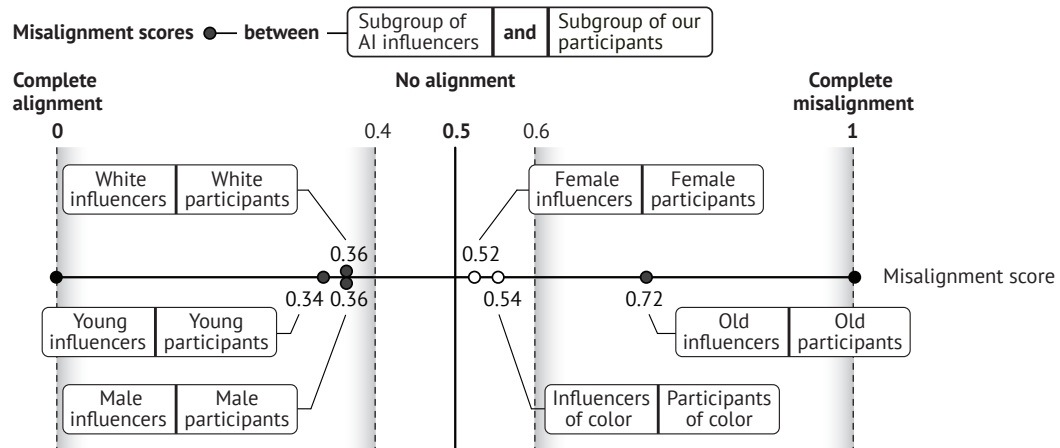


Figure 7: Misalignment scores between subgroups of AI influencers and subgroups of our participants, representative of the U.S. population. Young influencers’ views are most closely aligned with those of young participants.

RQ₄: What themes emerge from the most (or least) aligned views?

Themes From the Most Aligned Views. The key themes emerging from AI influencers closely aligned with public opinion revolve around cautious optimism and focus on three types of control (Tables 1 - 2). First, aligning AI with human values is essential, with a strong emphasis on fairness, equity, and respect for human rights. Ensuring that AI systems operate in ways that reflect these values helps address concerns about bias and ethical misuse. Second, transparency in AI decision-making is crucial for building trust. Many worry about the irresponsible use of AI, bias, and loss of privacy, so making AI systems more transparent, auditable, and understandable is vital for public confidence. People are also willing to accept a slower pace of AI development if it ensures greater safety and responsibility. Finally, concerns about AI centralizing power underscore the need to prevent its control from becoming concentrated in the hands of a powerful few.

Themes From the Least Aligned Views. Older influencers and billionaires diverge significantly from public opinion on AI. Older influencers prioritize democratic AI development, emphasizing diverse voices in its oversight, while the public tends to focus less on these governance issues. They also express unique concerns about AI exacerbating global instability, reflecting a more long-term and geopolitical perspective. Meanwhile, billionaires are optimistic about AI enhancing human productivity, creativity, and leisure, viewing it as a tool for boosting labor efficiency and personal fulfillment. This contrasts with public fears of job losses and economic inequality. There is also a notable difference in how the public and influencers articulate the benefits of AI. Participants tend to describe AI’s advantages in vague, general terms, such as “doing positive things for mankind” (P184), often highlighting medicine and the workplace. In contrast, influencers offer more detailed and articulate descriptions, such as “building AI simulations to improve driverless cars in challenging road scenarios”, covering a wider range of benefits across different fields.

Table 1: Top 25% (Q4) of sentences ranked by participants’ votes (from most to least important) in the large-scale study.

Rank	Sentence	Type
1	Ensure AI upholds human and fundamental rights	Hope
2	Use AI to achieve scientific breakthroughs	Hope
3	Avoid letting AI operate without human oversight	Fear
4	Set clear objectives and strict guardrails for AI	Hope
5	Avoid creating AI that can be used to cause violence	Fear
6	Prevent AI from destroying humanity	Fear
7	Use AI to make people more productive and efficient	Hope
8	Prevent the use of poorly understood and problematic datasets in AI development	Fear
9	Prevent centralizing power from AI in few hands	Fear
10	Build AI that addresses the needs of local communities	Hope

Table 2: Bottom 25% (Q1) of sentences ranked by participants’ votes (from most to least important) in the large-scale study.

Rank	Sentence	Type
30	Remove biases in AI-supported hiring practices	Fear
31	Promote participation in AI from people in under-represented communities	Hope
32	Avoid exploitation of workers in the Global South for data generation	Fear
33	Promote inclusive and democratic AI development	Hope
34	Avoid using AI to develop novel pathogens	Fear
35	Raise more interest in AI among society and states	Hope
36	Include more young people on AI advisory boards	Hope
37	Use AI to improve hiring practices by connecting companies with top talent	Hope
38	Prevent AI from being evaluated using simulated measurements instead of real-world data	Fear
39	Avoid AI-driven competition between democratic and authoritarian countries	Fear
40	Avoid AI slowing down due to stringent regulations	Fear

6 Discussion

By collecting views of 330 U.S. participants using the “The Hall of AI Fears and Hope” platform, we found that older influencers were the least aligned with the public, while younger influencers and academics presented better alignment. Additionally, influencers and the public have distinct concerns, with the public fearing losing control of AI and seeing a limited number of benefits, while influencers focus on controlling AI through regulations and are able to envision more benefits.

6.1 Consistencies with Previous Literature

Our results have confirmed the findings of previous research in three key areas: the relationship between gender, political affiliation, AI literacy and levels of hopefulness about AI; drivers of AI-phobia; and the misalignment between public views and those of AI influencers.

In terms of the first area, there was a greater sense of hope toward AI than fear across all demographic groups, consistent with Nader et al. [71], who found that most U.S. respondents were optimistic about AI’s future and its impact. Additionally, males were generally more hopeful about AI than females, as noted by Sartori and Bocca [90] and Rainie et al. [85]. We found that individuals identifying with the Republican Party were more likely to express negative views about AI compared to Democrats, corroborating findings from Rainie et al. [85]. We also found that low AI literacy leads to lower levels of AI hopefulness, further confirming the link between fear and anxiety, as identified by Schiavo et al. [94].

In terms of the second area, the thematic analysis of public views confirmed the main drivers of AI-phobia identified by Lee and Park [61]: AI substitutability (the fear of being replaced by AI), AI accountability (uncertainty over who is responsible when AI is involved), AI literacy (limited understanding of how AI works), and AI fever (the compulsive rush to adopt AI without a clear purpose or understanding). These drivers highlight key anxieties the public faces when interacting with or thinking about AI technologies. Participants were particularly concerned with AI’s impact on two domains: employment and healthcare. These areas, which directly affect individuals’ livelihoods and well-being, are repeatedly reported as major areas of concern [36, 71].

In terms of the third area, our findings show that views of under-represented subgroups—such as women and people of color—are not well-aligned with the general public’s views or even with members of their own demographic groups among the influencers (see Figure 7). These results echo critiques from liberal feminism literature, which argues that increasing representation without making structural changes often perpetuates existing biases. Representatives from these groups tend to conform to the dominant, powerful AI culture rather than challenge it [3]. Similarly, Noble [74] notes that representatives of people of color in big tech companies often adapt to prevailing norms rather than advocating for the needs of their communities. This lack of genuine representation is further compounded by the concentration of power among a few wealthy individuals. Billionaires, who control substantial investments in AI technology, hold enormous influence over its future direction and applications, which may not always align with the interests of the

broader public. For example, women receive only 0.7-3.4% of funding in AI startups in UK [45], meaning these technologies primarily reflect the priorities of wealthy white men. This funding disparity explains why women’s perspectives are often underrepresented in AI, as the technologies are shaped by the interests of those in positions of economic power.

6.2 Inconsistencies with Previous Literature

Our results have provided new insights into two key areas: the relationship between age, occupational preparation levels, and attitudes toward AI, as well as the alignment between public opinions and those of AI influencers.

Research has traditionally suggested that older adults tend to be more cautious, skeptical, or anxious about adopting new technologies, including AI. However, our findings contradict this by showing that older individuals may, in fact, exhibit more optimism about AI in specific contexts, particularly healthcare. Previous studies have indicated that older populations often view AI positively when it is framed as a tool to enhance quality of life, improve healthcare services, or assist with caregiving tasks [85]. This suggests that context plays a significant role in shaping their attitudes toward AI, especially when the technology directly benefits them in a practical, personal way. In contrast, individuals in occupations requiring extensive preparation (e.g., those needing advanced degrees such as a master’s or Ph.D.) were more negative about AI compared to those in roles requiring less preparation. These occupations often involve coordinating, training, or managing others, where advanced communication and organizational skills are essential. This negativity contrasts with prior research, which has generally posited that highly educated and skilled individuals are more receptive to technological advancements. Our results challenge this view, revealing that despite their high levels of expertise, these professionals may feel more threatened by AI’s potential disruption of their specialized roles.

In relation to the opinions of influential people on AI, previous research has not extensively assessed their perspectives. Our findings indicate that AI influencers, like the general public, tend to be more hopeful than fearful about AI. However, the themes within their opinions differ significantly. While both groups share concerns about AI’s impact on the job market, AI influencers focus on broader, societal implications, such as the economic shifts AI might cause and the nuanced benefits of its implementation across industries. In contrast, the public often approaches these issues from a personal perspective, shaped by everyday AI experiences.

This divergence in focus stems from the depth of knowledge that AI influencers possess. Drawing from professional training, academic literature, conferences, and direct involvement in AI development, influencers are equipped with a more sophisticated understanding of AI’s potential. One significant point that arises from this disparity is the concept of “AI-fabrication” [61], where AI-literate individuals, due to their advanced technical skills, are in a position to design and manipulate AI systems to serve particular interests, often without the general public’s knowledge or understanding. This creates an information asymmetry, where the public remains unaware of whether AI systems have been altered to favor certain outcomes, leaving them vulnerable to biased algorithms.

Our findings also reveal a contrast between the alignment of public views with academia and their divergence from those of tech elites, such as Silicon Valley workers and billionaires. During the COVID-19 pandemic, academia actively worked to bridge the gap between expert knowledge and public understanding through accessible scientific communication [115]. While it is unclear if similar efforts could contribute to the alignment of views on AI, this highlights the important role academics can play in educating the public and shaping AI development in ways that prioritize societal well-being over corporate or personal interests.

6.3 Implications

From a theoretical perspective, our study contributes to the growing research on public views of AI. It is important not only to understand what the public thinks about AI but also to explore how and where these views are formed [13, 81]. This deeper understanding is essential for bridging the gap between influencers and the public. While past studies have focused on public views [13, 51, 90, 91], the formation of these views has received less attention. However, without a clear understanding of public concerns and values, AI policies risk becoming disconnected from societal needs and may ultimately lack legitimacy. This disconnect could lead to increased public resistance or distrust toward AI technologies, undermining the societal benefits of AI and diminishing public willingness to support AI-driven initiatives. The public's limited capacity to articulate the benefits (hopes) of AI against its risks (fears) seems tied to their lack of knowledge about AI's uses and capabilities. To address this, we must consider that public opinion, although often shaped by media and influencers, represents real and evolving attitudes that can impact the social acceptance and ethical implications of AI technologies. The design of more interactive, participatory platforms that engage the public in the co-creation and critique of AI could be crucial. By incorporating concepts such as “value levers” [98], where organizational processes make value judgments explicit and open to public debate, AI literacy initiatives can transform abstract concerns into practical considerations for system design. These “value levers” can serve as a bridge between AI influencers and the public, fostering more transparent dialogue about the trade-offs in AI development and use. Additionally, AI literacy can be increased through seminars, literature, and development experiences that are more accessible and appealing to the general public. Increasing public understanding and involvement in AI-related discussions will help ensure that AI technologies are not only responsibly developed but also socially accepted, which is essential for their long-term success and alignment with societal values [58]. This also aligns with the goals of responsible AI [103], where practitioners are encouraged not only to follow value priorities but to justify when their choices deviate from commonly held preferences [46].

From a practical perspective, our platform served as a data collection tool to gather public opinions on AI, going beyond traditional methods. While the tool used should not limit how we understand the public's hopes and fears about AI, most participants in our study preferred our platform over traditional surveys. Although it does not solve all the problems with traditional surveys [104] and is not a one-size-fits-all solution, it represents a step toward more engaging ways of collecting opinions. Many participants, including 62% who

generally preferred the platform over traditional surveys, stated that using the tool felt more enjoyable and exploratory, rather than like a typical research study (Appendix A.4.2). Our platform also addresses calls for improved methods of gathering public opinion through scientifically designed experiments that clearly explain ethical dilemmas and help shape AI policies based on these insights [5, 34]. For example, Polis is a real-time system designed to gather the views of large groups in their own words and identify common ground and areas of disagreement [100]. Similarly, BetterBeliefs is a platform promoting virtuous online behaviors by encouraging evidence-based and rational discourse [23].

6.4 Limitations and Future Work

Our work comes with four main limitations that call for future research efforts. The first limitation lies in our methods for compiling the dataset of AI influencers' fears and hopes. We selected the Time100 AI list [108] list for its rigorous editorial process and inclusion of structured, in-depth interviews. However, public media interviews, while useful for accessing diverse perspectives, are influenced by biases such as agenda-setting, media framing, and social desirability effects [33, 39, 65]. Future studies could benefit from systematically curated datasets of AI influencers. While a definitive “ground truth” list may not be possible, it is essential to diversify the selection to ensure a range of perspectives and the inclusion of highly reputable voices such as leading scientists, technologists, policymakers, and thought leaders. For example, Struckman and Kupiec [102] reached out to over 100 first signatories of the open letter calling for a six-month pause on the training of powerful AI systems, demonstrating a method of targeting key figures engaged in public AI debates. The Expert Survey on Progress in AI [37] systematically surveyed authors who published in major machine learning conferences such as NeurIPS and ICML (2016, 2022), and in 2023 expanded to include ICLR, AAAI, IJCAI, and JMLR. Public investigations have also been conducted into participants of international AI standard-setting bodies [18], and national policymaking bodies such as the U.S. Congress [95]. While alternative sampling methods for influencers can be explored, several challenges remain in gathering reliable data on their AI-related views. Surveys targeting high-profile individuals like industry executives are often hindered by limited access due to time constraints or gatekeeping, reducing the chances of securing comprehensive responses. For example, Struckman and Kupiec [102] secured interviews with just 21 individuals from over 50,000 open letter signatories. Surveys targeting larger groups of AI researchers often face low participation rates. The Expert Survey on Progress in AI [37], conducted every few years after major conferences, achieves between 15–17% response rates—typical for large expert surveys but insufficient for sustained engagement. To mitigate that, AI conferences could integrate opinion-gathering tools into submission processes or use interactive screens for quick surveys during events. Collecting influencers' views from public platforms like YouTube or X (formerly Twitter) could expand datasets but raises concerns about authenticity, reliability, and platform bias. For example, AI-related discussions by policymakers on Twitter often emphasize economic growth and innovation [95], potentially skewing the data toward specific narratives.

The second limitation relates to the exposure of participants to the opinions of those featured in the Time magazine list. Our findings are limited to these influencers' perspectives and cannot be generalized to all viewpoints. Additionally, the comparison between public and influencer opinions was based on data collected using different methods—direct for the public and indirect for influencers. To mitigate this, we post-processed the influencer dataset to align with the public dataset's format. Future work should focus on replicating this experiment in different contexts with various sets of influencers, such as members of the United Nations High-level Advisory Body on Artificial Intelligence [73], members of the Working Party on Artificial Intelligence Governance from the Organization for Economic Cooperation and Development [35], or nominees from the 100 Brilliant Women in AI Ethics list [116].

The third limitation concerns the participant sample, consisting of 330 U.S.-based individuals, representative in terms of age, sex, ethnicity, and political affiliation. However, relying solely on U.S. participants limits the study's generalizability, as insights and patterns may differ across countries [36, 51]. Future work should expand the methodology to include larger and more diverse international samples for broader applicability.

The fourth limitation concerns the design of our platform for data collection. While it generated an engaging visualization that appeals to the broader public beyond traditional surveys, developing such tools is time-consuming due to the need for literature reviews, co-design, and visualization implementation. The platform is also more difficult to deploy and modify compared to standard surveys. Additionally, the platform's focus on fears and hopes might oversimplify public opinion into a black-and-white view. However, we believe this approach helps people think more deeply about AI and is suited for larger, well-planned studies requiring detailed analysis, rather than small-scale data collection.

7 Conclusion

Our study highlights a significant disconnect between the priorities of some members of the U.S. public and those of influential figures in AI, particularly older AI influencers and billionaires who are less in tune with public concerns. This misalignment suggests that the voices shaping AI development may not fully represent the broader societal concerns. Future research should explore how these disparities influence AI outcomes and consider mechanisms for incorporating public opinion more effectively into AI Governance. Specifically, HCI can play a crucial role in designing participatory mechanisms that enable diverse public voices to engage in AI policymaking. By leveraging user-centered design principles, HCI researchers and practitioners can help create tools and processes that make public input both scalable and actionable in shaping AI governance.

Acknowledgments

The work has been funded by Nokia Bell Labs. It has also been partially supported by the European Union's Horizon 2020 Research and Innovation Programme under Grant Agreement No 739578 and the Government of the Republic of Cyprus through the Deputy Ministry of Research, Innovation and Digital Policy under grant agreement No 101061303.

References

- [1] Betty Adamou. 2018. *Games and Gamification in Market Research*. Kogan Page.
- [2] Samira Adus, Jillian Macklin, and Andrew Pinto. 2023. Exploring Patient Perspectives on How They Can and Should Be Engaged in the Development of Artificial Intelligence (AI) Applications in Health Care. *BMC Health Services Research* 23, 1 (2023), 1163.
- [3] C. Arruzza, T. Bhattacharya, and N. Fraser. 2019. *Feminism for the 99 Percent: A Manifesto*. Verso Books.
- [4] Association for Computing Machinery. 2025. *ACM Digital Library*. Retrieved February 05, 2025 from <https://dl.acm.org/>
- [5] Edmond Awad, Michael Anderson, Susan Leigh Anderson, and Beishui Liao. 2020. An Approach for Combining Ethical Principles With Public Opinion To Guide Public Policy. *Artificial Intelligence* 287 (2020), 103349. <https://doi.org/10.1016/j.artint.2020.103349>
- [6] Stevie Bergman, Nahema Marchal, John Mellor, Shakir Mohamed, Iason Gabriel, and William Isaac. 2024. STELA: A Community-Centred Approach to Norm Elicitation for AI Alignment. *Scientific Reports* 14, 1 (2024), 6616. <https://doi.org/10.1038/s41598-024-56648-4>
- [7] Pei-Luen Patrick Rau Bingcheng Wang and Tianyi Yuan. 2023. Measuring User Competence in Using Artificial Intelligence: Validity and Reliability of Artificial Intelligence Literacy Scale. *Behaviour & Information Technology* 42, 9 (2023), 1324–1337. <https://doi.org/10.1080/0144929X.2022.2072768> arXiv:<https://doi.org/10.1080/0144929X.2022.2072768>
- [8] Abeba Birhane, William Isaac, Vinodkumar Prabhakaran, Mark Diaz, Madeleine Clare Elish, Iason Gabriel, and Shakir Mohamed. 2022. Power to the People? Opportunities and Challenges for Participatory AI. In *Proceedings of the ACM Conference on Equity and Access in Algorithms, Mechanisms, and Optimization (EAAMO)*. 1–8. <https://doi.org/10.1145/3551624.3555290>
- [9] Edyta Bogucka, Marios Constantinides, Sanja Šćepanović, and Daniele Quercia. 2024. Co-designing an AI Impact Assessment Report Template with AI Practitioners and AI Compliance Experts. *Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society* 7, 1 (Oct. 2024), 168–180.
- [10] Nathan Calvin and Jade Leung. 2020. Who Owns Artificial Intelligence? A Preliminary Analysis of Corporate Intellectual Property Strategies and Why They Matter. https://www.fhi.ox.ac.uk/wp-content/uploads/Patents_FHI-Working-Paper-Final-.pdf. Accessed: 2024-08-14.
- [11] Steve Campbell, Melanie Greenwood, Sarah Prior, Toniele Shearer, Kerrie Walkem, Sarah Young, Danielle Bywaters, and Kim Walker. 2020. Purposive Sampling: Complex or Simple? Research Case Examples. *Journal of Research in Nursing* 25, 8 (2020), 652–661. <https://doi.org/10.1177/1744987120927206>
- [12] Ricardo JGB Campello, Davoud Moulavi, and Jörg Sander. 2013. Density-Based Clustering Based on Hierarchical Density Estimates. In *Pacific-Asia Conference on Knowledge Discovery and Data Mining*. Springer, 160–172. https://doi.org/10.1007/978-3-642-37456-2_14
- [13] Stephen Cave, Kate Coughlan, and Kanta Dihal. 2019. "Scary Robots" Examining Public Responses to AI. In *Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society (AIES)*. 331–337. <https://doi.org/10.1145/3306618.3314232>
- [14] Stephen Cave and Kanta Dihal. 2019. Hopes and Fears for Intelligent Machines in Fiction and Reality. *Nature Machine Intelligence* 1, 2 (2019), 74–78. <https://doi.org/10.1038/s42256-019-0020-9>
- [15] Center for Deliberative Democracy and Crowdsourced Democracy Team, Stanford University. 2024. *The Stanford Online Deliberation Platform*. <https://deliberation.stanford.edu/tools-and-resources/online-deliberation-platform>
- [16] Marios Constantinides, Edyta Bogucka, Daniele Quercia, Susanna Kallio, and Mohammad Tahaei. 2024. RAI Guidelines: Method for Generating Responsible AI Guidelines Grounded in Regulations and Usable by (Non-)Technical Roles. *Proc. ACM Hum.-Comput. Interact.* 8, CSCW, Article 388 (Nov. 2024), 28 pages. <https://doi.org/10.1145/3686927>
- [17] Eric Corbett, Emily Denton, and Sheena Erete. 2023. Power and Public Participation in AI. In *Proceedings of the ACM Conference on Equity and Access in Algorithms, Mechanisms, and Optimization (EAAMO)*. 1–13. <https://doi.org/10.1145/3617694.3623228>
- [18] Corporate Europe Observatory. 2025. *Bias Baked In. How Big Tech Sets Its Own AI Standards*. <https://corporateeurope.org/en/2025/01/bias-baked>
- [19] Fernando Delgado, Stephen Yang, Michael Madaio, and Qian Yang. 2023. The Participatory Turn in AI Design: Theoretical Foundations and the Current State of Practice. In *Proceedings of the ACM Conference on Equity and Access in Algorithms, Mechanisms, and Optimization (EAAMO)*. 1–23. <https://doi.org/10.1145/3617694.3623261>
- [20] Wesley Hanwen Deng, Solon Barocas, and Jennifer Wortman Vaughan. 2024. Supporting Industry Computing Researchers in Assessing, Articulating, and Addressing the Potential Negative Societal Impact of Their Work. arXiv:2408.01057 [cs.HC] <https://arxiv.org/abs/2408.01057>
- [21] Nathan Dennler, Anaelia Ovalle, Ashwin Singh, Luca Soldaini, Arjun Subramanian, Huy Tu, William Agnew, Avijit Ghosh, Kyra Yee, Irene Font Peradejordi, et al. 2023. Bound by the Bounty: Collaboratively Shaping Evaluation Processes for Queer AI Harms. In *Proceedings of the AAAI/ACM Conference on AI, Ethics,*

- and Society (AIES). 375–386. <https://doi.org/10.1145/3600211.3604682>
- [22] Advait Deshpande and Helen Sharp. 2022. Responsible AI Systems: Who Are the Stakeholders?. In *Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society (AIES)*. ACM, 227–236. <https://doi.org/10.1145/3514094.3534187>
- [23] S. Kate Devitt, Tamara R. Pearce, Alok Kumar Chowdhury, and Kerrie Mengersen. 2022. A Bayesian Social Platform for Inclusive and Evidence-Based Decision-Making. In *Social Virtue Epistemology*. Routledge, 480–513.
- [24] Tim Draws, Alisa Rieger, Oana Inel, Ujwal Gadiraju, and Nava Tintarev. 2021. A Checklist to Combat Cognitive Biases in Crowdsourcing. *Proceedings of the AAAI Conference on Human Computation and Crowdsourcing (HCOMP)* 9, 1 (Oct. 2021), 48–59. <https://doi.org/10.1609/hcomp.v9i1.18939>
- [25] Fan Du, Catherine Plaisant, Neil Spring, and Ben Shneiderman. 2017. Finding Similar People to Guide Life Choices: Challenge, Design, and Evaluation. In *Proceedings of the ACM Conference on Human Factors in Computing Systems (CHI)* (CHI '17). ACM, 5498–5544. <https://doi.org/10.1145/3025453.3025777>
- [26] Paul Ekman. 1992. An Argument for Basic Emotions. *Cognition & Emotion* 6, 3–4 (1992), 169–200.
- [27] Pardis Emami-Naeini, Joseph Breda, Wei Dai, Tadayoshi Kohno, Kim Laine, Shwetak Patel, and Franziska Roesner. 2023. Understanding People's Concerns and Attitudes Toward Smart Cities. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. 1–24.
- [28] European Commission. 2024. Regulation of the European Parliament and of the Council on Laying Down Harmonised Rules on Artificial Intelligence and Amending Regulations (EU) No 300/2008, (EU) No 167/2013, (EU) No 168/2013, (EU) 2018/858, (EU) 2018/1139 and (EU) 2019/2144 and Directives 2014/90/EU, (EU) 2016/797 and (EU) 2020/1828 (Artificial Intelligence Act). https://www.europa.europa.eu/doceo/document/TA-9-2024-0138-FNL-COR01_EN.pdf. Accessed: 2024-08-20.
- [29] Hugging Face. 2024. Sentence Transformers: all-mpnet-base-v2. <https://huggingface.co/sentence-transformers/all-mpnet-base-v2>. Accessed: 2024-08-14.
- [30] Leon Festinger. 1954. A Theory of Social Comparison Processes. *Human Relations* 7, 2 (1954), 117–140. <https://doi.org/10.1177/001872675400700202> arXiv:<https://doi.org/10.1177/001872675400700202>
- [31] Octavia Field Reid, Anna Colom, and Roshni Modhvia. 2023. What Do the Public Think About AI? Understanding Public Attitudes and How To Involve the Public in Decision-Making About AI. <https://www.adalovelaceinstitute.org/evidence-review/what-do-the-public-think-about-ai>. Accessed: 2024-08-14.
- [32] Figma. 2024. Figma: The Collaborative Interface Design Tool. <https://www.figma.com>. Accessed: 2024-08-20.
- [33] Robert J. Fisher. 1993. Social Desirability Bias and the Validity of Indirect Questioning. *Journal of Consumer Research* 20, 2 (09 1993), 303–315. <https://doi.org/10.1086/209351> arXiv:<https://academic.oup.com/jcr/article-pdf/20/2/303/5074014/20-2-303.pdf>
- [34] Luciano Floridi, Josh Cowsils, Monica Beltrametti, Raja Chatila, Patrice Chazerand, Virginia Dignum, Christoph Luetge, Robert Madelin, Ugo Pagallo, Francesca Rossi, Burkhard Schafer, Peggy Valcke, and Effy Vayena. 2018. AI4People—An Ethical Framework for a Good AI Society: Opportunities, Risks, Principles, and Recommendations. *Minds and Machines* 28, 4 (2018), 689–707. <https://doi.org/10.1007/s11023-018-9482-5>
- [35] Organization for Economic Cooperation and Development. 2024. Working Party on Artificial Intelligence Governance. <https://oecdgroups.oecd.org/Bodies/ShowBodyView.aspx?BodyID=7755&BodyPID=15757&Lang=en&Book=False>. Accessed: 2024-08-14.
- [36] Google. 2024. Our Life With AI: The Reality of Today and the Promise of Tomorrow. https://static.googleusercontent.com/media/publicpolicy.google/en/resources/our_life_with_ai_google_ipsos_report.pdf. Accessed: 2024-08-14.
- [37] Katja Grace, Harlan Stewart, Julia Fabienne Sandkühler, Stephen Thomas, Ben Weinstein-Raun, and Jan Brauner. 2024. Thousands of AI Authors on the Future of AI. arXiv:2401.02843 <https://arxiv.org/abs/2401.02843>
- [38] Simone Grassini and Alexander Sæviold Ree. 2023. Hope or Doom AI-titude? Examining the Impact of Gender, Age, and Cultural Differences on the Envisioned Future Impact of Artificial Intelligence on Humankind. In *Proceedings of the European Conference on Cognitive Ergonomics (ECCE)*. ACM, Article 15, 7 pages. <https://doi.org/10.1145/3605655.3605669>
- [39] Tim Groseclose and Jeffrey Milyo. 2005. A Measure of Media Bias. *The Quarterly Journal of Economics* 120, 4 (11 2005), 1191–1237. <https://doi.org/10.1162/003353505775097542> arXiv:<https://academic.oup.com/qje/article-pdf/120/4/1191/5431230/120-4-1191.pdf>
- [40] Greg Guest, Arwen Bunce, and Laura Johnson. 2006. How Many Interviews Are Enough? An Experiment With Data Saturation and Variability. *Field Methods* 18, 1 (2006), 59–82. <https://doi.org/10.1177/1525822X05279903>
- [41] Michel Hohendanner, Chiara Ullstein, Yosuke Buchmeier, and Jens Grossklags. 2023. Exploring the Reflective Space of AI Narratives Through Speculative Design in Japan and Germany. In *Proceedings of the ACM Conference on Information Technology for Social Good (GoodIT)*. 351–362. <https://doi.org/10.1145/3582515.3609554>
- [42] Yifan Hu, Yehuda Koren, and Chris Volinsky. 2008. Collaborative Filtering for Implicit Feedback Datasets. In *IEEE International Conference on Data Mining. IEEE*, 263–272. <https://doi.org/10.1109/ICDM.2008.22>
- [43] Saffron Huang, Divya Siddarth, Liane Lovitt, Thomas I Liao, Esin Durmus, Alex Tamkin, and Deep Ganguli. 2024. Collective Constitutional AI: Aligning a Language Model with Public Input. In *Proceedings of the ACM Conference on Fairness, Accountability, and Transparency (FAccT)*. 1395–1417. <https://doi.org/10.1145/3630106.3658979>
- [44] Saffron Huang, Divya Siddarth, Liane Lovitt, Thomas I. Liao, Esin Durmus, Alex Tamkin, and Deep Ganguli. 2024. Collective Constitutional AI: Aligning a Language Model with Public Input. In *Proceedings of the ACM Conference on Fairness, Accountability, and Transparency (FAccT)*. ACM, 1395–1417. <https://doi.org/10.1145/3630106.3658979>
- [45] The Alan Turing Institute. 2024. Rebalancing Innovation: Women, AI and Venture Capital in the UK Second Report. https://www.turing.ac.uk/sites/default/files/2024-03/rebalancing_innovation_second_report_2.pdf. Accessed: 2024-08-14.
- [46] Maurice Jakesch, Zana Buçinca, Saleema Amershi, and Alexandra Olteanu. 2022. How Different Groups Prioritize Ethical Values for Responsible AI. In *Proceedings of the ACM Conference on Fairness, Accountability, and Transparency (FAccT)*. 310–323. <https://doi.org/10.1145/3531146.3533097>
- [47] Sepandar D. Kamvar and Jonathan Harris. 2011. We Feel Fine and Searching the Emotional Web. In *Proceedings of the ACM International Conference on Web Search and Data Mining (WSDM)*. ACM, 117–126. <https://doi.org/10.1145/1935826.1935854>
- [48] Shivani Kapania, Oliver Siy, Gabe Clapper, Azhagu Meena Sp, and Nithya Sambasivan. 2022. "Because AI Is 100% Right and Safe": User Attitudes and Sources of AI Authority in India. In *Proceedings of the ACM Conference on Human Factors in Computing Systems (CHI)*. 1–18. <https://doi.org/10.1145/3491102.3517533>
- [49] Naveena Karusala, Sohini Upadhyay, Rajesh Veeraraghavan, and Krzysztof Z. Gajos. 2024. Understanding Contestability on the Margins: Implications for the Design of Algorithmic Decision-making in Public Services. In *Proceedings of the ACM Conference on Human Factors in Computing Systems (CHI)*. ACM, Article 478, 16 pages. <https://doi.org/10.1145/3613904.3641898>
- [50] Anna Kawakami, Amanda Coston, Haiyi Zhu, Hoda Heidari, and Kenneth Holstein. 2024. The Situate AI Guidebook: Co-Designing a Toolkit to Support Multi-Stakeholder, Early-stage Deliberations Around Public Sector AI Proposals. In *Proceedings of the ACM Conference on Human Factors in Computing Systems (CHI)*. ACM, Article 749, 22 pages. <https://doi.org/10.1145/3613904.3642849>
- [51] Patrick Gage Kelley, Yongwei Yang, Courtney Heldreth, Christopher Moessner, Aaron Sedley, Andreas Kramm, David T. Newman, and Allison Woodruff. 2021. Exciting, Useful, Worrying, Futuristic: Public Perception of Artificial Intelligence in 8 Countries. In *Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society (AIES)*. 627–637. <https://doi.org/10.1145/3461702.3462605>
- [52] Sage Kelly, Sherrie-Anne Kaye, and Oscar Oviedo-Trespalacios. 2023. What Factors Contribute to the Acceptance of Artificial Intelligence? A Systematic Review. *Telematics and Informatics* 77 (2023), 101925. <https://doi.org/10.1016/j.tele.2022.101925>
- [53] Joshua D. Kertzer. 2022. Re-Assessing Elite-Public Gaps in Political Behavior. *American Journal of Political Science* 66, 3 (2022), 539–553. <https://doi.org/10.1111/ajps.12583>
- [54] Kimon Kieslich and Marco Lünich. 2024. Regulating AI-Based Remote Biometric Identification. Investigating the Public Demand for Bans, Audits, and Public Database Registrations. In *Proceedings of the ACM Conference on Fairness, Accountability, and Transparency (FAccT)*. 173–185. <https://doi.org/10.1145/3630106.3658548>
- [55] Taeyun Kim, Maria D. Molina, Minjin (MJ) Rheu, Emily S. Zhan, and Wei Peng. 2023. One AI Does Not Fit All: A Cluster Analysis of the Laypeople's Perception of AI Roles. In *Proceedings of the ACM Conference on Human Factors in Computing Systems (CHI)*. ACM, Article 29, 20 pages. <https://doi.org/10.1145/3544548.3581340>
- [56] Genia Kostka, Léa Steinacker, and Miriam Meckel. 2023. Under Big Brother's Watchful Eye: Cross-Country Attitudes Toward Facial Recognition Technology. *Government Information Quarterly* 40, 1 (2023), 101761. <https://doi.org/10.1016/j.giq.2022.101761>
- [57] Klaus Krippendorff. 2019. *Content Analysis: An Introduction to Its Methodology*. SAGE Publications, Inc. <https://doi.org/10.4135/9781071878781>
- [58] Yara Kyrychenko, Ke Zhou, Edyta Bogucka, and Daniele Quercia. 2025. C3AI: Crafting and Evaluating Constitutions for Constitutional AI. In *Proceedings of the ACM Web Conference 2025*. ACM. <https://doi.org/10.1145/3696410.3714705>
- [59] Alexander Laffer. 2022. Using an Online Narrative Approach to Explore Diverse Participants' Understanding of Emerging Technology: Citizen's Perspectives on Living With Emotional AI. In *Sage Research Methods: Doing Research Online*.
- [60] Markus Langer, Tim Hunsicker, Tina Feldkamp, Cornelius J. König, and Nina Grgić-Hlača. 2022. "Look! It's a Computer Program! It's an Algorithm! It's AI!": Does Terminology Affect Human Perceptions and Evaluations of Algorithmic Decision-Making Systems?. In *Proceedings of the ACM Conference on Human Factors in Computing Systems (CHI)*. ACM, Article 581, 28 pages. <https://doi.org/10.1145/3491102.3517527>

- [61] Junesoo Lee and Jaehyuk Park. 2023. AI As “Another I”: Journey Map of Working With Artificial Intelligence From AI-Phobia to AI-Preparedness. *Organizational Dynamics* 52, 3 (2023), 100994. <https://doi.org/10.1016/j.orgdyn.2023.100994>
- [62] Min Kyung Lee, Daniel Kusbit, Anson Kahng, Ji Tae Kim, Xinran Yuan, Allissa Chan, Daniel See, Ritesh Noothigattu, Siheon Lee, Alexandros Psomas, et al. 2019. WeBuildAI: Participatory Framework for Algorithmic Governance. *Proceedings of the ACM on Human-Computer Interaction (CSCW)* 3 (2019), 1–35. <https://doi.org/10.1145/3359283>
- [63] Yang W. Lee, Diane M. Strong, Beverly K. Kahn, and Richard Y. Wang. 2002. AIMQ: A Methodology for Information Quality Assessment. *Information & Management* 40, 2 (2002), 133–146. [https://doi.org/10.1016/S0378-7206\(02\)00043-5](https://doi.org/10.1016/S0378-7206(02)00043-5)
- [64] Annette Markham. 2012. Fabrication As Ethical Practice. *Information, Communication & Society* 15, 3 (2012), 334–353. <https://doi.org/10.1080/1369118X.2011.641993> arXiv:https://doi.org/10.1080/1369118X.2011.641993
- [65] Maxwell E. McCombs and Donald L. Shaw. 1972. The Agenda-Setting Function of Mass Media. *Public Opinion Quarterly* 36, 2 (01 1972), 176–187. <https://doi.org/10.1086/267990> arXiv:https://academic.oup.com/poq/article-pdf/36/2/176/5242655/36-2-176.pdf
- [66] Nora McDonald, Sarita Schoenebeck, and Andrea Forte. 2019. Reliability and Inter-Rater Reliability in Qualitative Research: Norms and Guidelines for CSCW and HCI Practice. *Proceedings of the ACM on Human-Computer Interaction* 3, CSCW (2019). <https://doi.org/10.1145/3359174>
- [67] Midjourney. 2024. Midjourney Bot. <https://www.midjourney.com>. Accessed: 2024-08-14.
- [68] Matthew Miles and Michael Huberman. 1994. *Qualitative Data Analysis: A Methods Sourcebook*. Sage.
- [69] Kunihiro Miyazaki, Taichi Murayama, Takayuki Uchiba, Jisun An, and Haewoon Kwak. 2024. Public Perception of Generative AI on Twitter: An Empirical Study Based on Occupation and Usage. *EPJ Data Science* 13, 1 (Jan. 2024). <https://doi.org/10.1140/epjds/s13688-023-00445-y>
- [70] Emanuel Moss, Elizabeth Watkins, Ranjit Singh, Madeleine Clare Elish, and Jacob Metcalf. 2021. Assembling Accountability: Algorithmic Impact Assessment for the Public Interest. *SSRN Electronic Journal* (2021). <https://doi.org/10.2139/ssrn.3877437>
- [71] Karim Nader, Paul Toprac, Suzanne Scott, and Samuel Baker. 2024. Public Understanding of Artificial Intelligence Through Entertainment Media. *AI & Society* 39, 2 (2024), 713–726. <https://doi.org/10.1007/s00146-022-01427-w>
- [72] National Institute of Standards and Technology. 2023. AI Risk Management Framework. <https://www.nist.gov/itl/ai-risk-management-framework>. Accessed: 2024-08-20.
- [73] United Nations. 2024. Members of the High-level Advisory Body on Artificial Intelligence. <https://www.un.org/en/ai-advisory-body/members>. Accessed: 2024-08-14.
- [74] Safiya Umoja Noble. 2018. *Algorithms of Oppression: How Search Engines Reinforce Racism*. New York University Press.
- [75] Heather L. O'Brien, Paul Cairns, and Mark Hall. 2018. A Practical Approach to Measuring User Engagement With the Refined User Engagement Scale (UES) and New UES Short Form. *International Journal of Human-Computer Studies* 112 (2018), 28–39. <https://doi.org/10.1016/j.ijhcs.2018.01.004>
- [76] OECS. 2024. Working Party on Artificial Intelligence Governance (WPAIGO). <https://oecdgroups.oecd.org/Bodies/ShowBodyView.aspx?BodyID=7755&BodyPID=15757&Lang=en&Book=False>. Accessed: 2024-08-14.
- [77] Victor Ojewale, Ryan Steed, Briana Vecchione, Abeba Birhane, and Inioluwa Deborah Raji. 2024. Towards AI Accountability Infrastructure: Gaps and Opportunities in AI Audit Tooling. arXiv:2402.17861 <https://arxiv.org/pdf/2402.17861>
- [78] Occupational Information Network (O*NET). 2024. O*NET OnLine Help: Job Zones. <https://www.onetonline.org/help/online/zones>. Accessed: 2024-08-14.
- [79] Jonas Oppenlaender, Tahir Abbas, and Ujwal Gadiraju. 2024. The State of Pilot Study Reporting in Crowdsourcing: A Reflection on Best Practices and Guidelines. *Proceedings of the ACM on Human-Computer Interaction (CSCW)* 8, Article 184 (apr 2024), 45 pages. <https://doi.org/10.1145/3641023>
- [80] Anna-Marie Orloff, Matthias Fassl, Alexander Ponticello, Florin Martius, Anne Mertens, Katharina Krombholz, and Matthew Smith. 2023. Different Researchers, Different Results? Analyzing the Influence of Researcher Experience and Data Type During Qualitative Analysis of an Interview and Survey Study on Security Advice. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems (Hamburg, Germany) (CHI '23)*. Association for Computing Machinery, New York, NY, USA, Article 864, 21 pages. <https://doi.org/10.1145/3544548.3580766>
- [81] Leila Ouchchy, Allen Coin, and Veljko Dubljević. 2020. AI in the Headlines: The Portrayal of the Ethical Issues of Artificial Intelligence in the Media. *AI & Society* 35, 4 (March 2020), 927–936. <https://doi.org/10.1007/s00146-020-00965-5>
- [82] Robert Plutchik and Henry Kellerman. 2013. *Theories of Emotion*. Vol. 1. Academic Press.
- [83] Prolific. 2014. *Quickly Find Research Participants You Can Trust*. Retrieved September 05, 2024 from <https://www.prolific.com>
- [84] Prolific. 2024. *Prolific's Attention and Comprehension Check Policy*. Prolific. Retrieved September 05, 2024 from <https://researcher-help.prolific.com/en/article/fb63bb>
- [85] Lee Rainie, Cary Funk, Monica Anderson, and Alec Tyson. 2022. How Americans Think About Artificial Intelligence. *Pew Research Center* (2022).
- [86] Bogdana Rakova, Jingying Yang, Henriette Cramer, and Rumman Chowdhury. 2021. Where Responsible AI Meets Reality: Practitioner Perspectives on Enablers for Shifting Organizational Practices. *Proceedings of the ACM on Human-Computer Interaction (CSCW)* 5, Article 7 (2021), 23 pages. <https://doi.org/10.1145/3449081>
- [87] David A Robb, Muneeb I Ahmad, Carlo Tiseo, Simona Aracri, Alistair C McConnell, Vincent Page, Christian Dondrup, Francisco J Chiyah Garcia, Hai-Nguyen Nguyen, Éric Pairet, et al. 2020. Robots in the Danger Zone: Exploring Public Perception Through Engagement. In *Proceedings of the 2020 ACM/IEEE International Conference on Human-Robot Interaction*. 93–102. <https://doi.org/10.1145/3319502.3374789>
- [88] Johnny Saldaña. 2015. *The Coding Manual for Qualitative Researchers*. Sage.
- [89] Philip Saleses, Katja Schechtner, and César A Hidalgo. 2013. The Collaborative Image of the City: Mapping the Inequality of Urban Perception. *PloS one* 8, 7 (2013), e68400. <https://doi.org/10.1371/journal.pone.0068400>
- [90] Laura Sartori and Giulia Bocca. 2023. Minding the Gap(s): Public Perceptions of AI and Socio-Technical Imaginaries. *AI & Society* 38, 2 (2023), 443–458. <https://doi.org/10.1007/s00146-022-01422-1>
- [91] Teresa Scantamburlo, Atia Cortés, Francesca Foffano, Cristian Barrué, Veronica Distefano, Long Pham, and Alessandro Fabris. 2023. Artificial Intelligence Across Europe: A Study on Awareness, Attitude and Trust. *arXiv preprint arXiv:2308.09979* (2023).
- [92] Marietje Schaake. 2024. *The Tech Coup: How to Save Democracy from Silicon Valley*. Princeton University Press.
- [93] Morgan Klaus Scheuerman. 2024. In the Walled Garden: Challenges and Opportunities for Research on the Practices of the AI Tech Industry. In *Proceedings of the ACM Conference on Fairness, Accountability, and Transparency (FAccT)*. ACM, 456–466. <https://doi.org/10.1145/3630106.3658918>
- [94] Gianluca Schiavo, Stefano Businaro, and Massimo Zancanaro. 2024. Comprehension, Apprehension, and Acceptance: Understanding the Influence of Literacy and Anxiety on Acceptance of Artificial Intelligence. *Technology in Society* 77 (2024), 102537. <https://doi.org/10.1016/j.techsoc.2024.102537>
- [95] Daniel S. Schiff. 2024. Framing Contestation and Public Influence on Policymakers: Evidence From US Artificial Intelligence Policy Discourse. *Policy and Society* 43, 3 (04 2024), 255–288. <https://doi.org/10.1093/polso/puae007> arXiv:https://academic.oup.com/policyandsociety/article-pdf/43/3/255/58997867/puae007.pdf
- [96] Ali Akbar Septiandri, Marios Constantinides, and Daniele Quercia. 2024. The Impact of Responsible AI Research on Innovation and Development. In *Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society (AIES)*.
- [97] Taylor Shelton and Thomas Lodato. 2019. Actually Existing Smart Citizens: Expertise and (Non) Participation in the Making of the Smart City. *City* 23, 1 (2019), 35–52.
- [98] Katie Shilton. 2013. Values Levers: Building Ethics Into Design. *Science, Technology, & Human Values* 38, 3 (2013), 374–397. <https://doi.org/10.1177/0162243912436985>
- [99] Mona Sloane, Emanuel Moss, Olaitan Awomolo, and Laura Forlano. 2022. Participation Is Not a Design Fix for Machine Learning. In *Proceedings of the ACM Conference on Equity and Access in Algorithms, Mechanisms, and Optimization (EAAMO)*. 1–6. <https://doi.org/10.1145/3551624.3555285>
- [100] Christopher Small, Michael Björkegren, Timo Erkkilä, Lynette Shaw, and Colin Megill. 2021. Polis: Scaling Deliberation by Mapping High Dimensional Opinion Spaces. *Reverca: revista de pensament i anàlisi* 26, 2 (2021). <https://doi.org/10.6035/reverca.5516>
- [101] Kaitao Song, Xu Tan, Tao Qin, Jianfeng Lu, and Tie-Yan Liu. 2020. MPNET: Masked and Permuted Pre-training for Language Understanding. *Advances in Neural Information Processing Systems (NeurIPS)* 33 (2020), 16857–16867.
- [102] Isabella Struckman and Sofie Kupiec. 2023. *Why They're Worried: Examining Experts' Motivations for Signing the "Pause Letter"*. <https://arxiv.org/pdf/2306.00891>
- [103] Mohammad Tahaei, Marios Constantinides, Daniele Quercia, Sean Kennedy, Michael Muller, Simone Stumpf, Q. Vera Liao, Ricardo Baeza-Yates, Lora Aroyo, Jess Holbrook, Ewa Luger, Michael Madaio, Ilana Golbin Blumenfeld, Maria De-Arteaga, Jessica Vitak, and Alexandra Olteanu. 2023. Human-Centered Responsible Artificial Intelligence: Current & Future Trends. In *Extended Abstracts of the Conference on Human Factors in Computing Systems (CHI)*. ACM, Article 515, 4 pages. <https://doi.org/10.1145/3544549.3583178>
- [104] Mohammad Tahaei, Daricia Wilkinson, Alisa Frik, Michael Muller, Ruba Abu-Salma, and Lauren Wilcox. 2024. Surveys Considered Harmful? Reflecting on the Use of Surveys in AI Research, Development, and Governance. *Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society (AIES)* 7 (2024), 1416–1433.

- [105] Ningjing Tang, Jiayin Zhi, Tzu-Sheng Kuo, Calla Kainaroi, Jeremy J Northup, Kenneth Holstein, Haiyi Zhu, Hoda Heidari, and Hong Shen. 2024. AI Failure Cards: Understanding and Supporting Grassroots Efforts to Mitigate AI Failures in Homeless Services. In *The 2024 ACM Conference on Fairness, Accountability, and Transparency*. 713–732.
- [106] The Computational Democracy Project. 2018. *Polis. Input Crowd, Output Meaning*. <https://pol.is/home>
- [107] The Ministry of Digital Affairs of Taiwan. 2024. *Utilizing AI to Enhance Information Integrity: Citizens' Deliberative Assembly*. <https://moda.gov.tw/en/major-policies/alignment-assemblies/2024-deliberative-assembly/1521>
- [108] Time. 2024. TIME 100AI. <https://time.com/collection/time100-ai/>. Accessed: 2024-08-14.
- [109] Unesco. 2024. Members of the High-level Advisory Body on Artificial Intelligence. <https://www.un.org/en/ai-advisory-body/members>. Accessed: 2024-08-14.
- [110] U.S. Census Bureau. 2021. *Race and Ethnicity in the United States: 2010 Census and 2020 Census*. <https://www.census.gov/library/visualizations/interactive/race-and-ethnicity-in-the-united-state-2010-and-2020-census.html>
- [111] U.S. Census Bureau. 2022. *Population by Age and Sex. Annual Social and Economic Supplement*. <https://www.census.gov/library/visualizations/interactive/race-and-ethnicity-in-the-united-state-2010-and-2020-census.html>
- [112] Rama Adithya Varanasi and Nitesh Goyal. 2023. "It Is Currently Hodgepodge": Examining AI/ML Practitioners' Challenges During Co-Production of Responsible AI Values. In *Proceedings of the ACM Conference on Human Factors in Computing Systems (CHI)*. 1–17. <https://doi.org/10.1145/3544548.3580903>
- [113] David Watson and Auke Tellegen. 1985. Toward a Consensual Structure of Mood. *Psychological Bulletin* 98, 2 (1985), 219. <https://doi.org/10.1037//0033-2909.98.2.219>
- [114] Cedric Deslandes Whitney, Teresa Naval, Elizabeth Quepons, Simrandeep Singh, Steven R. Rick, and Lilly Irani. 2021. HCI Tactics for Politics From Below: Meeting the Challenges of Smart Cities. In *Proceedings of the ACM Conference on Human Factors in Computing Systems (CHI)*. 1–15. <https://doi.org/10.1145/3411764.3445314>
- [115] Simon Nicholas Williams and Kimberly Dienes. 2023. Public Views During the COVID Pandemic (PVCovid). *PsyArXiv* (2023).
- [116] Women in AI Ethics. 2025. *100 Brilliant Women in AI Ethics list – 2025*. <https://womeninaethics.org/the-list/>
- [117] Nur Yildirim, Changhoon Oh, Deniz Sayar, Kayla Brand, Supritha Challa, Violet Turri, Nina Crosby Walton, Anna Elise Wong, Jodi Forlizzi, James McCann, and John Zimmerman. 2023. Creating Design Resources to Scaffold the Ideation of AI Concepts. In *Proceedings of the 2023 ACM Designing Interactive Systems Conference (DIS)*. Association for Computing Machinery, 2326–2346. <https://doi.org/10.1145/3563657.3596058>
- [118] Baobao Zhang and Allan Dafoe. 2020. U.S. Public Opinion on the Governance of Artificial Intelligence. In *Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society (AI/ES)*. ACM, 187–193. <https://doi.org/10.1145/3375627.3375827>

Appendix

A Methods for Collecting AI Views from the Public

A.1 Summary of Interviews with Designers

To gather insights on how to effectively visualize the platform’s requirements, we collaborated with two experienced designers through semi-structured interviews. The designers stressed that understanding key values, which represent the personal gains users receive after completing the survey, is crucial for inspiring participants and increasing the quality of responses, particularly in gamified survey contexts [1]. Specifically, they referred to four values: *self-discovery value*, where users find out something about themselves; *transcendent value*, where users feel a sense of contribution to broader societal discussions; *knowledge value*, where users leave with new information gained through the visualizations; and *narrative value*, where users see their responses integrated into a larger, ongoing story, giving them a sense of connection and purpose [1]. With these four key values as our guiding principles, we built the first prototype of the platform (Figure 8)—a visual story that uses the metaphor of filter bubbles and balances data collection from participants with the presentation of others’ opinions on AI.

To meet R1 *Balanced representation of views*, we incorporated the self-discovery value by presenting people’s opinions as interactive bubbles, color-coded based on the most dominant emotion felt about AI. To obtain the data about ordinary people’s opinion, we first searched for news related to AI using an unofficial Google News API ³. We then used a hard-matching process to identify whether these articles expressed any of the emotions from Plutchik’s wheel of emotions [82]. Participants can create their own bubble by describing their feelings about AI and comparing it with others, providing a balanced view of the data. To meet R2 *Emotional appeal*, we integrated the transcendent value by signaling to participants in the introduction that their honest opinions about AI will help break down opinion bubbles and contribute to developing stronger safeguards to limit AI risks, fostering a deeper emotional connection to the visualization. To meet R3 *Engaging participation*, we implemented the knowledge value by including links to articles where individuals expressed emotions about AI, providing educational content and keeping participants engaged. To meet R4 *Facilitated deliberation*, we integrated the narrative value by adding popups with portraits and bios to the bubbles, helping participants empathize with different perspectives and reflect on how individual stories contribute to the broader societal narrative about AI.

A.2 Questionnaire Used During the Co-Design Process

The study was divided into four parts. In the first part, participants were asked to rate, on a scale from 1 to 5 (ranging from “Far below average” to “Far above average”), how skilled or knowledgeable they consider themselves to be in technology and AI compared to most people. In the second part, participants were asked to rate their interest, on a scale from 1 to 5 (ranging from “Not interested” to “Extremely interested”), in the following areas: participating in

discussions about AI development, learning about others’ opinions on AI, understanding how their opinions on AI align with the general public, and seeing how their opinions on AI compare to those of influencers. In the third part, participants were asked two open-ended questions: how they would like to contribute to AI development or oversight and what barriers prevent their involvement in AI discussions or decision-making. In the final part, participants were shown the current iteration of the visualization and informed that it is intended to gather public opinions on AI and is still in its early development stages.

A.3 Design Iterations for the Visualization Platform

In the first iteration, we identified two primary concerns. First concern was related to R4 *Facilitated deliberation*. Our participants showed equal interest in comparing how their opinions align with the general public and with influencers, both receiving an average rating of 3.8 (“Very interested”). As phrased by F10, “*it would definitely encourage me to participate in discussions if I am able to see expert’s responses and read what their hopes and fears may be*”. The second concern was related to the R2 *Emotional appeal* and the design of the platform. F1, supported by F7, mentioned that “*The colored circles feel too childish for this topic and need to be better organized*”. To address these issues, we implemented two key changes (Figure 9): first, we decided to include the opinions of AI influencers in the platform. Second, we refined the visual design of the platform by replacing color-coded bubbles with clustered black-and-white graphics of people expressing their emotions towards AI (Figure 9).

In the second iteration, we identified a new design requirement based on user feedback regarding the connection between emotions, underlying knowledge of AI and ability to act upon them. F12 observed that, “*Many may have a misconception of exactly what AI will initially afford the user of this adaptable technology when used properly. Seeing the grouping of others who share the same doubt and concern of implementing AI into their lives is scary for the uneducated individual of AI*.” F14 suggested, “*I think that more open-ended discussions would be helpful to move forward on how we improve AI and how we can work together to bring the extreme ends of these views closer together. Although people may be fearful, we can work on providing reasons why it is beneficial, whereas on the other end, those who are hopeful may need to see the many flaws that AI has as well as the problems with no proper regulations*.” In response, we expanded our design focus from merely appealing to emotions (R2 *Emotional Appeal*) to ensuring a broader appeal (R2 *Broad Appeal*) that resonates with people of varying levels of knowledge and emotions regarding AI.

In the second iteration, participants also reported two concerns related to the look and feel of the platform (R2 *Broad Appeal*), and the categorization of people based on their answers (R3 *Engaging participation*). First, the interface was criticized as “*a bit too elementary in appearance. You could always use more realistic pictures and make some elements stand out more, just to help separate information*” (F2). In response, we decided to limit the color palette of the platform and replaced the black-and-white graphics with realistic-looking portraits of people. Specifically, we employed a Hope vs. Fear framing, using a stark black, white, and grey color palette with

³<https://serpapi.com/google-news-api>

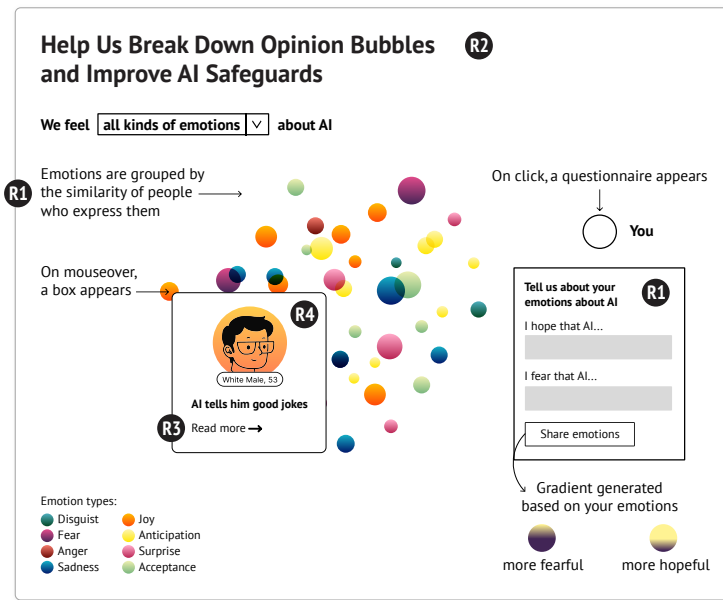


Figure 8: The first version of our platform for collecting public views on AI encourages participants to share their honest opinions to help break down opinion bubbles (R2). It features interactive bubbles representing people’s views (R1), with pop-ups (R4) displaying portraits, bios, and article links (R3). Participants can create their own bubbles to describe their feelings about AI and compare them with others, providing a balanced view of the data (R1).

high contrast to emphasize the emotional dichotomy and simplify the deliberation process. The portraits of influencers were collected from the public domain and restyled using Midjourney [67]. To ensure transparency and compliance with regulations like the EU AI Act [28], we disclosed the use of AI-generated content in an appropriate manner. This disclosure was embedded within the visualization, styled as a gallery image caption pop-up, ensuring that participants were informed without disrupting their experience. Second, one participant (F12) expressed concern that “a person could not be accurately categorized as hopeful or fearful about AI based on one question. I would like to see a more comprehensive survey to help align the results better.” To address this, we explored the use of provocative design elements such as investment game to collect more nuanced data in a way that doesn’t feel like a traditional survey but still allows for effective categorization.

By the third and final iteration, we focused on further improving *R2 Broad Appeal*. We enhanced the visual hierarchy by adjusting font sizes, repositioning interactive elements, and providing clearer annotations on how to navigate the platform. We also added more explanations of each person’s role in response to two comments similar to P30, “I didn’t recognize any of the names I was put with – if I did recognize them I think it would have been more exciting.”

These three iterations allowed us to revisit our design requirements and refine the platform based on user preferences. The process ultimately led to a finalized design where no further refinements were deemed necessary (Figure 3).

A.4 Evaluating the Platform for Collecting Views from the Public about AI

A.4.1 Pilot evaluation. We conducted a pilot study with 30 new participants, distinct from those in the earlier co-design sessions, to evaluate the final platform for collecting public views on AI and to simulate the upcoming large-scale study. By doing so, we aimed to identify any improvements needed before the full-scale launch and to determine appropriate monetary rewards for participants, following best practices for crowdsourcing research [79, 104].

The participants of the pilot study reflected the demographics of the U.S. population [110, 111] in terms of sex (15 males and 15 females) and ethnicity (18 White, 4 Black, 2 Asian, 6 Mixed or Other), with ages ranging from 19 to 71 years old. To evaluate the platform, we outlined five key statements on how effectively it: presented a balanced representation of views (R1), achieved broad appeal (R2), engaged individuals in data collection (R3), facilitated deliberation (R4), and whether similar visualizations would be preferred over traditional surveys for future AI opinion studies. We sourced these five statements from the dimensions of the Information Quality Assessment questionnaire [63] and one from the User Engagement Scale [75]. Each statement was rated on a 1-7 Likert scale, ranging from “Strongly disagree” to “Strongly agree”. Additionally, participants were asked to justify their responses, and these justifications were thematically analyzed to complement our quantitative findings. The thematic analysis allowed us to identify key patterns in participant feedback, which we further explore in the next paragraph through direct quotes (marked with F). The pilot study was approximately 20-60 minutes long and participants were paid on average about \$12 (USD) per hour.

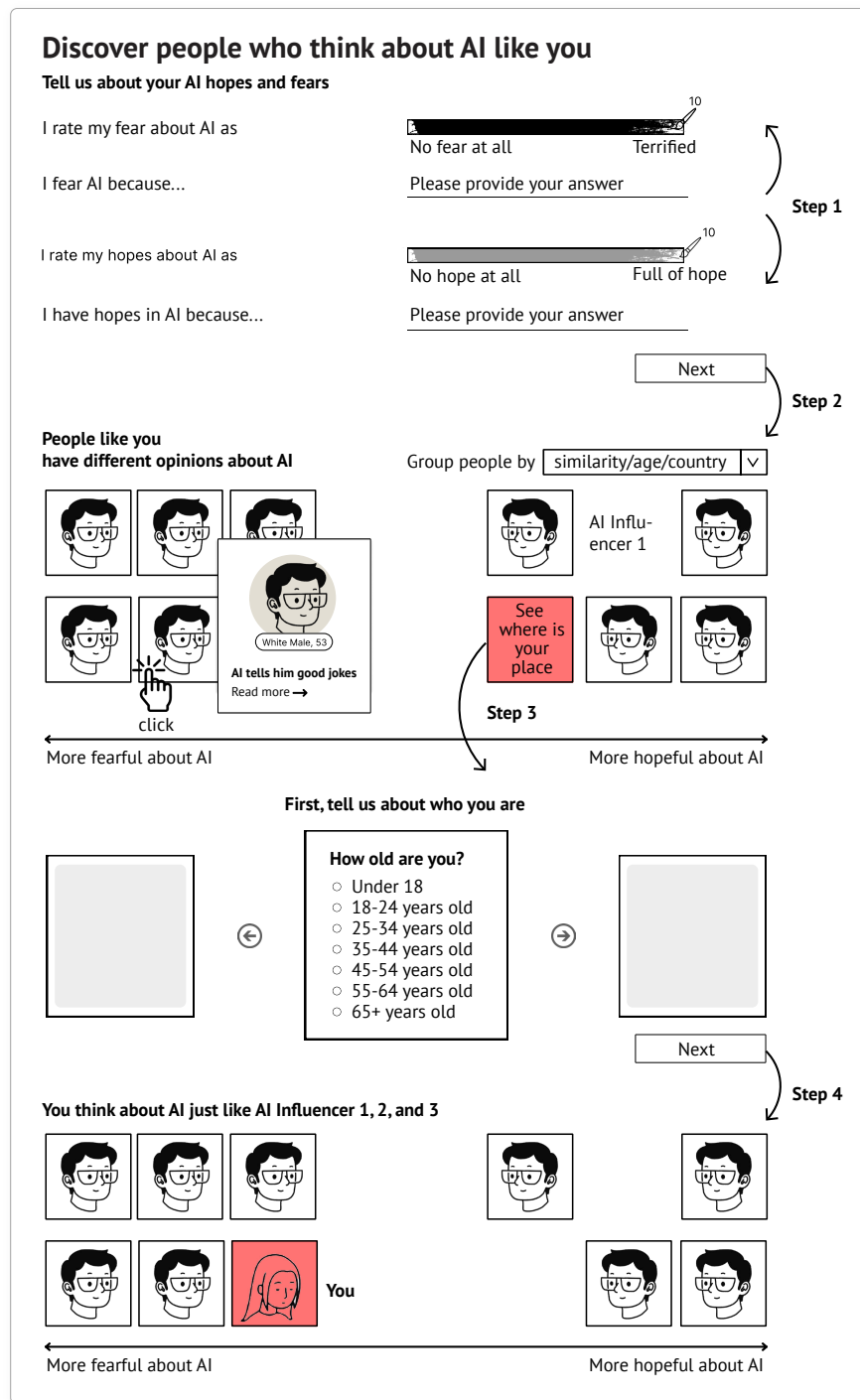


Figure 9: The second version of our platform helps participants find people who share similar opinions about AI. It begins with an interactive questionnaire where participants rate and describe their own fears and hopes about AI. After completing the questionnaire, participants see a gallery of portraits of AI influencers, with their opinions and article links available in pop-ups. Participants can then place themselves among these influencers by answering a short demographics questionnaire, which compares their own fears, hopes, and demographics with those of the influencers.

Table 3: Results from evaluating the platform for collecting public views on AI during the pilot study with 30 participants. The requirements were rated using a 7-point Likert scale, ranging from 0 (“strongly disagree”) to 7 (“strongly agree”).

Design requirement	Statement	Rating
R1 Balanced representation of views	The visualization presents a balanced view [63]	4.94 / 7
R2 Broad appeal	The visualization is easy to understand [63]	5.88 / 7
	The visualization is relevant to those without technical background [63]	5.54 / 7
R3 Engaging participation	The visualization is engaging [75]	5.66 / 7
R4 Facilitated deliberation	The visualization is useful for thinking about AI [63]	5.52 / 7

Table 4: Results from evaluating the platform for collecting public views on AI during the large-scale study with 330 participants. The requirements were rated using a 7-point Likert scale, ranging from 0 (“strongly disagree”) to 7 (“strongly agree”).

Design requirement	Statement	Rating
R1 Balanced representation of views	The visualization presents a balanced view [63]	5.13 / 7
R2 Broad appeal	The visualization is easy to understand [63]	5.48 / 7
	The visualization is relevant to those without technical background [63]	5.37 / 7
R3 Engaging participation	The visualization is engaging [75]	5.45 / 7
R4 Facilitated deliberation	The visualization is useful for thinking about AI [63]	5.20 / 7

Our proposed platform received high average ratings for each design requirement (Table 3). R1 averaged 4.94 (closer to “somewhat agree”), while R2–R4 ranged from 5.52 to 5.88 (closer to “agree”). 60% of participants stated they would generally prefer the platform over traditional surveys, with an additional 13% mentioning they would choose it in certain situations. Participants favored the platform for its interactivity and engagement over traditional surveys: “*its format was much more comfortable to navigate and interesting to engage with. I would like that approach to become more commonplace*” (F15). The clarity and ease of understanding were also important points: “[I] prefer the visualization instead of the traditional because it is easy to understand” (F2). Therefore, we concluded that our platform sufficiently supported our design requirements and could be used in a large-scale user study (Table 4).

However, to make the study even more effective, participants suggested improvements in two main areas: navigation through the platform and the phrasing of the questions. To address navigation issues, we added arrow annotations to guide participants on how to interact with the gallery of AI influencers’ views and improved the visibility of buttons to make it easier to move between sections. For question phrasing, we clarified the wording and improved the order of the questions. For example, instead of asking “*Please explain your answer*”, we used “*Please help us understand the reasons behind your answer*”, encouraging participants to share their thought process and foster their sense of contribution. Finally, after reviewing the quality of responses and the time spent during the pilot study, we decided to update the survey code to disable pasting from external sources and adjust the monetary rewards for participants over 70 years old, who spent nearly twice as much time (an average of 38 minutes) on the platform compared to younger participants, by including bonus payments.

A.4.2 Large-scale study evaluation. After deploying the platform in the large-scale study, R2: *Broad appeal* and R3: *Engaging participation* have average scores close to *agree*, while R1: *Balanced representation of views* and R4: *Facilitated deliberation* are closer to

somewhat agree (Table 4). Regarding R1, participants tended to be positive, and the main concerns were related to the presence of only well-known individuals: “[...] *it is important to include the views of people from all levels of society*.” However, most participants agreed that the representation of fears and hopes was mostly impartial: “*The visualization impartially presents both sides of the argument [...]*.” The majority of participants judged the visualization and interactions intuitive, having a *broad appeal* (R2) even for people with low technical skills: “*The visualization is intuitive and appeals to people with a little knowledge about AI (like me)*.” Some participants expressed their inability to judge how appealing the visualization would be to others. The interactive aspects of the visualization, especially clicking on the portraits, was one of the most important components for *engaging participation* (R3): “*I enjoyed reading all the different takes, and that you could click through for a more in-depth piece on a person to learn more*”. The visual appeal and design quality were also highlighted: “*Its an easy to use visualization and has high-quality design and pictures. I really like it*.” A few users expressed concerns about the volume of information presented: “*While the visual was nice, there was too many [portraits] [...]*”. Finally, participants highlighted that the quick contact with a large number and diversity of ideas helped expand their understanding of AI and ultimately *facilitated deliberation* (R4): “*The visualization brought of ideas I had not thought of before forming my own opinion about some of the hazards of AI in many aspects of life*.” Overlooked topics were also brought to the attention of some participants: “[...] *it gave me a variety of opinions to consider. It expanded my scope of thought, such as to how it might hurt people who are from Black, Hispanic, etc. and other marginalized communities*.”

When asked if similar visualizations would be preferred over traditional surveys for future AI opinion studies, 62% of participants stated they would generally prefer our platform, with an additional 15% indicating they would choose it in specific situations. However, 23% expressed a preference for traditional surveys over the platform.

We found 45 opinions related to being compared to influencers and identified distinct reasons behind positive, neutral, and negative opinions. Among the 31 positive opinions, participants most frequently appreciated seeing faces ($n = 7$) and views of others ($n = 7$), which made the visualization engaging ($n = 6$) and interesting ($n = 6$). They also highlighted that it felt real ($n = 5$), was easy to understand ($n = 3$), helpful ($n = 2$), and occasionally fun ($n = 1$), emphasizing its role in making abstract concepts about AI more personal and relatable. The neutral opinions ($n = 7$) acknowledged the visualization's presence but described its impact as limited, often noting that opinions were noticed but didn't evoke strong feelings ($n = 2$), while it provided perspectives ($n = 1$) but didn't deeply resonate. On the other hand, the 7 negative opinions emphasized issues with complexity or time consumption ($n = 2$), a lack of trust in the relevance or authenticity of the information ($n = 2$), and found the faces unhelpful ($n = 1$) or disturbing ($n = 1$).

We found 40 opinions about the reinforcement of beliefs and identified distinct reasons behind positive, neutral, and negative perspectives. Among the 27 positive opinions, participants most frequently appreciated how the visualization broadened perspectives ($n = 6$) and helped them form or refine their own opinions ($n = 5$). Many valued its ability to showcase a spectrum of views ($n = 4$) and provide a balance between pros and cons ($n = 3$). Others praised it as engaging and thought-provoking ($n = 3$), highlighting its educational value ($n = 3$), and noted that the diverse opinions encouraged reflection ($n = 3$). The neutral opinions ($n = 8$) often acknowledged the mix of opinions but indicated that it didn't lead to significant changes in beliefs ($n = 3$). Some described it as informative without being impactful ($n = 2$), while others felt it reinforced existing beliefs rather than introducing new perspectives ($n = 2$). One participant noted that the visualization was overwhelming but still recognized its range of views ($n = 1$). In contrast, the 5 negative opinions emphasized skepticism about the relevance of the shared opinions ($n = 3$) and described the content as overwhelming or too complex to engage with meaningfully ($n = 2$).

B Comparing the Views of the Public versus AI Influencers

B.1 Setup for the Web-Based Survey for a Large-Scale User Study To Collect Participants' Views on AI and Expose Them to the Opinions of AI Influencers

In our large-scale study on Prolific [83], we used a custom website and structured it into four steps (Figure 11, S1-S4). First, we provided participants with a brief introduction to the study and had them complete a warm-up task where they described Artificial Intelligence to a friend (Figure 11, S1). Next, participants interacted with our platform (Figure 11, S2). In the third step, they rated how well the platform met the four design requirements and provided feedback (Figure 11, S3). Finally, we redirected them to the Prolific confirmation page (Figure 11, S4).

B.2 Calculating the Misalignment Score

Different subgroups had varying numbers of AI influencers, each expressing a different number of fears and hopes. Figure 10 shows how the number of sentences in the unordered list impacted the misalignment score calculation for lists of 5, 10, and 30 sentences. As more sentences were included, the alignment scores tend to converge towards 0.5, which corresponds to a random ranking. To address this progressively skewed distribution, we normalized the misalignment scores using min-max normalization, adjusting for the minimum and maximum possible misalignment scores based on the specific number of sentences.

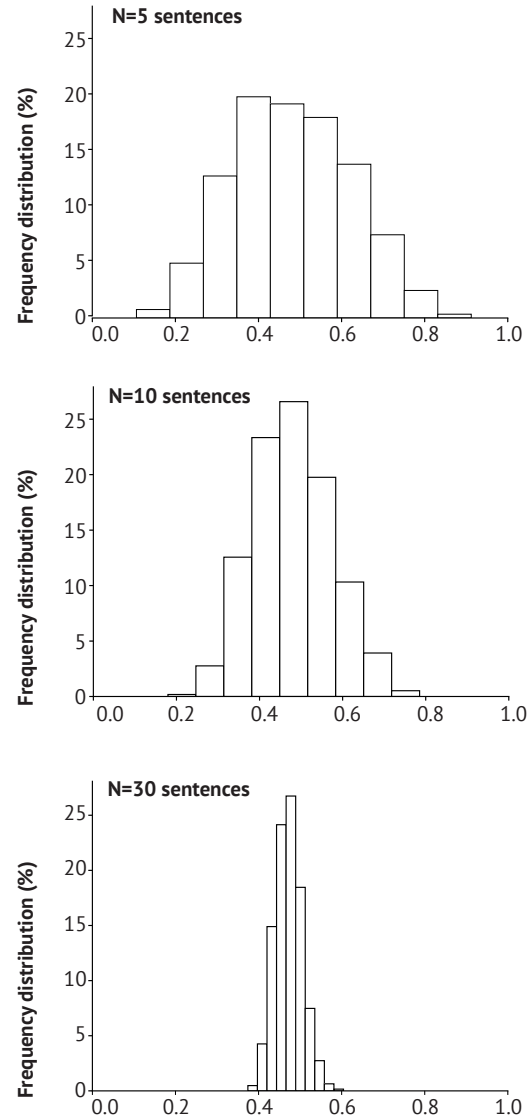


Figure 10: Impact of the number of sentences on the distribution of the misalignment score.

S1 Introduction to the study and a warm-up task

WE ARE EXPLORING THE DIFFERENT FEARS AND HOPES PEOPLE HAVE ABOUT AI

In this survey, you will be asked to:

- Task 1: Share your fears and hopes about AI.
- Task 2: Explore an interactive data visualization that highlights the fears and hopes about AI of famous people.
- Task 3: Provide a brief evaluation of the visualization.

All questions in this survey are mandatory. Read each of them thoroughly and provide your responses to the best of your ability. Copying responses from external sources is not permitted and disabled.

Please enter your Prolific ID

(14 characters)

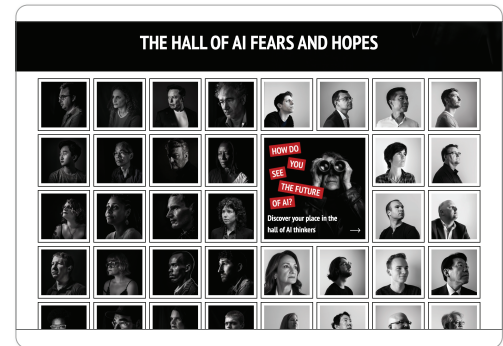
How would you describe Artificial Intelligence to a friend?

AI description

(2500 (Min. 70 characters))

[Start survey](#)

S2 Interaction with the platform



S3 Rating the platform for design requirements and providing feedback

EVALUATE THE HALL OF AI FEARS AND HOPES

BALANCED REPRESENTATION OF OPINIONS

The visualization presents a balanced view of opinions

☐ Strongly disagree ☐ Disagree ☐ Somewhat disagree ☐ Neutral ☐ Somewhat agree ☐ Agree ☐ Strongly agree

Please help us understand the reasons behind your answer

(250 (Min. 70 characters))

BROAD APPEAL

The visualization is easy to understand

☐ Strongly disagree ☐ Disagree ☐ Somewhat disagree ☐ Neutral ☐ Somewhat agree ☐ Agree ☐ Strongly agree

The visualization is relevant to those without technical background

☐ Strongly disagree ☐ Disagree ☐ Somewhat disagree ☐ Neutral ☐ Somewhat agree ☐ Agree ☐ Strongly agree

Please help us understand the reasons behind your answers

(250 (Min. 70 characters))

ENGAGING

The visualization is engaging

☐ Strongly disagree ☐ Disagree ☐ Somewhat disagree ☐ Neutral ☐ Somewhat agree ☐ Agree ☐ Strongly agree

Please help us understand the reasons behind your answer

(250 (Min. 70 characters))

FACILITATED DELIBERATION

The visualization is useful for thinking about AI

☐ Strongly disagree ☐ Disagree ☐ Somewhat disagree ☐ Neutral ☐ Somewhat agree ☐ Agree ☐ Strongly agree

Please help us understand the reasons behind your answer

(250 (Min. 70 characters))

Would you prefer using similar visualizations instead of traditional surveys for future studies on AI opinions? Why or why not?

(250 (Min. 70 characters))

Please enter your email if you would like to receive future updates about the project (optional)

[Submit my responses](#)

S4 Confirming receipt of responses and redirecting to the Prolific confirmation page

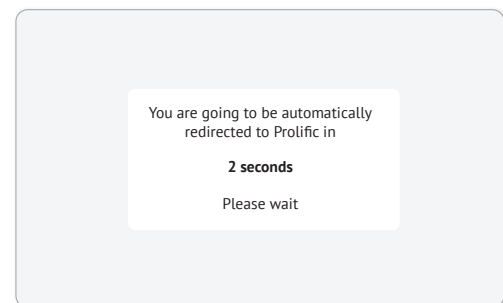


Figure 11: The large-scale study on Prolific consisted of four steps. In the first step (S1), participants received a brief introduction to the study and completed a warm-up task in which they described Artificial Intelligence to a friend. In the second step (S2), participants interacted with our platform. In the third step (S3), participants rated how well the platform met the four design requirements and provided feedback. Afterwards, they were redirected to the Prolific confirmation page (S4).

B.3 Themes of Fears and Hopes Among the Members of the Public

Table 5 presents the results of our thematic analysis of the fears and hopes shared by 330 participants, representative of the U.S. population in terms of age, sex, ethnicity, and political views. The table reports the identified themes along with their frequency of occurrence. Each theme includes fears and hopes from at least ten participants.

Table 5: Occurrences of fears and hopes related to AI among the public.

Fear	Occurrences
Reduced employment opportunities	144
Irresponsible use of AI (e.g., among countries or businesses)	66
Increased spread of misinformation	55
AI becoming uncontrollable	49
Decline in inventive thinking	42
Lack of transparency in AI decision-making	27
Inherent bias in AI (e.g., ethnic, gender bias)	26
Overreliance on AI	20
AI gaining consciousness	19
Loss of personal privacy	17
Increased global conflicts	17
Reduced social connectedness	11
Hope	Occurrences
Advance in social services (e.g., medical)	74
Ability to handle repetitive tasks	65
Increased access to information	50
Increased innovation (e.g., research)	40
Increased employment opportunities	22
Improved creative thinking	18
Improved social connectedness (e.g., increased inclusivity)	15