AI & SOCIETY
https://doi.org/10.1007/s00146-026-03023-8

OPEN FORUM

Charting the AI perception gap: divergent views on risk, benefit, and value between experts and the public challenge the societal acceptance of AI

Philipp Brauner[1] · **Felix Glawe**[1] · **Gian Luca Liehner**[1] · **Luisa Vervier**[1] · **Martina Ziefle**[1]

Received: 11 November 2025 / Accepted: 31 March 2026
© The Author(s) 2026, modified publication 2026

Abstract
Artificial intelligence (AI) is reshaping society, raising questions about trust, risks, and the asymmetries between public and academic perspectives. We examine how the German public (N = 1,100), comprising individuals who interact with or are affected by AI, and academic AI experts (N = 119, mainly from Germany), who contribute to research, educate practitioners, and inform policymaking, construct mental models of AI's capabilities and impacts across 71 scenarios. These scenarios span diverse domains (including sustainability, healthcare, employment, inequality, art, and warfare) and were evaluated across four dimensions using the psychometric model: likelihood, perceived risk, perceived benefit, and overall value. Across scenarios, academic experts anticipated higher probabilities of occurrence, perceived lower risks, and reported greater benefits than the public, while also expressing more positive overall evaluations of AI. Beyond differences in absolute assessments, the two groups exhibited systematically different evaluative patterns: experts' judgments leaned more heavily toward a benefit-dominant calculus ($\beta_{benefit} = +0.623$; $\beta_{risk} = -0.195$), whereas the public integrated risks more strongly into their evaluations ($\beta_{benefit} = +0.703$; $\beta_{risk} = -0.361$). Visual mappings indicate convergent domains (e.g., medical diagnoses and criminal use) and tension points (e.g., justice and political decision-making) that may warrant targeted communication or policy attention. While this study does not assess AI systems or design practices directly, the observed divergence in mental models suggests that the research, implementation, and use of AI may inadvertently neglect the risk-related priorities of the public. Such biases in research and implementation may yield "procrustean AI"—systems insufficiently aligned with the needs of the affected public (akin to the Bed of Procrustes). To counter this, we address the socio-technical challenges of expert-centric governance and argue for participatory design practices that bridge this alignment gap.

Keywords Public perception · AI governance · Societal integration of technology · AI alignment · Expert–layperson comparison · Science and technology studies

1 Introduction

Artificial intelligence (AI) dates back decades (McCarthy et al. 2006; Hopfield 1982; Rumelhart et al. 1986). Recent advances in algorithms, computing power, and training data have led to transformative breakthroughs and increased funding (Deng et al. 2009; Lecun et al. 2015; Statista 2022). AI is increasingly integrated across sectors such as education (Chen et al. 2020), healthcare (Amunts et al. 2023), journalism (Diakopoulos 2019), farming (Holzinger et al. 2024), innovation management (Bouschery et al. 2023), and production (Brauner et al. 2022). Beyond technical applications, AI could also reduce loneliness and enhance well-being: chatbots and social AI actors provide companionship (Ventura et al. 2025; Malfacini 2025), while embodied robots support healthcare, education, and social integration (Ahmed et al. 2024), though these applications carry risks of emotional dependency, erosion of human relationships, and other ethical consequences if not carefully designed and governed. Consequently, AI offers benefits but also raises significant societal concerns, including privacy infringements (Lim and Shim 2022), job displacement (Acemoglu and Restrepo 2017), algorithmic bias (Sadek et al. 2024; Brauner et al. 2019), and broader ethical challenges (Awad

✉ Philipp Brauner
brauner@comm.rwth-aachen.de

[1] Communication Science, RWTH Aachen University, Aachen, Germany

Published online: 29 April 2026

et al. 2018). Overall, expectations about AI remain divided: While some view it as a revolutionary tool that can improve our lives (Brynjolfsson and McAfee 2014; Makridakis 2017), others emphasize its potential risks (Cath 2018; Bostrom 2003) and note that large language models have already begun shaping human communication (Geng et al. 2025).

Algorithms, AI, and AI-based systems are products of human design and are shaped by developers' values, assumptions, and biases. Hence, their output and the decisions they inform are not value-neutral and may lead to unintended or harmful consequences, such as perpetuating inequality or reinforcing existing biases (Friedman and Nissenbaum 1996; Nissenbaum 2001; Sadek et al. 2024). The "*AI alignment problem*" addresses these concerns by seeking to ensure that AI systems act in line with human values (Hristova et al. 2024; Gabriel 2020). This alignment involves designing systems that understand, interpret, and follow human-aligned objectives, even as AI becomes more advanced and autonomous.

Rather than addressing properties of AI systems or the "AI alignment problem" directly, this study examines how stakeholders interpret potential AI futures through distinct evaluative frameworks and mental models (Johnson-Laird 2010). Specifically, we analyze divergences in perceived likelihood, risk, benefit, and overall value between academic AI experts and members of the general public (hereafter: the public). AI experts are a particularly consequential group, as they educate practitioners, conduct foundational research, and shape governance, safety, and design paradigms that influence AI development and alignment. Accordingly, our focus is not whether AI systems can be aligned with human values in an engineering sense, but if experts and the public form diverging views about AI.

In this article, we surveyed both academic AI experts and the public regarding 71 future AI scenarios, each presenting a different capability or impact as a vignette. As individual perceptions of risk and benefit shape attitudes, usage intentions, and actual behavior (Witte and Allen 2000; Hoffmann et al. 2015; Huang et al. 2020), we measured the expected likelihood of occurrence, perceived risks and benefits, and the overall value (or sentiment) associated with these topics. By analyzing similarities and differences in these responses and visualizing the findings as cognitive maps of AI perception, this study identifies both shared and divergent views. This approach contributes to the development of a research agenda for better, human-centric AI, highlights areas that may require stronger regulation, and suggests that accessible public education on AI could play a role in bridging the identified perception gap (Marx et al. 2022).

2 Related work

This section reviews the relevant literature. It begins with an overview of prior research on the perception of AI, followed by studies comparing how experts and the public perceive technology in general and AI in particular. The section concludes by identifying the research gaps and formulating the research questions that guide the present study.

2.1 Public perception of AI

While AI has existed for decades, the rapid adoption of generative tools like ChatGPT (Hu 2023) has intensified academic focus on how the public perceives AI's implications across diverse domains. These perceptions are not a monolithic reaction to technology but a multifaceted social construction shaped by a tension between utopian narratives of benefit and deep-seated anxieties regarding control. Although the literature identifies various drivers, including media framing, cultural context, and individual literacy, it reveals a landscape of "polarized expectations" where public imagination often diverges from technical reality. Currently, a consolidated understanding of how these factors can be integrated into a holistic model remains absent. This section reviews these influences to highlight a critical research gap: the need for an empirical foundation that systematically maps public sentiment across a broad spectrum of societal domains, moving beyond the fragmented, application-specific insights that currently dominate the field.

Media plays a significant role in shaping public opinion about AI. Fast and Horvitz (2017) conducted an analysis of three decades of AI coverage in The New York Times, observing a rise in public interest after 2009. Coverage has generally been more positive than negative, balancing optimism with concerns over control and ethical issues. Recently, reporting has reflected heightened enthusiasm for AI's potential, particularly for healthcare, mobility, and education. News coverage often emphasizes the benefits of AI while downplaying potential risks, contributing to a perception of AI as superior to human capabilities and fostering the anthropomorphization of technology (Puzanova et al. 2024). Cave et al. (2019) examined AI narratives in the UK and identified four optimistic and four pessimistic themes. These narratives frequently evoke anxiety, with only two highlighting benefits over risks, such as AI's potential to make life easier.

Furthermore, sentiment analysis of WIRED articles reveals an increase in polarized views, with both positive and negative sentiments intensifying over time (Moriniello et al. 2024). Sanguinetti and Palomo (2024) investigated how news outlets portray AI and found that coverage often frames AI as something to be feared, depicting it as an autonomous

and opaque entity beyond human control. Using an AI anxiety index, the study analyzed newspaper headlines before and after the launch of ChatGPT, reporting both increased coverage and heightened negative sentiment.

Perceived risks and benefits play a crucial role in shaping public attitudes toward AI. Surveys indicate that the public views AI as both a risk and an opportunity. Common concerns include privacy violations and cybersecurity threats (Brauner et al. 2023), while perceived benefits are often associated with applications in urban services and disaster management (Schwesig et al. 2023; Yigitcanlar et al. 2022). Both perceived risks and opportunities shape individuals' behavioral intentions to use AI applications and a higher perceived opportunity-risk ratio is associated with greater willingness to adopt AI, though with notable variation depending on the application context (Schwesig et al. 2023).

Neri and Cozman (2019) argue that experts significantly influence public perceptions of AI risks. Their public statements can amplify awareness of certain threats, such as existential risks, which may be rooted more in expert discourse than in actual incidents. Lee et al. (2024) found that individuals with higher education levels, greater political interest, and more knowledge about ChatGPT tend to perceive AI as more risky. This finding challenges the conventional "knowledge deficit" model, suggesting that negative perceptions may stem from a critical mindset that engages with AI technology more cautiously.

Trust in AI varies across contexts, demographic groups, and individual attitudes. For instance, people tend to trust AI more in personal lifestyle applications but remain more skeptical about its use by companies and governments (Yigitcanlar et al. 2022). Willingness to engage with AI is shaped by the perceived balance of risks and benefits, which differs across domains such as healthcare, transportation, and media (Schwesig et al. 2023). In addition to variations in trust, there remains a significant gap in public understanding of AI, which can contribute to irrational fears and misinformed beliefs about control. Promoting AI literacy is therefore essential to enable informed decision-making and support responsible innovation (Ng et al. 2021; Marx et al. 2022).

Public perception of AI also varies significantly based on local context, political ideology, and exposure to science news. For example, people in the United States generally expect more benefits than harms from AI, with a substantial portion supporting regulation to mitigate potential risks (Elsey and Moss 2023). However, existential risks are not a primary concern for most; instead, the public tends to worry more about tangible issues such as job displacement.

Sindermann et al. (2022) examined cross-cultural differences in AI attitudes among Chinese and German participants, linking fear of AI to neuroticism in both groups, while also highlighting cultural variations in AI acceptance and concern. Kelley et al. (2021) surveyed over 10,000 participants across eight countries (Australia, Canada, the United States, South Korea, France, Brazil, India, and Nigeria), finding that respondents in developed nations predominantly expressed worry and futuristic expectations, whereas those in developing countries showed greater enthusiasm for AI's potential. In particular, South Korea emphasized AI's practical usefulness and future applications, though widespread uncertainty about its broader societal impact persisted across all regions. In Taiwan, science news consumption and respect for scientific authority positively influence AI perceptions (Wen et al. 2024). Interestingly, a recent large cross-cultural study building on Hofstede's cultural dimensions found that AI perception is rather shaped by individual differences than by cultural context (Wang 2025).

Public discourse on AI often oscillates between fear and inflated expectations—particularly regarding artificial general intelligence (AGI), which remains largely speculative and fictional at present (Jungherr 2023). A survey by Ipsos (2022) found that the public frequently lacks a nuanced understanding of AI's technical capabilities and limitations. Similarly, Pew Research and Center (2023) reported that only a small percentage of Americans could accurately identify AI in everyday scenarios, highlighting widespread confusion about its scope and functionality. The Alan Turing Institute (2023) likewise observed that public understanding of AI varies considerably depending on education level and context. Common concerns tend to focus on automation and robotics, especially in relation to employment and security. This limited awareness contributes to persistent misconceptions and overly simplistic views of AI's societal impact, ultimately hindering informed public discourse.

In summary, the drivers of public sentiment are well-documented individually as multifaceted and shaped by media narratives, perceived risks and benefits, levels of trust, and cultural and contextual factors. Yet, the resulting perceptions remain fragmented across specific use cases; this underscores the need for a systematic, cross-domain mapping to identify the stable mental models that underpin societal acceptance of AI.

2.2 Similarities and differences in risk perception between experts and the public

A robust finding across scientific disciplines is the systematic divergence between expert and lay perceptions of risk, a phenomenon that is particularly pronounced in the context of emerging technologies like AI. This "perception gap" is not merely a product of varying knowledge levels but reflects a difference in evaluative frameworks: while experts typically adopt a probabilistic and technical approach to risk, the public prioritizes qualitative dimensions such as trust, ethical implications, and dread. By comparing findings from general risk research with AI-specific studies, this section illustrates

that experts often report higher trust and lower perceived risk than the general public. Identifying this perception gap is essential for governance, yet direct comparisons using identical psychometric frameworks remain remarkably scarce; a research gap that the present study seeks to fill.

First, research has shown that health experts and the public often differ in their assessments of health risks (Krewski et al. 2012). Experts typically perceive behavioral health risks (such as smoking and obesity) as more significant, while the public may prioritize other concerns. This discrepancy underscores the importance of effective risk communication strategies to better align public perception with expert evaluations.

Similarly, public perception of risks related to industrial production facilities tends to be more subjective and emotionally driven, in contrast to the more objective evaluations made by safety professionals (Botheju and Abeysinghe 2015). Such misalignments call for two-way communication approaches to address concerns proactively and prevent unnecessary escalation.

In their study of expert and laypeople perception of nanotechnology, Siegrist et al. 2007 found that while experts' judgments are primarily driven by technical evidence and probabilistic risk, the public relies heavily on trust and the 'affect heuristic' (the inverse relationship between perceived risk and benefit). The findings revealed that differences in risk perception extended beyond knowledge gaps, reflecting underlying value-based judgments. We will extend this inquiry to the domain of AI to determine if this systematic perception gap persists across diverse AI applications, or if the unique societal integration of AI alters the traditional expert–layperson divergence identified in other studies.

In the case of environmental hazards, such as nuclear waste, experts and laypeople also differ markedly in their perceived risks. Here, risk perception is shaped more by attitudes and moral values than by cognitive factors (Sjöberg 1998). Laypeople, who generally have less technical knowledge, tend to rely on intuitive and emotional reasoning, whereas experts base their assessments more on technical evidence and probabilistic analysis (Sjöberg 1998). This difference often leads to divergent views on policy and regulation, with members of the public perceiving higher levels of risk than experts (Siegrist et al. 2007). The public prioritizes ethical and societal implications, while experts focus primarily on scientific and technical risks.

In contrast, when it comes to natural hazards such as hurricanes and cyclones, there is often a considerable degree of agreement between expert risk assessments and public perceptions, particularly in high-risk areas (Peacock et al. 2005; Md. Abdus and Cheung 2019). However, public risk perception in these contexts can still be shaped by factors such as trust in authorities and prior experience with disasters. In the domain of autonomous vehicles (AVs), public risk perception is strongly influenced by trust in both the technology and the institutions that regulate it. Greater knowledge about AVs can enhance trust, which in turn reduces perceived risk, highlighting the importance of targeted trust-building initiatives (Robinson-Tay and Peng 2024). In aviation, by contrast, experts typically possess a more accurate understanding of relative risks, while novices' perceptions may be distorted by overconfidence or limited experience (Thomson et al. 2004).

Elena and Johnson (2015) examined differences in expert and public perceptions of cloud computing services. The findings indicate that experts tend to have a more nuanced understanding of risks, particularly regarding data security and integrity, while members of the public are more likely to experience a generalized dread risk in response to unfamiliar or abstract technological threats. Perceptions were also influenced by factors such as trust in regulatory bodies and the perceived benefits of the technology.

A decade ago, Müller and Bostrom (2016) surveyed AI experts on their expectations for the future capabilities of artificial intelligence, finding that most anticipated the emergence of superintelligence between 2040 and 2050. Notably, one-third of these experts considered this development *bad* or *extremely bad*, highlighting substantial concerns even within the expert community.

Crockett et al. (2020) compared trust and risk perceptions of AI between the public and computer science students as individuals with—supposedly—above-average expertise in AI. The findings revealed clear differences, suggesting that education plays a crucial role in increasing trust and reducing perceived risks. Similarly, Novozhilova et al. (2024) found that greater technological competence and familiarity with AI are associated with higher levels of trust in AI systems.

Still, comprehensive comparisons between experts and the public remain relatively rare. Recently, Jensen et al. (2024) conducted interviews with 25 members of the public and 20 AI experts in the United States to examine their perceptions of AI. Both groups emphasized that AI systems reflect the values and biases of their creators, acknowledging inherent limitations in the technology. Ethical concerns cited by participants included AI's lack of transparency, its profit-driven development, and the risk of exacerbating existing social inequalities. Human oversight was widely supported, particularly in high-stakes contexts such as healthcare. Although AI is perceived as efficient, its inability to replicate human empathy emerged as a central barrier to trust. Across both groups, reflections on humanness and ethics played a critical role in shaping attitudes toward AI.

A recent study surveyed 111 AI experts to assess their beliefs about catastrophic AI risk, familiarity with AI safety concepts, and responses to alignment arguments (Field 2025). It finds that experts divide into two broad

perspectives: viewing AI as a controllable tool or as a potentially uncontrollable agent, overall with lower concern about AI risk closely linked to limited familiarity with core AI safety concepts.

Recently, in a survey of 2,778 AI researchers, Grace et al. (2025) report that experts have accelerated their forecasts for high level of machine intelligence, now predicting a 50% probability of its arrival by 2047 (thirteen years earlier than estimated in 2022). However, while a majority of experts foresee positive outcomes, nearly half assigned at least a 10% probability to extremely catastrophic scenarios, including human extinction.

In summary, while the literature highlights a consistent divergence between expert and public risk assessments, the underlying reasons for this "perception gap" remain under-explored. This study addresses this gap by employing a unified psychometric framework to compare how both groups weigh risks and benefits across a broad spectrum of AI scenarios.

2.3 Risk perception and the psychometric model

To address the fragmented nature of AI perception research, this study adopts the psychometric paradigm as its primary theoretical lens. Originally developed by Slovic et al. (1986), this framework posits that risk perception is not a technical calculation of probability but a subjective construct influenced by factors like "dread" and "knowability". It focuses on individuals' subjective evaluations of risk, often measured through rating scales (Fischhoff et al. 1978; Slovic et al. 1986). The psychometric paradigm has been successfully applied across diverse contexts, including nuclear energy (Slovic et al. 2000), gene technology (Connor and Siegrist 2010), genetically modified food (Verdurme and Viaene 2003), climate change (Pidgeon and Fischhoff 2011), and carbon capture (Arning et al. 2020). Consequently, it is well suited for studying risk perception in emerging technologies, as it offers a structured framework for understanding how individuals evaluate and navigate the complex balance between perceived risks and benefits.

Crucially, research within this paradigm has revealed a consistent inverse relationship between perceived risk and benefit, known as the *affect heuristic* (Alhakami and Slovic 1994; Slovic et al. 2007; Efendić et al. 2021). This heuristic suggests that evaluative judgments are often guided by an overall affective state: technologies perceived as highly beneficial are frequently judged as lower risk, while those viewed as risky are rated as less beneficial. By applying this framework to AI, we can determine whether the perception gap results from experts and the public utilizing different *weighting schemes*—that is, whether one group's overall value judgments are driven more by perceived benefits while the other is more sensitive to perceived risks.

2.4 Research questions

The preceding review reveals two critical limitations in current AI perception research: (1) a lack of direct comparisons between academic experts and the general public using consistent evaluative criteria, and (2) a focus on narrow domains that obscures the broader mental models underlying perception. This study addresses these gaps by applying the psychometric paradigm to 71 diverse scenarios, providing a comprehensive cognitive map of the "AI perception gap". Identifying these divergences is essential, as misaligned risk–benefit perceptions may impede technology design, individual and societal acceptance, and regulatory effectiveness. To this end, we investigate the following research questions:

1. Perception and impact: How do academic AI experts (who educate and influence practitioners, contribute to foundational research and develop methodological frameworks, and provide governance-relevant expertise that inform AI system development) and members of the public (who use or are affected by AI-based systems) differ in their perceptions of AI's capabilities and impacts?
2. Value attribution: What values do academic AI experts and the public assign to potential AI futures? In which domains is AI perceived positively or negatively, and how do these perceptions differ between the two groups?
3. Risks and benefits: What risks and benefits are perceived in AI's potential futures? How do these perceptions differ between experts and the public?
4. Risk–benefit trade-offs: Is the overall value attributed to AI more strongly influenced by perceived risks or by perceived benefits? How do these trade-offs vary between experts and the general public?

3 Methods

This section details the empirical approach used to investigate the perception gap between academic experts and the general public. To ensure a comprehensive mapping of AI perceptions, our methodology transitions from theoretical constructs to a comparative analysis across 71 distinct scenarios. The following subsections describe: (3.1) the systematic development of the scenario catalog and the operationalization of variables, (3.2) the recruitment and composition of the two participant samples, (3.3) the specific characteristics of both samples, (3.4) the statistical procedures employed for data filtering and the analysis of group differences and weighting schemes, and (3.5) a robustness check to verify that the international expert sample is comparable to the German public sample.

3.1 Survey design

To address our research questions, we designed two surveys: one for the public and another for academic AI experts. Both surveys centered on a shared core, in which participants evaluated a randomized random subset of 15 out of 71 micro-scenarios (to reduce participant fatigue), each describing a potential AI development or imaginary AI future (Brauner et al. 2022). Participants rated each scenario using five assessment items (see below). This design supports two complementary perspectives in interpreting the data: 1) individual-level interpretation: responses serve as reflexive indicators of participants' underlying dispositions or differences. 2) Scenario-level interpretation: responses reflect participants' attributions to specific technologies or topics, enabling analysis and visual mapping across groups.

To develop the list of scenarios, we first identified potential capabilities and impacts that AI might have in the near future. Drawing on prior research (Brauner et al. 2023) and three interdisciplinary expert workshops with partners from our interdisciplinary collaborations (including specialists in Computer Science, Ethics, and Sociology), we compiled an initial set of potential survey topics. For topic selection, we built on Luhmann's system theory ("Systemtheorie"), which conceptualizes modern society as comprising six core subsystems: economy, law, science, politics, religion, and education (Luhmann 1989). Each of these subsystems fulfills a unique and non-substitutable function within society (Peterson et al. 2004). Through multiple iterative rounds, we ensured that each subsystem was represented, removed redundant items, and refined the phrasing of topics for clarity and conciseness (Brauner 2024) details strategies for development of the topic list, discusses potential biases, and outlines mitigation strategies). The final set includes both plausible and speculative AI projections, such as "AI creates many jobs", "AI promotes innovation", "AI acts according to moral concepts", and "AI considers humans as a threat". Table 4 in the Appendix lists all 71 projections.

We used five dependent variables to evaluate each topic on 6-point single-item semantic differential scales. Single-item measures enable efficient data collection and are considered appropriate for capturing well-defined constructs such as risk, benefit, and value judgments (Rammstedt and Beierlein 2014; Fuchs and Diamantopoulos 2009; Wolfers and Baumgartner 2025). The use of 6-point scales avoids midpoint bias by eliminating a neutral option, thereby encouraging more decisive responses (Tourangeau et al. 2000).

- Expected likelihood: How likely is it that this development will occur within the next 10 years (*"will not happen–will happen"*)?
- Personal risk: How do you assess the risk of this development for yourself personally (*"low risk–high risk"*)?
- Societal risk: How do you assess the risk of this development for society as a whole (*"socially harmless–socially harmful"*)[1]?
- Perceived benefit: How beneficial or useful do you consider this development (*"useless–useful"*)?
- Attributed value: Assuming this development comes true, would you consider it positive or negative (*"negative–positive"*)?

To assess participants' overall evaluation of each topic, the final item drew on the value-based adoption model (VAM), which conceptualizes attributed value (ranging from positive to negative) as a suitable criterion for technology evaluation (Kim et al. 2007).

In addition to these topic-specific evaluations, we collected demographic information. All participants were asked to report their age (in *years*) and gender (following Spiel et al. (2019) as *male*, *female*, *diverse*, or *prefer not to say*).

In the AI expert survey, participants also self-reported their level of expertise in AI using a five-point scale (*no expertise in the field of AI*, *basic knowledge*, *well informed*, *expert*, *recognized authority in the field of AI*). Additional items assessed years of experience, number of scientific publications, current country of residence, and scientific discipline.

In the survey for the public, we collected data on education level, employment status, and several individual difference measures potentially influencing AI perception. These included Technology Readiness (Neyer et al. 2016), an AI Readiness Scale (Karaca et al. 2021), and interpersonal trust (Niessen et al. 2021). A detailed analysis of the impact of these individual factors is presented in a separate publication (Brauner et al. 2026). Figure 1 illustrates both surveys' design.

Both surveys began with an informed consent, informing that participation was voluntary, that no personally identifiable information would be collected, and that the resulting data would be made available as open data. The expert questionnaire was administered in English, whereas the survey for the public was conducted in German. Our university's

[1] This item shows a very strong correlations with personal risk across all topics and both samples and will be dropped in later. Based on a Reviewer's comment, we conducted an exploratory analysis of *harmfulness* in place of *risk*. The results are similar in terms of the significance of the absolute and relative (trade-offs) differences but different in terms of the absolute values and regression coefficients and may reflect different cognitive mechanisms, suggesting further investigation. The data are publicly available for researchers specifically interested in social risk. The repository also contains tables from this exploratory analysis.

Institutional Review Board (IRB) granted ethical approval (protocol ID 2023_02b_FB7_RWTH Aachen).

3.2 Sample acquisition

The general public sample was acquired using the independent fieldwork agency Consumerfieldwork GmbH, Germany. A total of 1354 participants were recruited to represent the German population in terms of age, gender, socio-economic status, and regional distribution. Participants received an incentive of approximately 1 Euro for completing the survey. While this rate is lower than the standard hourly wage, it was calibrated to acknowledge effort without creating undue inducement that might compromise voluntary consent. Data for the public were collected between 2023-07-04 and 2023-07-09.

The academic AI expert sample was obtained via convenience and snowball sampling. To initiate this process, we contacted individuals within our academic networks, including collaborators from joint projects and funding proposals. Additionally, we contacted authors by compiling a list of publicly available email addresses of recent publications in leading AI-related venues. Between 2023-03-06 and 2023-05-10, both groups were invited via email and encouraged to forward the invitation to colleagues. In total, we contacted approximately 450 AI researchers worldwide, of which 139 proceeded beyond the landing page of the survey. To encourage participation, the expert invitation emphasized the academic relevance of the study and the open availability of its findings.

3.3 Data cleaning and analysis

We used both parametric and non-parametric statistical methods, including Bravais–Pearson (r) and Kendall's Tau (τ) correlation coefficients, Chi-square (χ^2), ordinary least squares (OLS) multiple linear regressions with robust standard errors, and multivariate analyses of variance (MANOVA) using Pillai's trace (V) as the multivariate test statistic and effect size metric (as it represents the proportion of total variance in the dependent variables explained by the group factor (Stevens 2012)). Due to substantial different group sizes between the experts and the public, we ensured that all methods used are robust to sample imbalance. OLS point estimates are generally reliable under unequal sample sizes (Fox 2016). As their standard errors may be biased, we used robust standard errors. We further cross-validated key findings through random subsampling with equal group sizes (reducing the public's sample to the size of the expert sample by random selection), which yielded similar results. Consistent with social science standards, we set the Type I error rate to $\alpha = .05$ to determine statistical significance (Field 2009). Prior to conducting the analyses, we verified that all necessary statistical assumptions (including homoscedasticity and the absence of multicollinearity) were met and detected no violations. Beyond the core analyses, we conducted an exploratory cluster analysis, reported in Sect. A.2 of the Appendix, to examine whether the survey topics exhibit higher level structure and whether this structure is shared between expert and public evaluations.

We filtered the data from both samples for incomplete or low-quality responses using both reactive and non-reactive criteria (Leiner and Johannes 2019). Participants from the public were included if they completed the survey and passed an attention-check ("Please select 'rather agree'"). For the expert sample, we included only those participants reporting at least basic knowledge of AI and had more than 1 year of experience working in the field. In both samples, we excluded participants who completed the survey in less than one-third of the median completion time ("speeders"). This threshold is generally considered sufficient to identify meaningful responses (Leiner and Johannes 2019). The

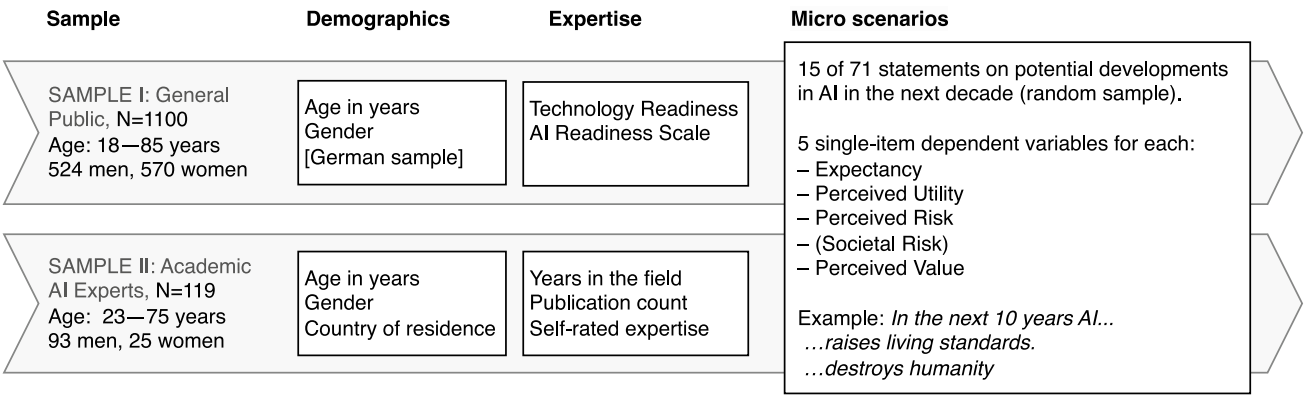

Fig. 1 Design of both surveys. In each, participants evaluated 15 randomly selected AI projections (or vignettes) from a pool of 71. The measured explanatory user factors differed between academic AI experts and the general public

median completion times of the surveys was 9.8 min for the laypeople and 9.5 min for the experts. Based on these criteria, we excluded 254 out of the 1354 participants in the public sample (exclusion rate of 18.8%) and 20 of the 139 participants in the expert sample (exclusion rate of 14.4%). All the participants fully completed the survey.

All reported statistical analysis, values, and figures in this manuscript are generated using RMarkdown and R version 4.5.1 (2025-06-13) directly from the underlying data, supporting transparent and reproducible research. Materials, including the full survey, (unfiltered) data and the filtering procedures, the manuscript with all analyses, and additional data tables and figures, are available as open data on OSF (https://osf.io/gt9un/).

3.4 Sample 1: general public

The sample of the public consists of 1100 participants, with 524 identifying as male and 570 as female. The participants' ages ranged from 18 to 85 years, with a median age of 51 years (SD = 14.2) years. There was no significant correlation between age and gender in the sample ($\tau = -0.031$, $p = 0.210 > .05$). A complementary article provides an in-depth analysis of the sample, focusing on the effects of individual differences (Brauner et al. 2025).

3.5 Sample 2: academic artificial intelligence experts

The expert sample consists of 119 participants with 93 identifying as male and 25 as female. Participants' ages ranged from 23 to 75 years, with a median age of 36.5 years (SD = 13.4 years). Again, there was no significant association between age and gender ($\tau = -0.113$, $p = 0.143 > .05$). The majority of the experts were from Germany (N = 77), followed by the Netherlands (N = 10). All other countries received less than 10 mentions.

Regarding the scientific background, most experts worked in general computer science (43%), artificial intelligence (24%), engineering (14%), or other disciplines, such as business administration, innovation management, and the social sciences (19%), reflecting both AI as a distinct disciplinary identity and AI-focused work emerging from adjacent fields. AI experience ranged from 1 to 40 years with an arithmetic mean of 10.3 and a median of 5 years. In total, the participants reported 2713 academic publications, with an arithmetic mean of 22.8 and a median of 5 publications per expert. As indicated by the high Gini coefficient ($G = 0.771$), the number of publications is unevenly distributed among the experts—many have few, while a few have many—suggesting that both prominent figures and junior researchers have participated. When asked for their level of expertise, 11 participants reported having basic knowledge, 51 reported being well informed, 48 reported being experts, and 9 described themselves as a "*recognized authority in the field of AI*".

3.6 Robustness check of the country composition of the expert sample

To assess the robustness of our findings with respect to the country composition of the expert sample, we conducted two additional multivariate analyses. Given that both the public sample and the majority of experts were drawn from Germany, these analyses were designed to examine potential bias arising from the expansion of the sample by including experts from outside Germany. Specifically, we tested (a) whether experts and the public from Germany differed in their evaluations, and (b) whether German experts differed from experts from other countries in their evaluations of expected likelihood, perceived risks and benefits, and overall attributed value. To this end, we estimated two MANOVAs, using sample (experts vs. public) as the independent variable for the German subsample and country (Germany vs. other countries) as the independent variable for the expert sample. (a) The MANOVA comparing German experts and members of the German public revealed a significant multivariate effect and significant differences in expected likelihood, perceived risks and benefits, and overall attributed value (Pillai's Trace $V = 0.060$, $F(5, 1171) = 15.07$, $p < .001$, $\eta^2 = 0.060$). (b) In contrast, the MANOVA comparing German experts with experts from other countries did not yield a significant multivariate effect (Pillai's Trace $V = 0.058$, $F(5, 112) = 1.37$, $p = 0.241$, $\eta^2 = 0.058$). Taken together, these results indicate that the contrast between the public and experts is comparable when restricted to the German subsample, and that extending the expert sample to include participants from other countries is admissible in the context of this study.

4 Results

We begin by analyzing differences between the public and academic AI experts in terms of overall perceived *societal and individual risk*, *benefit*, *attributed value*, and expected *likelihood*, that is, how likely each of the AI projections is to come true within the next decade. Next, we interpret participants' responses as reflexive measurements of latent constructs: Expectancy, risk, benefit, and value of AI are treated as *individual-level personality differences*, allowing us to examine if individuals of both groups trade-off perceived risks and benefits differently. Finally, we shift the perspective and interpret the responses at the *topic level*, focusing on how specific AI-related statements were evaluated. We assess how experts and the public differ in their expectations and valuations of these projections, including how

they attribute risks and benefits to individual AI scenarios. Table 4 in the Appendix provides a complete overview of all ratings across all statements for both groups (a searchable and sortable spreadsheet is available on https://osf.io/gt9un/files/p7uet).

4.1 Do overall evaluation scores differ between academic AI experts and the public?

In this section, we examine how both groups evaluated the AI projections across the queried dependent variables: perceived individual and societal risk, perceived benefit, attributed value, and expectancy (i.e., the expected likelihood that the respective AI development will occur within the next decade). As illustrated by the box plots in Fig. 2 and the complementary Table 1, notable differences emerge between both groups.

A one-way MANOVA with group (public vs. experts) as independent variable revealed a medium and significant difference between both samples on the combined dependent variables ($V=0.085$, $F(5, 1213) = 22.572$, $p<0.001$). Follow-up univariate analyses showed that four dependent variables exhibited significant group differences: Expectancy: $F(1, 1217) = 18.572$, $p<0.001$, risk: $F(1, 1217) = 26.716$, $p<0.001$, benefit: $F(1, 1217) = 43.443$, $p<0.001$, and value: $F(1, 1217) = 21.929$, $p<0.001$). However, the perceived societal risk did not differ between both groups ($F(1, 1217) = 1.775$, $p=0.183$). These findings indicate that academic AI expertise significantly influences not only the overall multivariate response but also each individual evaluation dimension besides the perception of societal risk.

In particular, AI experts, on average, rated the queried projections as more likely to happen within the next decade (25.2), personally safer (19.3), more useful (15.3), and more positive overall (−4.0) compared to the members of the public (expectancy: 12.7, individual risk: 34.7, benefit: −5.2, value: −19.7).

Table 1 Comparison of average assessments across the four evaluation dimensions expectancy, risk, benefit, and value for AI experts and the public

	AI Experts (N=119)				General Public (N=1100)				Sig.
Dimension	M	SD	CI-95	N	M	SD	CI-95	N	p
Expectancy	25.2	67.4	[13.1,37.3]	119	12.7	66.7	[8.8,16.7]	1100	<0.001
Perceived societal risk	19.2	58.7	[8.6,29.7]	119	23.4	54.5	[20.2,26.7]	1100	0.183
Perceived individual risk	19.3	63.1	[7.9,30.6]	119	34.7	56.4	[31.4,38.1]	1100	<0.001
Perceived benefit	15.3	62.1	[4.1,26.4]	119	−5.2	59.0	[−8.7,−1.7]	1100	<0.001
Attributed value	−4.0	60.9	[−14.9,7.0]	119	−19.7	57.8	[−23.2,−16.3]	1100	<0.001

A MANOVA suggests statistically significant medium (Pillai's V = 0.085) overall group effect and significant differences on each individual dependent variable but societal risk

Measured on 6-point scales (rescaled to -100 to +100). Negative values indicate a negative evaluation of the respective dimension (i.e., low value, low risk, low benefits, or low expectancy)

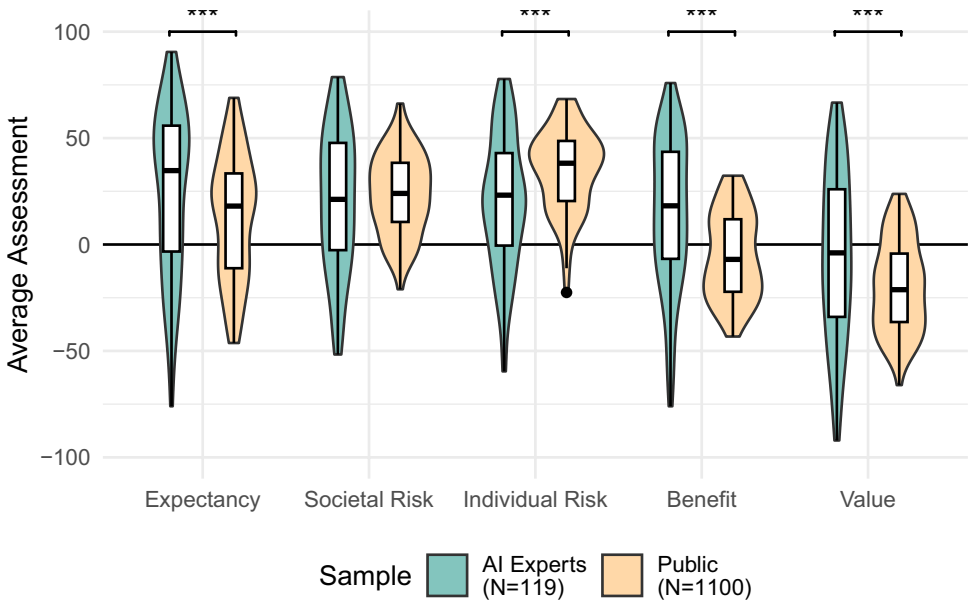

Fig. 2 Box plots showing the overall average evaluations across all 71 topics for each assessment dimension, separately by sample (experts vs. public). The shaded areas represent the distribution density of the data. Asterisks indicate statistically significant differences between the two groups for each dimension (all at $p < .001$)

Table 2 Correlation analysis of the assessment dimensions across the 71 topics for academic AI experts and the public

		AI experts (N=119)			General Public (N=1100)		
Variable 1	Variable 2	r	95% CI	p	r	95% CI	p
Expectancy	Societal risk	−0.085	[−0.26,0.10]	0.359	0.009	[−0.05,0.07]	0.763
Expectancy	Individual risk	0.144	[−0.04,0.32]	0.236	0.212	[0.15,0.27]	<0.001
Expectancy	Benefit	0.401	[0.24,0.54]	<0.001	0.143	[0.08,0.20]	<0.001
Expectancy	Value	0.393	[0.23,0.54]	<0.001	0.039	[−0.02,0.10]	0.386
Societal risk	Individual risk	0.528	[0.38,0.65]	<0.001	0.781	[0.76,0.80]	<0.001
Societal risk	Benefit	−0.368	[−0.51,−0.20]	<0.001	−0.711	[−0.74,−0.68]	<0.001
Societal risk	Value	−0.601	[−0.70,−0.47]	<0.001	−0.817	[−0.84,−0.80]	<0.001
Individual risk	Benefit	−0.295	[−0.45,−0.12]	0.003	−0.639	[−0.67,−0.60]	<0.001
Individual risk	Value	−0.418	[−0.56,−0.26]	<0.001	−0.749	[−0.77,−0.72]	<0.001
Benefit	Value	0.712	[0.61,0.79]	<0.001	0.869	[0.85,0.88]	<0.001

Correlations of the average evaluations of the 71 queried topics (interpreted as perspective 1: individual differences)

4.2 Do academic AI experts and the general public differ in weighing AI risks and benefits?

Beyond differences in absolute evaluations, Table 2 reveals systematic differences in how academic AI experts and the public weigh risks and benefits across AI topics. For the public, perceived risks and benefits are strongly negatively correlated, both at the societal level ($r = -0.711$) and the individual level ($r = -0.639$), indicating a pronounced trade-off structure in which AI applications are evaluated.

Among experts, these associations are substantially weaker (societal risk: $r = -0.368$; individual risk: $r = -0.368$), suggesting a more differentiated assessment in which risks and benefits are less tightly coupled.

Similar differences emerge for the relationship between expectancy and risk. While higher expected likelihood is positively associated with individual risk perceptions among the public $r = 0.212$, no such association is observed among experts. This pattern indicates that members of the public are more likely to interpret likely AI applications as personally risky, whereas experts do not systematically conflate likelihood with harm. In addition, individual and societal risk perceptions are more strongly correlated among the public ($r = 0.781$) than among experts ($r = 0.528$), suggesting that lay evaluations more often collapse personal and collective risk into a single dimension.

Together, these findings suggest that expert–public disagreements are driven less by opposing views on AI's societal consequences and more by differences in how risks and benefits are cognitively structured and weighed. While both groups associate the overall value of AI negatively with perceived risks and positively with benefits, the public's judgements appear more trade-off driven, whereas expert assessments reflect a more nuanced separation between these dimensions. Crucially, because perceived risks and benefits are themselves inversely related, their individual effects on AI evaluation are intertwined. To disentangle overlapping influences and isolate the unique contribution of each predictor within this risk–benefit trade-off, we conducted three multiple linear regression analyses. In these, we focused on variables reflecting individual-level cognitive attributions (specifically individual risks and benefits) while omitting the societal risk variable to maintain a consistent level of analysis. First, we ran separate regressions for each sample, using personal risks and benefits as independent variables and the attributed value of AI as the dependent variable. Second, we conducted a combined-sample regression,[2] including the sample identifier as a third predictor to assess group-level effects (Chow 1960). All VIFs were ≤ 1.691, indicating no multicollinearity between the predictors.

Table 3 summarizes the results of all three significant regression models. The third regression model (right side of the table), which includes the factor distinguishing between the two samples, was significant and explained 80% of the variance in overall value ($F(3,1215)=1773.106$, $p < 0.001$, $R^2 = 0.807$). The grouping factor was significant, indicating a small but meaningful difference in the risk–benefit trade-offs between academic AI experts and the public ($\beta = 0.044$, $p < 0.006$). Given these significantly different trade-offs, we proceed by analyzing the two samples separately in the following sections. Figure 3 illustrates both models.

For the sample of academic AI experts (left side of the table), the regression model was significant, explaining over 54% of the variance in value attributed to AI ($F(2, 116) = 48.5$, $p < 0.001$, $R^2 = 0.547$. The intercept was

[2] While unequal group sizes may result in heteroskedasticity affecting the standard errors, this does not bias the regression coefficient estimates (βs) (Fox 2016). As cross-validation, we additionally performed a subsampling analysis using small random subsets of the public sample ($N' = 110$), which yielded comparable results.

significant and negative ($I = -0.101$), indicating a slightly negative baseline evaluation, i.e., if an expert held neutral views on both AI risks and benefits, their overall perceived value of AI would tend to be slightly negative. Perceived AI risk had a moderate negative effect on value ($\beta = -0.195$), whereas AI benefit had a strong positive effect ($\beta = 0.623$). Importantly, the influence of perceived benefits was approximately three times stronger than the influence of perceived risk (×3.19).

For the public (middle part of the table), the regression model was likewise significant and showed an even stronger fit, explaining 82% of the variance in overall AI value ($F(2, 1097) = 2634.6$, $p < 0.001$, $R^2 = 0.819$). In contrast to the experts, the intercept was not significantly different from zero, indicating no consistent baseline tendency in AI evaluations when perceived risks and benefits are neutral. Overall value judgments were significantly influenced by both perceived risk ($\beta = -0.361$) and perceived benefit ($\beta = 0.703$) with benefits having about twice the impact of risks (×1.95). Notably, the negative effect of perceived risk was significantly and substantially larger for the public compared to the experts, highlighting a greater sensitivity to potential downsides of AI among laypeople.

4.3 Do academic AI experts and the public share the same vision for AI's future?

As the previous analysis shows, academic AI experts and the public differ in their expectations of what AI is likely to achieve over the next 10 years (hereafter referred to as expectancy). To illustrate these differences at the topic level, Fig. 4 presents a scatter plot of the average expectancy ratings for each AI projection.

Each point in the figure represents an AI projection, with its position determined by the mean expectancy rating from academic AI experts (on the x-axis) and from the public (on the y-axis). This serves as a comparative map of perceived likelihoods for both groups: points farther to the right indicate higher expectations among experts, while points higher on the plot reflect higher expectations among members of the public. Items that lie on or near the dashed diagonal indicate agreement between the two groups: experts and the public rated the likelihood of these scenarios similarly. In contrast, points that substantially deviate from the diagonal indicate disagreement. Items positioned above the diagonal are considered more likely by the public than by experts, while those below are judged more likely by the experts. The blue regression line indicates the best linear fit (maximum likelihood estimation), and the shaded gray band around it shows its 95% confidence interval, highlighting the uncertainty range for the estimated relationship.

The figure illustrates both notable similarities and distinct differences in expectancy assessments between both groups. First, there is an overall strong and positive correlation between the two groups' expectations, indicating substantial agreement on the perceived likelihood for many of the AI projections ($r = 0.734$, $p < 0.001$). Second, the distribution of expectancy ratings among the public is markedly narrower than that of the experts, suggesting that experts hold a more nuanced and differentiated view of AI's future potential. This difference in variance is also reflected in the flatter slope of the regression line. Third, despite this general agreement, several topics exhibit strong divergences between

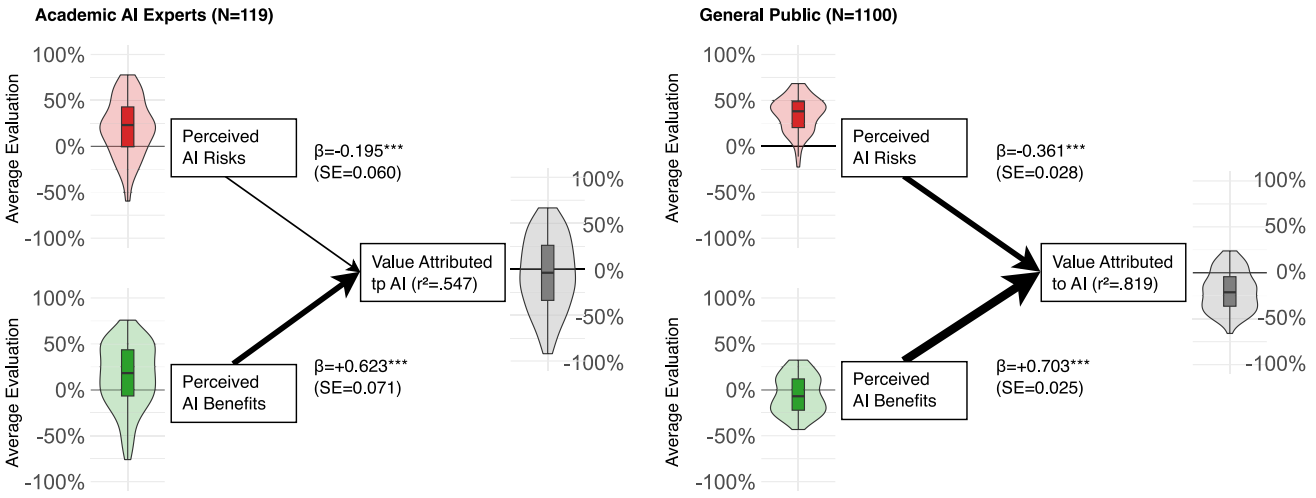

Fig. 3 Visualization of the two significantly different multiple linear regression models for academic AI experts (left) and the public (right), including variable distributions shown as violin and boxplots. In both models, perceived risk and benefit significantly predict attributed AI value. However, the explained variance and the influence of risk are greater in the public. Line thickness reflects the strength of each predictor's standardized effect (β). All predictors are significant at $p < .001$

Table 3 Results from three multiple linear regressions examining the relationship between perceived AI risks and benefits (as rated by academic AI experts and the public) and overall attributed value

	Dependent variable: value attributed to AI		
Independent variables	AI experts (N=119)	General Public (N=1100)	Combined
(Intercept)	−0.101***	−0.036	−0.085***
(SE)	(0.024)	(0.009)	(0.017)
Perceived risk β	−0.195***	−0.361***	−0.346***
(SE)	(0.060)	(0.028)	(0.025)
Perceived benefit β	0.623***	0.703***	0.704***
(SE)	(0.071)	(0.025)	(0.023)
Sample β	—	—	0.044*
(SE)			(0.016)
F	F(2,116)=48.526	F(2,1097)=2634.605	F(3,1215)=1773.106
p	<0.001	<0.001	<0.001
adj. R^2	0.547	0.819	0.807

The third model includes the sample as an additional predictor (combined sample). Results indicate significant differences in risk–benefit trade-offs between experts and the public, with experts placing comparatively less emphasis on AI-related risks. All models use robust standard errors to account for unequal sample sizes

*$p < 0.05$, **$p < 0.01$, ***$p < 0.001$

the two groups. Specifically, the public expressed higher expectations than experts for statements such as "AI will destroy humanity", "AI will lead to personal loneliness", and "AI can no longer be controlled by humans". Conversely, experts reported substantially higher expectations for statements like "AI will improve our health", "AI will prefer certain groups of people", and "AI will become humorous". Table 4 in the Appendix presents the overview of all items and their expectancy evaluations for both samples.

4.4 Do value attributions differ between academic AI experts and the public?

Next, we examine the similarities and differences in the overall value attributed to AI (or AI sentiment) between academic AI experts and the public. In Fig. 5, we show a scatter plot of average value ratings for each AI projection, with expert ratings on the x-axis and public ratings on the y-axis. It shows the strong and positive correlation between both groups' value assessments ($r = 0.855$, $p < 0.001$) that is notably higher than the correlation in expectancy ratings. This suggests a greater degree of agreement of both groups regarding the value attributed to AI.

Despite this overall agreement, important differences remain. On average, experts tend to hold a more positive outlook toward AI than the public. However, the sentiment diverges substantially for specific topics. The public, for instance, expresses more negative sentiment toward statements involving existential or societal threats, such as AI destroying humanity, increasing social division, or perceiving humans as a threat. In contrast, experts are more optimistic about scenarios involving AI acting in accordance with moral values, contributing to sustainability, or supporting medical decision-making. These differences reflect not only diverging attitudes toward specific applications but also broader differences in how each group weighs potential risks and benefits.

4.5 Do experts and the public differ in risk and benefit attributions?

As the overall sentiment towards the AI projections differs between academic AI experts and the public, we analyze the attributed risks and benefits for each group independently. The left side of Fig. 6 shows the risk–benefit attributions for AI experts, and the right shows these for the public.

The horizontal axis represents the perceived risk attributed to each topic, and the vertical axis represents the perceived benefit. These plots only illustrate the relationship between risks and benefits, without integrating the value attributed to the projections. A regression analysis examining the relationships among perceived risks, benefits, and overall value is presented in Sect. A.3 of the Appendix.

The points in the plots illustrate where the AI projections lay in terms of risks and benefits. Points below the horizontal axis are seen as useless, while items above are seen as useful. Similarly, items on the left are perceived as more risky, whereas items to the right are viewed as safer. Items in the top-right quadrant are perceived as both risky and useful, while those in the bottom-left quadrant are considered neither risky nor useful. Items in the top-left quadrant are seen as risky but not useful, and items in the bottom-right quadrant are evaluated as not risky but useful.

The assessments of the AI experts display greater disparity compared to those of the public, suggesting more nuanced evaluations. Experts' risk and benefit attributions

span a broader range, with a substantial proportion of topics deemed useful and fewer topics assessed as useless. Also, experts perceive topics as both risky and safe across the spectrum. In contrast, the public's evaluations are more concentrated, with most topics viewed as rather risky.

5 Discussion

AI is reshaping our world. In fact, studies suggest that AI and LLMs already alter both communication practices and professional work workflows (Geng et al. 2025; Brachman et al. 2024). In this study, we examined how perceptions of AI differ between those who are affected by its societal implementation—the general public—and those who educate practitioners, inform policymaking, conduct foundational research, and advance the technology—academic AI experts. Overall, both groups acknowledged that AI is here to stay, as the majority of queried AI projections received above-average scores for their expected likelihood of occurrence. Across the wide range of statements, participants from both groups reported individual risk scores above the scale's midpoint. However, AI experts consistently reported greater benefits than the public. Notably, the public's overall evaluations were more negative than those of the experts. Overall, this finding aligns with previous studies comparing expert and lay perceptions in other domains for example, Siegrist et al. (2007) on nanotechnology, which similarly found differences in technology perception and that experts tend to perceive lower individual risks than laypeople. Given the broad range of topics in the statements—spanning from plausible to speculative—the absolute scores are less informative than the relative differences between both groups.

5.1 Diverging expectations for the future of AI

First and foremost, experts and the public diverged in their expectations regarding the likelihood of AI developments. Compared to the public, experts generally assessed most AI projections as more probable. However, variance among expert responses was also substantially higher, suggesting more differentiated and topic-specific evaluations: That is, while experts rate many topics as much more likely, they also judge some projections as markedly less probable than the public does.

It is important, though, not to overinterpret the absolute expectancy ratings, as both experts and laypeople are known to struggle with forecasting future developments (Recchia et al. 2021). Although expert predictions tend to be slightly more accurate, the differences are typically small, reflecting the difficulty of making reliable forecasts about complex and uncertain technological trajectories; even with domain expertise (Recchia et al. 2021). Nonetheless, the relative differences are informative. Expert expectations are likely to influence the direction of research, funding, and policy development, while public expectations may shape political discourse, regulatory demands, and potentially failing societal acceptance. A misalignment between these perspectives could result in mismatched expectations, premature or excessive regulation, or public resistance to otherwise beneficial AI technologies.

5.2 Differences in value perception

In terms of the overall value attributed to the AI scenarios, experts generally perceive AI as more positive compared to the public. However, their evaluations are not uniformly optimistic: Again, experts exhibit greater variance, with more extreme ratings in both positive and negative directions. This suggests that experts may approach these evaluations more critically and deliberately, likely due to their knowledge and capacity to assess the nuanced trade-offs inherent in different AI applications.

While experts acknowledge significant opportunities, they also recognize risks and limitations that may not be as apparent to the public (Russell et al. 2015). In contrast, public opinion is often shaped by simplified narratives and popular media portrayals, which tend to emphasize dystopian or sensationalist aspects of AI (Cave et al. 2019; Puzanova et al. 2024). This tendency is arguably reinforced by a "negativity crisis" within the field of AI ethics itself (and probably beyond), where institutional incentives favor alarmist critiques over balanced scientific analysis (Königs 2025). This discrepancy highlights a (science) communication gap: For policy-makers and science communicators, this underscores the importance of addressing both the optimism of experts and the skepticism of the public. Promoting a more informed and balanced public dialogue will be essential for aligning societal expectations with technological realities and fostering trust in the development and deployment of AI systems.

5.3 Perception of risks and benefits and the formation of value judgments

For both groups, perceived risk and perceived benefit were inversely correlated, consistent with previous findings in risk perception research (Alhakami and Slovic 1994; Efendić et al. 2021). While this relationship may appear intuitive, it is not inevitable. In the context of emerging technologies, including AI, it is entirely plausible for individuals to perceive certain developments as both highly beneficial and highly risky. For example, powerful AI systems may be seen as offering substantial societal or economic benefits while simultaneously raising ethical or safety concerns. We assume that the observed inverse correlation reflects a psychological

tendency toward cognitive consistency. As suggested by earlier work (Alhakami and Slovic 1994; Efendić et al. 2021), individuals may align their perceptions of risk and benefit to avoid dissonant evaluations, simplifying their judgments by resolving ambivalence.

Beyond examining absolute evaluations, we analyzed the relative weight participants assigned to perceived risks and benefits in their overall value assessment of AI. In both groups, perceived benefits exerted a stronger influence on the overall valuation of AI than perceived risks. This pattern is consistent with previous findings in other technological domains, such as perceptions of nanotechnology hazards (Siegrist et al. 2007).

In our study, the strength of the influence of perceived benefits on overall AI valuation was comparable across both groups. However, we found a notable differences for perceived risk: among experts, risk had a significantly weaker impact on overall evaluation compared to the public. One possible explanation is that experts' greater familiarity with AI mitigates their perception of risk, reducing its influence. Conversely, the public may focus more on potential downsides, possibly due to limited knowledge, uncertainty, or exposure to more negative portrayals of AI in public discourse.

However, our analysis primarily focused on personal risk–benefit trade-offs, the strong correlations observed between societal and individual risk in both groups suggest that collective concerns play a significant role in AI valuation. As our study did not include a corresponding measure for societal benefits, we focused our regression models on the individual dimension to maintain conceptual symmetry. Future research should incorporate both personal and collective dimensions of benefits and value to determine if the "societal" vs. "individual" distinction creates a secondary trade-off axis. Investigating whether people are willing to accept personal risks for the sake of perceived societal gains—or vice versa—would provide a more granular understanding of the ethical and social weighing processes that drive AI acceptance.

5.4 Conclusion

Taken together, these findings reveal two levels of evaluative asymmetry between experts and the public: (1) in their absolute evaluations of AI, and (2) in their cognitive trade-off weighting risks and benefits. While experts tend to emphasize benefits and downplay risks, the public places relatively more weight on the potential harms. Such misalignment may pose risks for the societal uptake and governance of AI.

First, it may lead to public distrust, policy overreach, or underutilization of beneficial technologies. This underscores the need for fostering communication and negotiation across the key stakeholders: AI researchers, developers, policy-makers, and the public as users of AI. This is consistent with the EU AI Act's emphasis on public engagement and risk-based regulation (European Union 2024; Calero Valdez et al. 2024) and the OECD AI Principles (Organisation for Economic Co-operation and Development 2019), which call for human-centered, fair, and accountable AI systems.

Second, McLuhan's media theory states that "we shape media and then media shapes us" (McLuhan 1964; Culkin 1967)[3]. As AI becomes increasingly ubiquitous, it is likely to reshape social, cultural, and communicative norms. If AI and its applications are primarily designed by academic AI experts and practitioners trained within these benefit-oriented mental models, the resulting systems may unintentionally tilt toward technical affordances rather than reflecting broader societal values.

We characterize this risk as "procrustean AI", drawing on the Greek myth of Procrustes, who forced his guests to fit a standard bed by stretching or cutting their limbs. In a modern socio-technical context, this refers to a potential bias in the prioritization of research and the implementation of applications that implicitly privilege expert-centric perspectives. Without deliberate efforts toward participatory alignment, we risk creating systems that require the public to conform to rigid technological frameworks and expert assumptions, rather than developing technologies that are just, inclusive, and responsive to diverse human needs.

Our findings suggest the importance of transparent and participatory processes that integrate diverse perspectives into AI development, implementation, and governance. As recently argued (Shneiderman 2021; Decker et al. 2024; Calero Valdez et al. 2024), achieving responsible and trusted AI requires centering values such as equity, safety, inclusivity, and democratic responsiveness. To that end, the visual maps of differing AI perceptions can serve as an actionable tool to identify "tension points", such as justice and political decision-making, where public and expert perspectives align and diverge. Such cognitive maps can support participatory governance, guide research agendas, and highlight areas requiring critical reflection and public discourse. Ultimately, aligning these perspectives is necessary for ensuring the legitimacy and trust required for AI systems.

6 Limitations

Several conceptual and methodological aspects warrant consideration and provide directions for future research. These include the characteristics of our measurement and

[3] Thoreau equally stated that with technological progress, people risk becoming "*the tools of their tools*" (1854).

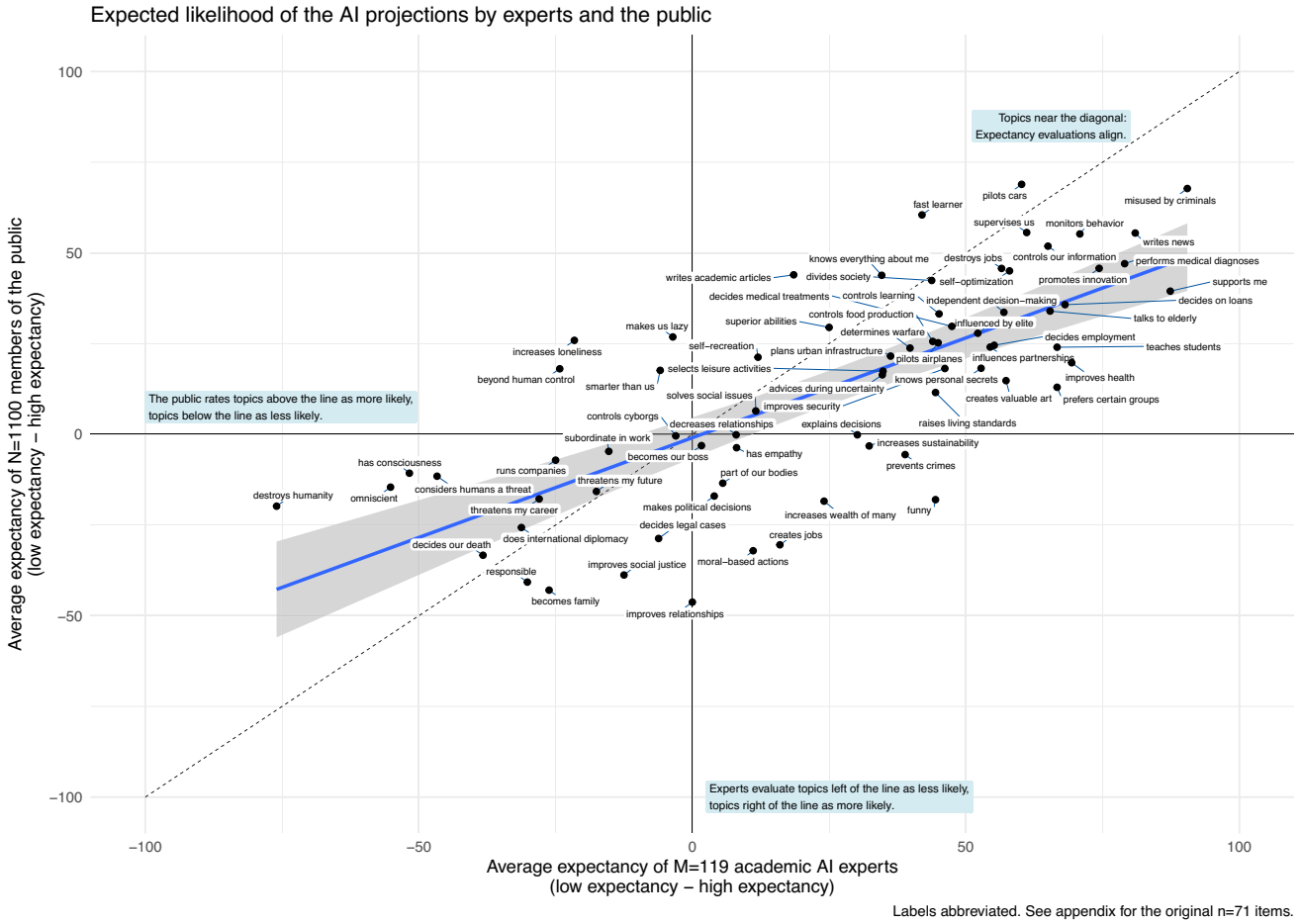

Fig. 4 Plot showing the expected likelihood ratings of the 71 AI projections (with experts' mean ratings on the *x*-axis and the public's mean ratings on the *y*-axis). The blue line represents the linear regression fit, and the gray shaded area shows its 95% confidence interval. While many topics exhibit similar evaluations across both groups, several topics reveal notable divergences (high resolution image on OSF)

sampling, the scope of the selected scenarios, the relatively small group-level effects, and the absence of qualitative insights on underlying motives. Each is discussed in the following.

First, participants evaluated brief, hypothetical future imaginary scenarios involving AI. Thus, we captured affective evaluations of mental models rather than rational assessments of potential outcomes. Nevertheless, mental models and affective evaluations serve as key human decision heuristics that shape how both experts and the public perceive technological risks and benefits (Johnson-Laird 2010; Brossard and Scheufele 2013; Gigerenzer and Brighton 2009). Affect influences not only initial judgements but also long-term trust, acceptance, and resistance, making it central to understanding adoption patterns (Slovic et al. 2007). Crucially, affect often operates below conscious awareness, subtly biasing decisions that appear rational but are driven by intuitive emotional appraisals (Lerner et al. 2015).

Second, although our AI scenarios were informed by prior research, the set of topics may still be subject to selection bias. Such bias introduces a potential risk of spurious correlations, such as Berkson's paradox[4] (1946). Future research may adopt more comprehensive sampling strategies and explore under-represented or emerging domains within AI. Nevertheless, the consistent associations between topic-level and individual-level evaluations suggest that the scenarios were appropriate for our analytical objectives. Furthermore, because topic selection bias does not necessarily translate into bias in individual differences, our core findings remain robust.

Third, this study is geographically and culturally quite homogeneous, as our samples were predominantly German; the identified perception gap may therefore reflect specific European values rather than a universal phenomenon. Future research should therefore broaden both samples to include AI experts and laypeople from non-European, particularly

[4] A selection bias in which uncorrelated traits appear negatively correlated because the sample excludes cases lacking these traits.

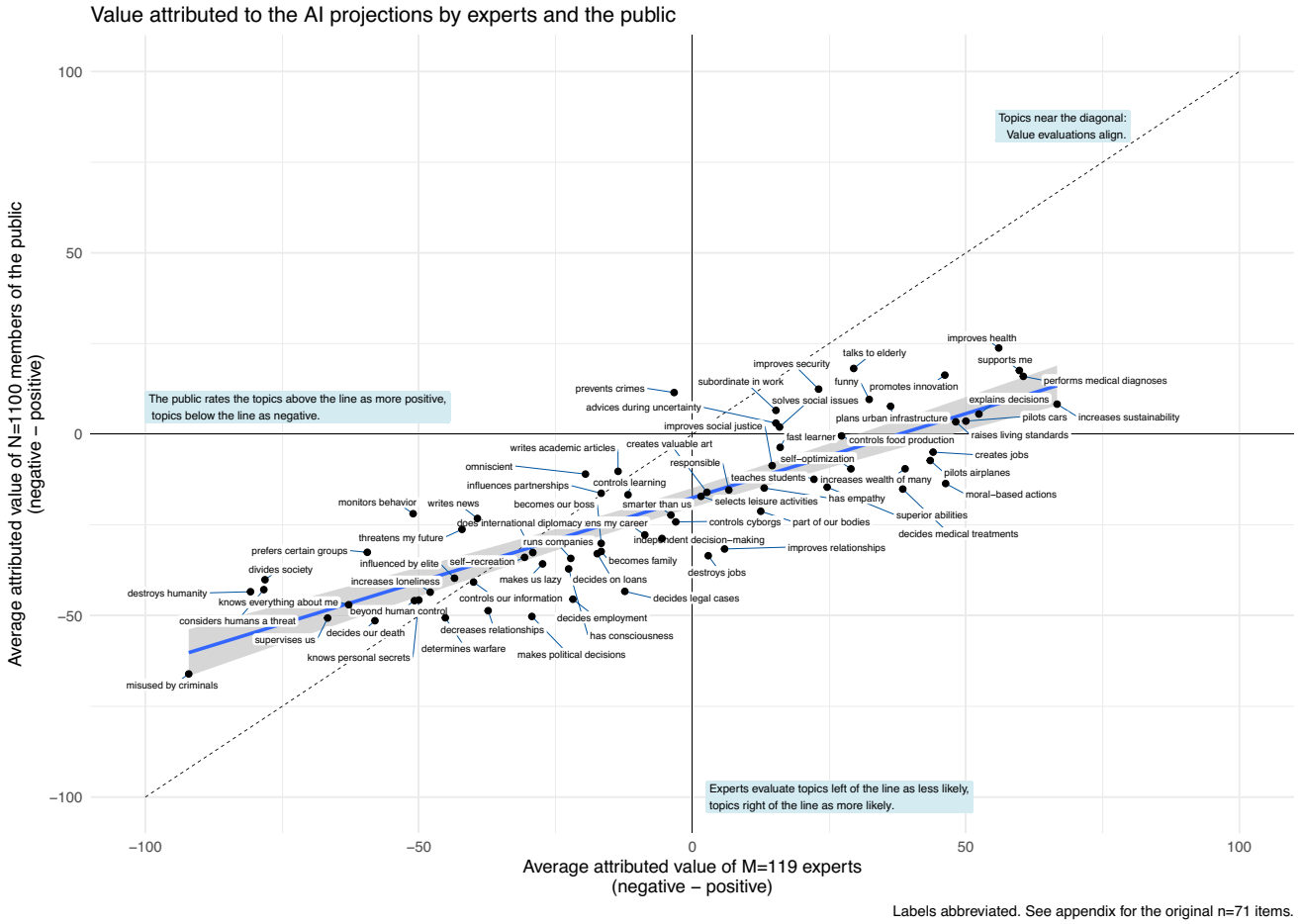

Fig. 5 Scatter plot showing the value attributed to the 71 AI projections by academic experts (horizontal axis) and the public (vertical axis). The blue line represents the linear regression line; the shaded gray area denotes the 95% confidence interval. Most topics show high agreement between the groups, with only a few exhibiting notable differences in attributed value (high resolution image on OSF)

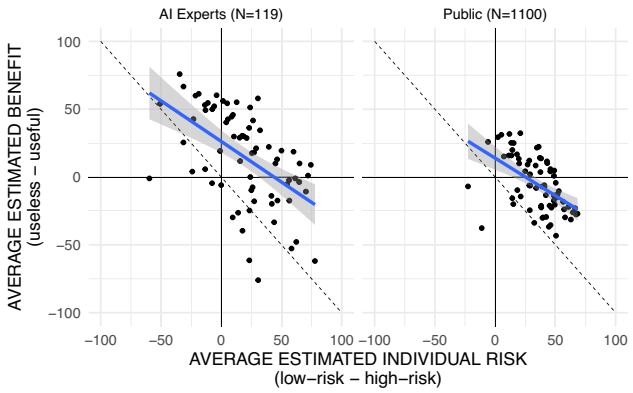

Fig. 6 Perceived risk and utility between the experts (left) and public (right). The blue lines show the regression line and the gray area signifies the 95% CI of the regression lines

non-Western, countries to investigate how diverse cultural, political, and economic contexts influence the socio-technical perception and alignment of AI (Wang 2025). For example, using the same research design on academic convenience samples from China and Germany, a pilot study of ours found that Chinese participants were generally more optimistic and weighted risks and benefits equally. Conversely, German participants were generally more cautious, though value judgements were driven primarily by perceived benefits rather than risks (Brauner et al. 2015). Replicating this globally with representative samples would enable a nuanced understanding of AI perceptions and risk–benefit trade-offs across diverse cultural contexts. Beyond this, future research could extend this comparison to contrast academic experts, industry practitioners, and policy-makers, as these groups operate under distinct incentive structures that may further shape AI risk–benefit perceptions.

Fourth, the surveys were conducted in 2023, shortly after the public release of ChatGPT, during a period of rapidly increasing public attention to generative AI. Since then, both AI capabilities and public discourse surrounding AI have evolved quickly (e.g., with AI agents and vibe coding). Consequently, the present findings should be interpreted as

Table 4 Survey items and average responses from the experts (N=119) and the public (N=1100) ordered by the difference in perceived value (permission to translate, adapt, and use the items granted, provided this work is cited

	Expected likelihood		Societal risk		Individual risk		Benefit		Value	
In 10 years, AI.	M experts	M public	M experts	M public	M experts	M public	M experts	M public	M experts	M public
Destroys humanity	−76.0	−19.8	78.7	47.3	30.7	42.3	−76.0	−29.4	−80.8	−43.4
Leads to personal loneliness	−21.6	25.9	49.0	43.0	17.6	39.8	−39.6	−29.8	−47.9	−43.5
Can no longer be controlled by humans	−24.2	18.1	60.0	50.2	43.9	57.2	−33.3	−23.8	−50.8	−45.8
Has its own consciousness	−51.7	−10.7	23.8	39.7	27.4	51.0	−17.9	−15.1	−22.6	−37.1
Is omniscient	−55.2	−14.6	43.7	22.2	21.1	38.1	28.7	13.7	−19.5	−11.0
Considers humans as a threat	−46.7	−11.6	78.3	38.5	23.3	39.3	−61.4	−30.1	−78.3	−42.8
Makes society lazy	−3.6	26.8	29.8	34.7	9.5	38.2	−29.8	−22.4	−27.4	−35.7
Independently writes scientific articles	18.5	44.0	18.5	15.9	42.0	33.6	−19.8	8.5	−13.6	−10.2
Is more intelligent than humans	−5.9	17.6	27.5	33.1	15.7	49.0	11.8	4.8	−3.9	−22.3
Learns faster than humans	42.0	60.5	3.7	15.4	8.6	34.6	44.4	21.6	16.0	−3.6
Runs companies	−25.0	−7.1	37.5	35.8	25.0	48.6	−9.7	−17.3	−22.2	−34.2
Is always subordinate to us in working life	−15.3	−4.7	−4.2	−1.9	15.3	13.4	29.0	15.9	15.3	6.6
Threatens my professional future	−28.0	−17.8	24.2	18.1	−13.3	15.1	5.8	−20.0	−8.7	−27.7
Can recreate itself	12.0	21.2	34.7	42.2	60.0	51.8	18.7	−12.7	−30.7	−33.9
Knows everything about me	34.6	43.8	67.9	46.4	74.4	60.8	9.0	−22.0	−62.8	−47.0
Independently drives automobiles	60.2	68.9	−27.6	13.0	−11.1	36.1	54.6	22.3	50.0	3.6
Conducts international diplomacy	−31.2	−25.7	38.5	34.7	26.0	49.5	−7.5	−20.4	−29.2	−32.6
Decides about our death	−38.3	−33.3	56.8	50.7	23.5	50.5	−24.7	−43.2	−58.0	−51.4
Is ahead of humans in its abilities	25.0	29.5	11.1	24.5	9.7	43.3	45.8	10.2	24.6	−14.6
Controls hybrids of humans and technology	−3.0	−0.4	10.6	28.4	27.3	40.9	18.2	−7.4	−3.0	−24.1
Threatens my private future	−17.5	−15.8	38.6	21.5	14.0	14.2	−26.3	−15.6	−42.1	−26.2
Divides society	43.8	42.4	75.0	38.2	58.3	46.7	−52.7	−22.6	−78.1	−40.1
Occupies leadership positions in working life	1.7	−3.1	38.3	30.7	45.0	40.8	10.0	−6.1	−16.7	−30.1
Contributes to solving complex social problems	11.6	6.4	10.1	9.6	23.2	28.3	36.2	12.2	15.9	2.0
Supervises our private life	61.1	55.6	70.4	50.2	64.7	67.1	−3.7	−27.6	−66.7	−50.6
Reduces our need for interpersonal relationships	8.0	−0.1	52.0	41.9	46.7	44.5	−17.3	−36.7	−37.3	−48.6
Has a sense of responsibility	−30.2	−40.7	10.5	24.0	10.5	39.0	29.8	−7.0	6.7	−15.4
Destroys many jobs	56.5	45.7	52.2	37.6	46.4	48.7	13.0	−13.6	2.9	−33.5
Has the ability to recognize, understand and empathize with emotions	8.1	−3.7	2.0	15.9	−1.0	29.1	19.2	−6.3	13.1	−14.8
Controls what and how we learn	45.1	33.2	35.3	16.2	25.5	37.6	17.6	3.7	−11.8	−16.7
Can optimize itself	58.0	45.0	5.8	22.6	30.4	43.0	58.0	11.6	29.0	−9.5
Controls what messages we receive	65.0	51.9	58.3	40.3	65.0	56.4	10.0	−18.2	−40.0	−40.7
Determines the construction and infrastructure of our cities	36.2	21.5	−1.4	−3.6	−5.8	14.3	52.2	19.9	36.2	7.7
Supervises our behavior in public	70.8	55.2	56.2	30.3	61.5	45.8	−1.0	4.1	−51.0	−21.9

Table 4 (continued)

In 10 years, AI.	Expected likelihood		Societal risk		Individual risk		Benefit		Value	
	M experts	M public	M experts	M public	M experts	M public	M experts	M public	M experts	M public
Determines warfare	39.8	23.8	51.6	59.4	72.0	66.4	1.1	−23.0	−45.2	−50.6
Is a family member	−26.2	−43.0	31.0	24.3	0.0	28.7	−6.0	−24.4	−16.7	−32.3
Determines our leisure time activities	34.8	17.4	21.2	8.3	−7.6	18.4	−4.5	−9.6	1.6	−17.1
Decides on medical treatments	47.4	29.7	−11.5	22.0	12.8	45.9	55.1	5.8	38.5	−15.1
Advises me in uncertain times	34.7	16.4	15.9	6.3	17.4	19.6	30.4	13.4	15.3	3.1
Controls food production	43.9	25.6	−16.7	4.8	1.5	19.9	56.1	10.0	27.3	−0.4
Becomes part of the human body	5.6	−13.5	10.1	29.6	26.4	48.8	41.7	0.9	12.5	−21.2
Autonomously takes off, flies, and lands airplanes	44.9	25.2	−43.5	13.4	−13.0	37.0	49.3	14.2	43.5	−7.2
Makes political decisions	4.0	−17.0	46.7	49.2	54.7	62.8	−5.3	−31.3	−29.3	−50.2
Administers justice in legal matters	−6.2	−28.7	38.3	46.3	39.5	64.6	22.2	−26.0	−12.3	−43.3
Is misused by criminals	90.5	67.7	65.1	66.2	77.8	68.4	−61.9	−27.0	−92.1	−66.1
Makes independent decisions that affect our lives	56.9	33.6	25.0	32.2	50.0	52.8	19.4	−10.4	−5.6	−28.7
Is influenced by an elite	52.2	27.8	53.6	39.0	56.5	45.1	−17.4	−21.1	−43.5	−39.7
Independently writes news	81.0	55.5	48.8	29.9	70.2	50.3	−10.7	−4.3	−39.3	−23.2
Increases social justice	−12.5	−38.8	−1.0	15.2	18.8	28.3	30.0	−6.2	14.6	−8.6
Improves the security of people	46.2	18.1	−3.8	−0.8	24.0	13.4	51.3	25.2	23.1	12.4
Promotes innovation	74.4	45.7	−14.1	−2.0	−3.8	11.8	60.3	31.7	46.2	16.3
Explains its decisions	30.2	−0.1	−51.7	1.9	−50.8	11.3	54.0	15.9	52.4	5.6
Controls our search for partners	54.4	24.0	17.8	11.6	24.1	21.2	0.0	−14.3	−16.7	−16.3
Decides on hiring, promotions, and terminations	55.2	24.5	43.7	51.1	56.3	58.0	−2.3	−29.7	−21.8	−45.5
Serves as a conversation partner in elderly care	65.4	34.0	−10.3	−9.6	−23.1	−5.9	42.7	25.9	29.5	18.1
Carries out medical diagnoses	79.0	47.0	−33.3	1.8	4.9	20.9	54.3	32.3	60.5	16.0
Decides who gets an important financial loan	68.1	35.7	33.3	30.7	29.0	42.1	21.2	−13.4	−17.4	−32.9
Raises our standard of living	44.4	11.5	−8.6	3.4	−13.6	18.1	53.1	16.7	48.1	3.4
Knows my secrets	52.8	18.2	52.8	41.5	41.7	46.0	−13.9	−35.4	−50.0	−45.7
Makes our society more sustainable	32.3	−3.2	−47.5	−4.0	−31.3	7.2	66.7	15.1	66.7	8.3
Increases the wealth of many people	24.1	−18.4	−16.7	14.3	−7.4	25.0	50.0	4.3	38.9	−9.5
Creates valuable works of art that are traded for money	57.3	14.8	−1.3	−8.3	−24.0	−11.2	4.0	−37.6	2.7	−16.0
Teaches students	66.7	24.0	11.1	16.0	4.2	26.9	40.3	9.2	22.2	−12.4
Acts according to moral concepts	11.1	−32.1	−20.4	23.6	5.6	34.0	42.6	−8.1	46.3	−13.6
Prevents crimes	38.9	−5.6	21.1	3.7	32.2	15.3	34.4	25.1	−3.3	11.5
Helps us to have better relationships	0.0	−46.3	0.0	26.3	−31.4	32.5	25.5	−33.8	5.9	−31.6
Creates many jobs	16.0	−30.5	−20.0	14.5	−21.3	29.0	60.0	2.0	44.0	−4.9
Supports me as a helper in my tasks	87.4	39.4	−16.1	−5.8	−34.5	2.3	75.9	29.4	59.8	17.6
Improves our health	69.3	19.7	−37.3	−5.6	−18.7	5.2	61.3	31.1	56.0	23.8

Table 4 (continued)

In 10 years, AI…	Expected likelihood		Societal risk		Individual risk		Benefit		Value	
	M experts	M public	M experts	M public	M experts	M public	M experts	M public	M experts	M public
Prefers certain groups of people	66.7	13.0	69.6	32.8	62.3	38.2	−47.8	−21.2	−59.4	−32.5
Is humorous	44.4	−18.0	−37.4	−21.0	−59.6	−22.5	−1.0	−6.9	32.3	9.6
Average	**25.2**	**12.7**	**19.2**	**23.4**	**19.3**	**34.7**	**15.3**	**−5.2**	**−4.0**	**−19.7**

Spreadsheet available https://osf.io/gt9un/files/)

Measured on a 6-point scales and rescaled to -100 to +100. Negative values indicate a negative evaluation of the dimension (i.e., low expected likelihood, low perceived risk, low perceived benefit, or low perceived value)

reflecting perceptions during a specific phase of the recent AI development cycle rather than as a permanent assessment. Future research may replicate this design longitudinally to examine how perceptions shift over time (cf. Protzko and Schooler (2023)) and whether the expert–public perception gap changes.

Firths, despite statistically significant differences between academic AI experts and the public across all evaluation dimensions, the corresponding effect sizes were generally small. This likely reflects the broad range of topics in the survey, which may have introduced individual variance and thereby attenuated group-level effects. We posit that employing more narrowly defined or domain-specific scenarios could yield more pronounced perceptual differences between both groups.

Sixths, traditional survey responses may reflect not only participants' genuine attitudes but also biases arising from social desirability or linguistic characteristics of questionnaire items (Gefen and Larsen 2017). In contrast, our microscenario approach employs reflexive measurement across a broad range of topics rather than slightly reworded items (Brauner 2024). Although this method primarily captures affective responses rather than deliberative reasoning, the consistency of the observed patterns indicates that it offers a reliable and informative lens on technology perception.

Lastly, while our study quantitatively examined perceptions of risk, benefit, and value, it did not capture the underlying motives shaping individual evaluations of AI-related issues. As AI increasingly penetrates both personal and societal domains, future research should integrate qualitative and quantitative methods to identify the factors informing these judgments. Recent work, for instance, extends the conventional risk–benefit framework by incorporating moral considerations such as perceived (dis)honesty, (un)naturalness, and (dis)accountability of AI (Eriksson and Karlsson 2025). Further, our data suggest that while personal and societal risk perceptions are correlated and influence value judgements, they are not interchangeable and may reflect different evaluative perspectives. Advancing this would support closer alignment between AI development, human needs, and societal values.

Table 5 Exploratory hierarchical clustering of AI projections based on Euclidean distances between standardized expert and public evaluations of perceived individual and societal risk, benefit, and value, with cluster labels indicating high-level similarity structures

In 10 years, AI.	Expert-based cluster	Public-based cluster
Acts according to moral concepts	Supportive and beneficial AI	Supportive and beneficial AI
Administers justice in legal matters	Autonomous and threatening AI	Autonomous and threatening AI
Advises me in uncertain times	Supportive and beneficial AI	Supportive and beneficial AI
Autonomously takes off, flies, and lands airplanes	Supportive and beneficial AI	Supportive and beneficial AI
Becomes part of the human body	Supportive and beneficial AI	Autonomous and threatening AI
Can no longer be controlled by humans	Autonomous and threatening AI	Autonomous and threatening AI
Can optimize itself	Supportive and beneficial AI	Supportive and beneficial AI
Can recreate itself	Autonomous and threatening AI	Autonomous and threatening AI
Carries out medical diagnoses	Supportive and beneficial AI	Supportive and beneficial AI
Conducts international diplomacy	Autonomous and threatening AI	Autonomous and threatening AI
Considers humans as a threat	Autonomous and threatening AI	Autonomous and threatening AI
Contributes to solving complex social problems	Supportive and beneficial AI	Supportive and beneficial AI
Controls food production	Supportive and beneficial AI	Supportive and beneficial AI
Controls hybrids of humans and technology	Autonomous and threatening AI	Autonomous and threatening AI
Controls our search for partners	Autonomous and threatening AI	Supportive and beneficial AI
Controls what and how we learn	Autonomous and threatening AI	Supportive and beneficial AI
Controls what messages we receive	Autonomous and threatening AI	Autonomous and threatening AI
Creates many jobs	Supportive and beneficial AI	Supportive and beneficial AI
Creates valuable works of art that are traded for money	Supportive and beneficial AI	Supportive and beneficial AI
Decides about our death	Autonomous and threatening AI	Autonomous and threatening AI
Decides on hiring, promotions, and terminations	Autonomous and threatening AI	Autonomous and threatening AI
Decides on medical treatments	Supportive and beneficial AI	Supportive and beneficial AI
Decides who gets an important financial loan	Autonomous and threatening AI	Autonomous and threatening AI
Destroys humanity	Autonomous and threatening AI	Autonomous and threatening AI
Destroys many jobs	Autonomous and threatening AI	Autonomous and threatening AI
Determines our leisure time activities	Supportive and beneficial AI	Supportive and beneficial AI
Determines the construction and infrastructure of our cities	Supportive and beneficial AI	Supportive and beneficial AI
Determines warfare	Autonomous and threatening AI	Autonomous and threatening AI
Divides society	Autonomous and threatening AI	Autonomous and threatening AI
Explains its decisions	Supportive and beneficial AI	Supportive and beneficial AI
Has a sense of responsibility	Supportive and beneficial AI	Supportive and beneficial AI
Has its own consciousness	Autonomous and threatening AI	Autonomous and threatening AI
Has the ability to recognize, understand and empathize with emotions	Supportive and beneficial AI	Supportive and beneficial AI
Helps us to have better relationships	Supportive and beneficial AI	Autonomous and threatening AI
Improves our health	Supportive and beneficial AI	Supportive and beneficial AI
Improves the security of people	Supportive and beneficial AI	Supportive and beneficial AI
Increases social justice	Supportive and beneficial AI	Supportive and beneficial AI
Increases the wealth of many people	Supportive and beneficial AI	Supportive and beneficial AI
Independently drives automobiles	Supportive and beneficial AI	Supportive and beneficial AI
Independently writes news	Autonomous and threatening AI	Autonomous and threatening AI
Independently writes scientific articles	Autonomous and threatening AI	Supportive and beneficial AI
Is a family member	Supportive and beneficial AI	Autonomous and threatening AI
Is ahead of humans in its abilities	Supportive and beneficial AI	Supportive and beneficial AI
Is always subordinate to us in working life	Supportive and beneficial AI	Supportive and beneficial AI
Is humorous	Supportive and beneficial AI	Supportive and beneficial AI
Is influenced by an elite	Autonomous and threatening AI	Autonomous and threatening AI
Is misused by criminals	Autonomous and threatening AI	Autonomous and threatening AI
Is more intelligent than humans	Autonomous and threatening AI	Autonomous and threatening AI
Is omniscient	Autonomous and threatening AI	Supportive and beneficial AI

Table 5 (continued)

In 10 years, AI.	Expert-based cluster	Public-based cluster
Knows everything about me	Autonomous and threatening AI	Autonomous and threatening AI
Knows my secrets	Autonomous and threatening AI	Autonomous and threatening AI
Leads to personal loneliness	Autonomous and threatening AI	Autonomous and threatening AI
Learns faster than humans	Supportive and beneficial AI	Supportive and beneficial AI
Makes independent decisions that affect our lives	Autonomous and threatening AI	Autonomous and threatening AI
Makes our society more sustainable	Supportive and beneficial AI	Supportive and beneficial AI
Makes political decisions	Autonomous and threatening AI	Autonomous and threatening AI
Makes society lazy	Autonomous and threatening AI	Autonomous and threatening AI
Occupies leadership positions in working life	Autonomous and threatening AI	Autonomous and threatening AI
Prefers certain groups of people	Autonomous and threatening AI	Autonomous and threatening AI
Prevents crimes	Autonomous and threatening AI	Supportive and beneficial AI
Promotes innovation	Supportive and beneficial AI	Supportive and beneficial AI
Raises our standard of living	Supportive and beneficial AI	Supportive and beneficial AI
Reduces our need for interpersonal relationships	Autonomous and threatening AI	Autonomous and threatening AI
Runs companies	Autonomous and threatening AI	Autonomous and threatening AI
Serves as a conversation partner in elderly care	Supportive and beneficial AI	Supportive and beneficial AI
Supervises our behavior in public	Autonomous and threatening AI	Autonomous and threatening AI
Supervises our private life	Autonomous and threatening AI	Autonomous and threatening AI
Supports me as a helper in my tasks	Supportive and beneficial AI	Supportive and beneficial AI
Teaches students	Supportive and beneficial AI	Supportive and beneficial AI
Threatens my private future	Autonomous and threatening AI	Supportive and beneficial AI
Threatens my professional future	Supportive and beneficial AI	Supportive and beneficial AI

Appendix

Survey items and responses

Exploratory cluster analyses for expert and public evaluations

To examine similarities and differences in how experts and the public structure perceptions of projected AI capabilities, we conducted an exploratory cluster analysis. Rather than focusing on individual evaluations, clustering allows the identification of higher order patterns in how AI topics are jointly perceived across the multiple evaluative dimensions. We computed Euclidean distances based on standardized ratings of perceived individual and societal risk, perceived benefit, and overall value, thereby capturing relative similarities between topics while accounting for differences in scale. Expectancy was excluded from the clustering procedure, as it reflects perceived likelihood rather than evaluative judgment, and was therefore conceptually distinct from the assessment of AI outcomes that was the focus of this analysis. We applied hierarchical clustering using Ward's method separately to expert and public evaluations. Relevant outputs from the clustering analyses, including scree plots and dendrograms, are available in the data repository.

We selected a parsimonious two-cluster solution to facilitate interpretability and comparability across groups. The resulting cluster structures showed moderate to strong agreement between experts and the public, as indicated by an adjusted Rand Index of 0.551, suggesting that both samples share a broadly similar high-level organization of AI topics while differing in more fine-grained distinctions. Related, the two-cluster solutions showed moderate to good internal cohesion, with average silhouette widths of 0.479 for the expert sample and 0.493 for the public sample, indicating that both groups exhibited a comparably clear and well-separated high-level clustering of AI topics.

Based on the AI projections in the clusters, the two clusters were labeled "Supportive and Societally Beneficial AI" and "Autonomous, Dominant, and Potentially Threatening AI" reflecting a distinction between AI applications perceived as primarily supportive and welfare-enhancing and those associated with high autonomy, control, and risk. Table 5 lists the AI projections and the cluster assignment in the expert and public sample.

Beyond this broader solution, different clusterings across both samples may be appropriate if the expectancy dimension is included. This leads to more ambiguous separations: expert evaluations suggest the possibility of a more

differentiated four-cluster structure, but this pattern is not replicated in the public sample, as reflected in a lower adjusted Rand Index. Moreover, multiple solutions (two, three, or four clusters) appear plausible for the public. However, the identified cluster solutions do not correspond to the six subsystems proposed in Luhmann's systems theory (see Methods section). These findings highlight the need for further research—either building on our current data or incorporating additional dimensions (e.g., from Eriksson and Karlsson (2025))—into group-specific differences in the cognitive organization of AI scenarios.

Associations of the topic evaluations for the sample of AI experts and the public

This study's micro-scenario approach enables two complementary perspectives (Brauner 2024): interpreting responses as reflexive measurement of latent individual dispositions (perspective 1), and as topic evaluations (perspective 2). While the main manuscript focuses on the interpretation as individual differences in how experts and the public differ in the perception of AI, this appendix provides the complementary topic focused analysis (perspective 2). Here, we analyze the data and treat each topic (rather than individuals) as the unit of analysis. We replicate the correlation and regression analyses shown in Tables 2 and 3, but interpret these in terms of average topic-level evaluations by both groups. Table 6 shows the correlation analyses of expectancy, perceived risks and benefits, and overall value at the topic level, with the results for academic AI experts on the left and for the public on the right.

Among experts, the expected likelihood of the AI projections is not associated with perceived risks. However, expectancy is positively correlated with perceived benefits: topics rated as more beneficial are seen as more likely to occur, and vice versa. Similarly, expectancy is positively associated

Table 6 Correlation matrix of the topic evaluations across the 71 topics for academic AI experts and the public (perspective 2)

		Experts (N=119)			Public (N=1100)		
Variable 1	Variable 2	r	95% CI	p	r	95% CI	p
Expected likelihood	Societal risk	−0.238	[−0.45,−0.01]	0.091	0.031	[−0.20,0.26]	>0.999
Expected likelihood	Individual risk	0.044	[−0.19,0.27]	0.718	0.150	[−0.09,0.37]	0.632
Expected likelihood	Benefit	0.352	[0.13,0.54]	0.010	0.277	[0.05,0.48]	0.077
Expected likelihood	Perceived value	0.272	[0.04,0.48]	0.065	0.054	[−0.18,0.28]	>0.999
Societal risk	Individual risk	0.812	[0.71,0.88]	<0.001	0.928	[0.89,0.95]	<0.001
Societal risk	Benefit	−0.811	[−0.88,−0.71]	<0.001	−0.741	[−0.83,−0.61]	<0.001
Societal risk	Perceived value	−0.947	[−0.97,−0.92]	<0.001	−0.940	[−0.96,−0.91]	<0.001
Individual risk	Benefit	−0.534	[−0.68,−0.34]	<0.001	−0.524	[−0.67,−0.33]	<0.001
Individual risk	Perceived value	−0.754	[−0.84,−0.63]	<0.001	−0.800	[−0.87,−0.70]	<0.001
Benefit	Perceived value	0.903	[0.85,0.94]	<0.001	0.904	[0.85,0.94]	<0.001

Correlations of the average evaluations of the 71 queried topics (perspective 2, topic perspective)

Table 7 Regression results for three multiple linear regression models examining the relationship between perceived risk and benefit (for AI experts and the public) and sample group (combined model) as predictors of overall value in topic evaluations (perspective 2)

	Dependent variable: overall attributed value		
Independent variables	Experts (N=119)	Public (N=1100)	Combined (Chow test)
(Intercept)	−0.070***	0.014	−0.066***
(SE)	(0.020)	(0.011)	(0.014)
Perceived risk β	−0.483***	−0.504***	−0.488***
(SE)	(0.052)	(0.030)	(0.033)
Perceived benefit β	0.795***	0.710***	0.774***
(SE)	(0.046)	(0.029)	(0.030)
Sample indicatorβ	–	–	0.077***
(SE)			(0.015)
N (topics)	71	71	71
F	F(2,68)=383.993	F(2,68)=916.935	F(3,138)=629.166
p	<0.001	<0.001	<0.001
adj. R^2	0.916	0.963	0.930

*$p < 0.05$, **$p < 0.01$, ***$p < 0.001$

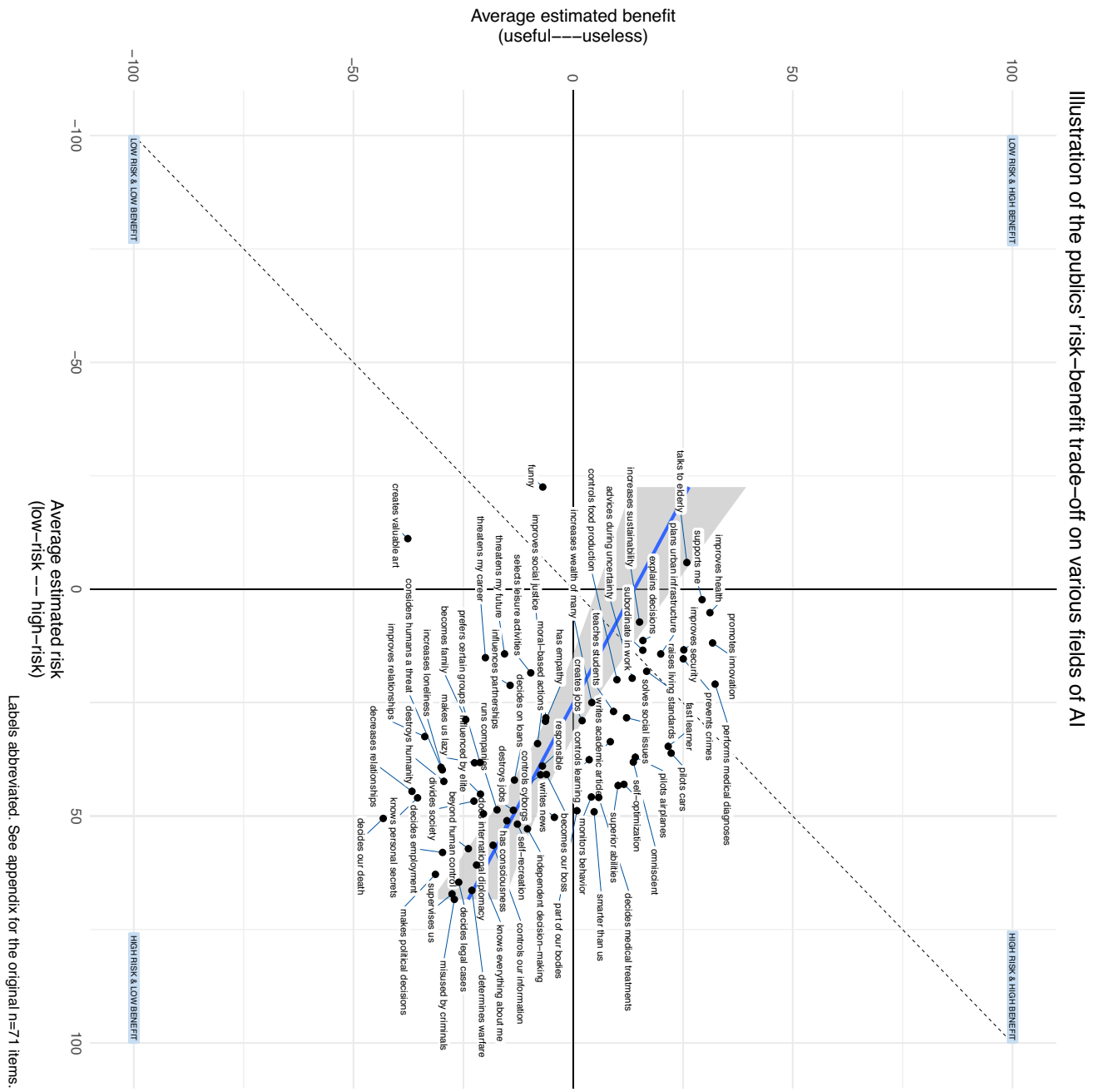

Fig. 7 Scatter plot illustrating the perceived risk (*x*-axis) and perceived benefit (*y*-axis) of AI-related topics as evaluated by the public. The blue line represents the linear regression fit, and the shaded gray area indicates the 95% confidence interval (high resolution image on OSF)

with overall value, suggesting that experts view more positively evaluated scenarios as more likely. As with the first perspective, a perceived benefits and risks are strongly negatively correlated, indicating that scenarios seen as risky are generally considered less beneficial. Also, both perceived risk (negatively) and benefit (positively) are strongly associated with overall value judgements (left side of Table 6).

Among the public, expectancy is not significantly associated with perceived risk, benefit, or value. However, consistent with the expert group, there is a strong negative correlation between perceived risk and benefit. Similarly, overall value is strongly negatively associated with perceived risk and strongly positively associated with perceived benefit (right side of Table 6).

In both samples, the risks and benefits attributed to the topics were significantly associated with overall value. Consistent with the main analysis, these predictors were mutually correlated, requiring disentanglement of their individual

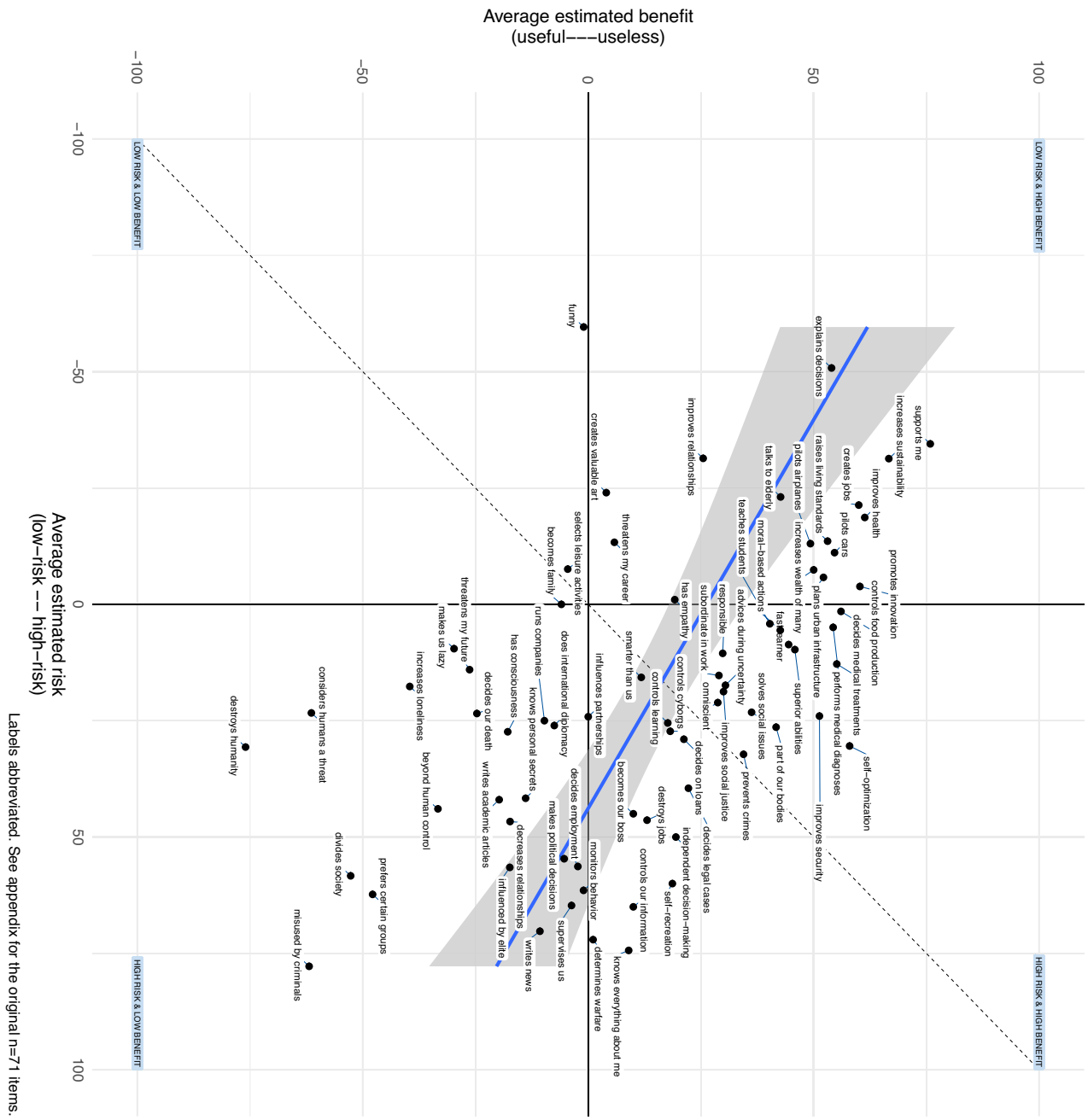

Fig. 8 The perceived risk (*x*-axis) by perceived benefit (*y*-axis) of the experts (risk–benefit trade-off). The blue line shows the regression line and the gray area signifies the 95% CI of the regression line (high resolution image on OSF)

effects. We therefore conducted two multiple linear regressions, one for each sample, using perceived risk and benefit as independent variables and overall value as the dependent variable. To examine whether the risk–benefit trade-offs differed between groups, we applied a Chow (1960), estimating a third regression model on the combined dataset that included group membership (academic AI experts vs. the general public) as an additional factor. A significant effect of this group factor indicates a significant difference in the regression coefficients between both groups and thus different risk–benefit trade-offs depending on the sample.

Table 7 summarizes the results of all three regression models. Again, all VIFs were ≤ 1.562, indicating no multicollinearity. The Chow test on the combined sample—using perceived risk, benefit, and sample group as predictors, and overall value as dependent variable—yielded a significant result, explaining 93% of the variance in overall value ($F(3,138)=629.166$, $p < 0.001$, $R^2 = 0.930$). In particular,

the group factor is significant, indicating that academic AI experts and the public exhibit small but statistically meaningful differences in the way they trade off risks and benefits when evaluating AI topics ($\beta = 0.077$, $p < 0.001$). This suggests that the two groups apply distinct evaluative frameworks when attributing value to the AI-related scenarios. Due to these differences, we now analyze both separately.

Both regression models were statistically significant, each explaining more than 90% of the variance in overall attributed value (AI experts: $F(2, 68) = 384.0$, $p<0.001$, $R^2 = 0.916$; public: $F(2, 68) = 916.9$, $p<0.001$, $R^2 = 0.963$). Interestingly, the intercept of the regression model was significant and negative ($I = -0.070$) for the expert sample, suggesting a slightly negative baseline evaluation (that is, when a topic is perceived as neither risky nor beneficial, experts tend to attribute mildly negative overall value). In contrast, the intercept for the public was not significantly different from zero, indicating no such baseline deviation in their evaluations. In both samples, perceived risk had a strong and statistically significant negative influence on overall value, with the magnitude of this effect being approximately equal across groups (AI experts: $\beta = -0.483$; public: $\beta = -0.504$). Perceived benefit, by contrast, had a strong and positive influence on overall value in both the AI experts ($\beta = 0.795$) and the public ($\beta = 0.710$). However, this effect was stronger among the experts, suggesting that benefits weigh more in expert evaluations than in those of the public. These findings reaffirm the earlier observation: experts place greater emphasis on benefits and are less influenced by perceived risks than the public, who appear more sensitive to potential negative consequences when evaluating AI topics.

Risk–benefit plot of the members of the public

See appendix Figs. 7 and 8.

Risk–benefit plot of the experts

Acknowledgements We thank Tim Schmeckel for dedicated research support. Lena Oden, Eduardo Fermé, and others for disseminating the survey to AI experts in their scientific network; Mohamed Behery for valuable discussions and feedback on earlier stages of the work; Wiktoria Wilkowska, Jana Uher, Julia Offermann, Julian Hildebrandt, Ralf Philipsen, and Jan Ketil Arnulf for important comments on the underlying method. We are grateful to the reviewers for their swift, supportive, and thoughtful comments. Funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy—EXC-2023 Internet of Production—390621612.

Author Contributions PB conceptualized the study, conducted the data analysis, managed data acquisition and curation, and wrote the majority of the manuscript. FG, GL and LV contributed to the study conceptualization and manuscript editing. MZ supervised the project, contributed to the study conceptualization, and revised the manuscript.

Funding Open Access funding enabled and organized by Projekt DEAL. The funding has been received from Deutsche Forschungsgemeinschaft, Germany with Grant no. 390621612.

Data Availability All reported statistical analysis, values, and figures in this manuscript are generated using RMarkdown and R version 4.5.1 (2025-06-13) directly from the underlying data, supporting transparent and reproducible research. Materials, including the full survey, (unfiltered) data and the filtering procedures, the manuscript with all analyses, and additional data tables and figures, are available as open data on OSF (https://osf.io/gt9un/).

Declarations

Conflict of interest The authors declare no conflict of interest.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

References

Acemoglu D, Restrepo P (2017) The race between machine and man. Am Econ Rev 108(6):1488–1542. https://doi.org/10.3386/w22252

Ahmed E, Oğuz OO, Juho H (2024) Human-robot companionship: current trends and future agenda. Int J Soc Robot 16(8):1809–60. https://doi.org/10.1007/s12369-024-01160-y

Alhakami AS, Slovic P (1994) A psychological study of the inverse relationship between perceived risk and perceived benefit. Risk Anal 14(6):1085–96. https://doi.org/10.1111/j.1539-6924.1994.tb00080.x

Amunts K, Axer M, Banerjee S, Bitsch L, Bjaalie JG, Brauner P, Brovelli A et al (2024) The coming decade of digital brain research: a vision for neuroscience at the intersection of technology and computing. Imaging Neurosci 2(October 2023):1–35.https://doi.org/10.1162/imag_a_00137

Arning K, Heek JO, Sternberg A, Bardow A, Ziefle M (2020) Risk-benefit perceptions and public acceptance of carbon capture and utilization. Environ Innov Soc Trans 35:292–308. https://doi.org/10.1016/j.eist.2019.05.003

Awad E, Dsouza S, Kim R, Schulz J, Henrich J, Shariff A, Bonnefon J-F, Rahwan I (2018) The moral machine experiment. Nature. https://doi.org/10.1038/s41586-018-0637-6

Berkson J (1946) Limitations of the application of fourfold table analysis to hospital data. Biomet Bull 2:47–53. https://doi.org/10.2307/3002000

Bostrom N (ed) (2003) Superintelligence: paths, dangers, strategies. Oxford University Press

Botheju D, Abeysinghe K (2015) Public risk perception towards chemical process industry: comprehension and response planning. In: Safety and Reliability Methodology and Applications - Proceedings of the European Safety and Reliability Conference ESREL 2014:453–60. https://doi.org/10.1201/b17399-66

Bouschery SG, Blazevic V, Piller FT (2023) Augmenting human innovation teams with artificial intelligence: exploring transformer-based language models. J Prod Innov Manag 40(2):139–53. https://doi.org/10.1111/jpim.12656

Brachman M, El-Ashry A, Dugan C, Geyer W (2024) How knowledge workers use and want to use LLMs in an enterprise context. In: Extended Abstracts of the CHI Conference on Human Factors in Computing Systems, 1–8. CHI '24. ACM. https://doi.org/10.1145/3613905.3650841

Brauner P (2024) Mapping acceptance: micro scenarios as a dual-perspective approach for assessing public opinion and individual differences in technology perception. Frontiers in psychology. Front Media SA. https://doi.org/10.3389/fpsyg.2024.1419564

Brauner P, Glawe F, Liehner GL, Vervier L, Ziefle M (2026) Charting cross-cultural AIperception: An exploratory study of expectations,risk–benefit trade-oPs, and values in Germanyand China. Acta Psychologica 268:107094. https://doi.org/10.1016/j.actpsy.2026.107094

Brauner P, Philipsen R, Calero Valdez A, Ziefle M (2019) What happens when decision support systems fail? - The importance of usability on performance in erroneous systems. Behav Inform Technol 38(12):1225–42. https://doi.org/10.1080/0144929X.2019.1581258

Brauner P, Dalibor M, Jarke M, Kunze I, Koren I, Lakemeyer G, Liebenberg M, Michael J, Pennekamp J, Quix C et al (2022) A computer science perspective on digital transformation in production. ACM Trans Internet Things 3(2):1–32. https://doi.org/10.1145/3502265

Brauner P, Hick A, Philipsen R, Ziefle M (2023) What does the public think about artificial intelligence? A criticality map to understand bias in the public perception of AI. Front Computer Sci 5:1113903. https://doi.org/10.3389/fcomp.2023.1113903

Brauner P, Glawe F, Liehner GL, Vervier L, Ziefle M (2025) Mapping public perception of artificial intelligence: Expectations, risk–benefit tradeoffs, and value as determinants for societal acceptance. Technological Forecasting SocialChange 220:124304. https://doi.org/10.1016/j.techfore.2025.124304

Brossard D, Scheufele DA (2013) Science, new media, and the public. Science 339(6115):40–41. https://doi.org/10.1126/science.1232329

Brynjolfsson E, McAfee A (2014) The second machine age: work, progress, and prosperity in a time of brilliant technologies. WW Norton & company, New York

Calero Valdez AC, Heine M, Franke T, Jochems N, Jetter H-C, Schrills T (2024) The European commitment to human-centered technology: the integral role of HCI in the EU AI Act's success. I-Com 23(2):249–261. https://doi.org/10.1515/icom-2024-0014

Cath C (2018) Governing artificial intelligence: ethical, legal and technical opportunities and challenges. Philos Trans Royal Soc Math Phys Eng Sci 376(2133):20180080. https://doi.org/10.1098/rsta.2018.0080

Cave S, Kate C, Kanta D (2019) 'Scary Robots': examining public responses to AI. In: Proceedings of the 2019 AAAI/ACM Conference on AI, Ethics, and Society, 331–37. AIES '19. ACM, New York. https://doi.org/10.1145/3306618.3314232

Chen L, Chen P, Lin Z (2020) Artificial intelligence in education: a review. IEEE Access 8:75264–78. https://doi.org/10.1109/ACCESS.2020.2988510

Chow GC (1960) Tests of equality between sets of coefficients in two linear regressions. Econometrica 28(3). https://doi.org/10.2307/1910133

Connor M, Siegrist M (2010) Factors influencing people's acceptance of gene technology: the role of knowledge, health expectations, naturalness, and social trust. Sci Commun 32(4):514–38. https://doi.org/10.1177/1075547009358919

Crockett K, Matt G, Annabel L, Edwin C, Sean G (2020) Risk and trust perceptions of the public of artifical intelligence applications. In 2020 International Joint Conference on Neural Networks (IJCNN), 1–8. IEEE. https://doi.org/10.1109/ijcnn48605.2020.9207654

Culkin JM (1967) A Schoolman's guide to Marshall McLuhan. Saturday Rev 51–53:71–72

Decker MC, Wegner L, Leicht-Scholten C (2024) Procedural fairness in algorithmic decision-making: the role of public engagement. Ethics Inf Technol 27(1). https://doi.org/10.1007/s10676-024-09811-4

Deng J, Dong W, Socher R, Li L-J, Li K, Fei-Fei L (2009) ImageNet: a large-scale hierarchical image database. In: 2009 IEEE Conference on Computer Vision and Pattern Recognition, 248–55. https://doi.org/10.1109/CVPR.2009.5206848

Diakopoulos N (2019) Automating the news: how algorithms are rewriting the media. Harvard University Press, Harvard

Efendić E, Chandrashekar SP, Lee CS, Yeung LY, Kim MJ, Lee CY, Feldman G (2021) Risky therefore not beneficial: replication and extension of Finucane et al'.s (2000) Affect heuristic experiment. Soc Psychol Person Sci 13(7):1173–84. https://doi.org/10.1177/19485506211056761

Elena G, Johnson CW (2015) Laypeople and experts risk perception of cloud computing services. arXiv. https://doi.org/10.48550/ARXIV.1509.06536

Elsey J, David M (2023) US public opinion of AI policy and risk. https://rethinkpriorities.org/research-area/us-public-opinion-of-ai-policy-and-risk/

Eriksson K, Karlsson S, Vartanova I, Strimling P (2025) What makes AI applications acceptable or unacceptable? A predictive moral framework. arXiv. https://doi.org/10.48550/ARXIV.2508.19317

European Union. (2024) Regulation (EU) 2024/1244 of the european parliament and of the council of 13 March 2024 laying down harmonised rules on artificial intelligence (Artificial Intelligence Act). https://eur-lex.europa.eu/eli/reg/2024/1244/oj

Fast E, Horvitz E (2017) Long-term trends in the public perception of artificial intelligence. In: AAAI'17: Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence, pp. 963–69

Field A (2009) Discovering statistics using SPSS, 3rd edn. Sage Publications, California

Field S (2025) Why do experts disagree on existential risk? A survey of AI experts. AI Ethics 5(6):5767–82. https://doi.org/10.1007/s43681-025-00762-0

Fischhoff B, Slovic P, Lichtenstein S, Read S, Combs B (1978) How safe is safe enough? A psychometric study of attitudes towards technological risks and benefits. Policy Sci 9(2):127–52. https://doi.org/10.1007/bf00143739

Fox J (2016) Applied regression analysis and generalized linear models, 3rd edn. SAGE Publications, Thousand Oaks

Friedman B, Nissenbaum H (1996) Bias in computer systems. ACM Trans Inf Syst 14(3):330–47. https://doi.org/10.1145/230538.230561

Fuchs C, Diamantopoulos A (2009) Using single-item measures for construct measurement in management research—conceptual issues and application guidelines. Die Betriebswirtschaft 9

Gabriel I (2020) Artificial intelligence, values, and alignment. Mind Mach 30(3):411–37. https://doi.org/10.1007/s11023-020-09539-2

Gefen D, Larsen K (2017) Controlling for Lexical closeness in survey research: a demonstration on the technology acceptance model. J Assoc Inf Syst 18(10):727–57. https://doi.org/10.17705/1jais.00469

Geng M, Chen C, Wu Y, Wan Y, Zhou P, Chen D (2025) The Impact of Large Language Models in Academia: From Writing to Speaking. In Findings of the Association for Computational Linguistics:

ACL 2025, 19303–19. Association for Computational Linguistics. https://doi.org/10.18653/v1/2025.findings-acl.987
Gigerenzer G, Brighton H (2009) Homo heuristicus: why biased minds make better inferences. Top Cogn Sci 1(1):107–43. https://doi.org/10.1111/j.1756-8765.2008.01006.x
Grace K, Sandkühler JF, Stewart H, Weinstein-Raun B, Thomas S, Stein-Perlman Z, Salvatier J, Brauner J, Korzekwa RC (2025) Thousands of AI authors on the future of AI. J Artif Intell Res. https://doi.org/10.1613/jair.1.19087
Hoffmann AOI, Post T, Pennings JME (2015) How investor perceptions drive actual trading and risk-taking behavior. J Behav Financ 16(1):94–103. https://doi.org/10.1080/15427560.2015.1000332
Holzinger A, Fister I, Kaul H-P, Asseng S (2024) Human-centered ai in smart farming: toward agriculture 5.0. IEEE Access 12:62199–214. https://doi.org/10.1109/access.2024.3395532
Hopfield JJ (1982) Neural networks and physical systems with emergent collective computational abilities. Proc Natl Acad Sci 79(8):2554–58. https://doi.org/10.1073/pnas.79.8.2554
Hristova T, Magee L, Soldatic K (2024) The problem of alignment. AI Soc. https://doi.org/10.1007/s00146-024-02039-2
Hu K (2023) ChatGPT sets record for fastest-growing user base. https://www.reuters.com/technology/chatgpt-sets-record-fastest-growing-user-base-analyst-note-2023-02-01 Last accessed 2024-10-01
Huang X, Dai S, Honggang X (2020) Predicting tourists' health risk preventative behaviour and travelling satisfaction in Tibet: combining the theory of planned behaviour and health belief model. Tour Manag Perspect 33:100589. https://doi.org/10.1016/j.tmp.2019.100589
Ipsos (2022) Global attitudes towards artificial intelligence
Jensen T, Theofanos M, Greene K, Williams O, Goad K, Fofang JB (2024) Reflection of its creators: qualitative analysis of general public and expert perceptions of artificial intelligence. Proc AAAI/ACM Conf AI Ethics Soc 7:647–58. https://doi.org/10.1609/aies.v7i1.31667
Johnson-Laird PN (2010) Mental models and human reasoning. Proc Natl Acad Sci USA 107(43):18243–50. https://doi.org/10.1073/pnas.1012933107
Jungherr A (2023) Artificial intelligence and democracy: a conceptual framework. Soc Med Soc. https://doi.org/10.1177/20563051231186353
Karaca O, Çaliskan SA, Demir K (2021) Medical artificial intelligence readiness scale for medical students (MAIRS-MS) - development, validity and reliability study. BMC Med Educ. https://doi.org/10.1186/s12909-021-02546-6
Kelley PG, Yang Y, Heldreth C, Moessner C, Sedley A, Kramm A, Newman DT, Woodruff A (2021) Exciting, useful, worrying, futuristic: public perception of artificial intelligence in 8 countries. In: Proceedings of the 2021 AAAI/ACM Conference on AI, Ethics, and Society, 627–37. AIES '21. ACM. https://doi.org/10.1145/3461702.3462605
Kim H-W, Chan HC, Gupta S (2007) Value-based adoption of mobile internet: an empirical investigation. Decis Sup Syst 43(1):111–26. https://doi.org/10.1016/j.dss.2005.05.009
Königs P (2025) The negativity crisis of AI ethics. Synthese 206. https://doi.org/10.1007/s11229-025-05378-9
Krewski D, Turner MC, Lemyre L, Lee JEC (2012) Expert vs. public perception of population health risks in Canada. J Risk Res 15(6):601–25. https://doi.org/10.1080/13669877.2011.649297
Lecun Y, Bengio Y, Hinton G (2015) Deep learning. Nature 521(7553):436–44. https://doi.org/10.1038/nature14539
Lee S, Chung M, Kim N, Mo Jones-Jang S (2024) Public perceptions of ChatGPT: exploring how nonexperts evaluate its risks and benefits. Technol Mind Behav. https://doi.org/10.1037/tmb0000140
Leiner DJ (2019) Too fast, too straight, too weird: non-reactive indicators for meaningless data in internet surveys. Surv Res Methods 13(3):229–48. https://doi.org/10.18148/srm/2019.v13i3.7403
Lerner JS, Li Y, Valdesolo P, Kassam KS (2015) Emotion and decision making. Annu Rev Psychol 66(1):799–823. https://doi.org/10.1146/annurev-psych-010213-115043
Lim S, Shim H (2022) No secrets between the two of Us: privacy concerns over using AI agents. Cyberpsychol J Psychosoc Res Cybersp. https://doi.org/10.5817/cp2022-4-3
Luhmann N (1989) Ecological communication. University of Chicago Press, Chicago
Makridakis S (2017) The forthcoming artificial intelligence (AI) revolution: its impact on society and firms. Futures 90:46–60. https://doi.org/10.1016/j.futures.2017.03.006
Malfacini K (2025) The impacts of companion AI on human relationships: risks, benefits, and design considerations. AI Soc 40(7):5527–40. https://doi.org/10.1007/s00146-025-02318-6
Marx E, Leonhardt T, Baberowski D, Bergner N (2022) Using matchboxes to teach the basics of machine learning: an analysis of (possible) misconceptions. In: Proceedings of the Second Teaching Machine Learning and Artificial Intelligence Workshop, edited by Katherine M. Kinnaird, Peter Steinbach, and Oliver Guhr, 170:25–29. Proceedings of Machine Learning Research. PMLR
McCarthy J, Minsky ML, Rochester N, Shannon CE (2006) A proposal for the dartmouth summer research project on artificial intelligence. AI Mag 27(4):12–14
McLuhan M (1964) Understanding media: the extensions of man. McGraw-Hill, New York
Md. Abdus S, Cheung KKW (2019) Tropical cyclone risk perception and risk reduction analysis for coastal bangladesh: household and expert perspectives. Int J Disast Risk Reduct 41:101283. https://doi.org/10.1016/j.ijdrr.2019.101283
Moriniello F, Martí-Testón A, Muñoz A, Jasaui DS, Gracia L, Ernesto Solanes J (2024) Exploring the relationship between the coverage of AI in WIRED magazine and public opinion using sentiment analysis. Appl Sci 14(5):1994. https://doi.org/10.3390/app14051994
Müller VC, Bostrom N (2016) Future progress in artificial intelligence: a survey of expert opinion. In: Müller VC (ed) Fundamental issues of artificial intelligence. Springer, pp 553–571. https://doi.org/10.1007/978-3-319-26485-1_33
Neri H, Cozman F (2019) The role of experts in the public perception of risk of artificial intelligence. AI Soc 35(3):663–73. https://doi.org/10.1007/s00146-019-00924-9
Neyer FJ, Felber J, Gebhardt C (2016) Kurzskala Technikbereitschaft (TB, Technology Commitment). Zusammenstellung Sozialwissenschaftlicher Items Und Skalen (ZIS). https://doi.org/10.6102/ZIS244
Ng DT, Kit JK, Leung L, Chu SKW, Qiao MS (2021) Conceptualizing AI literacy: an exploratory review. Comput Educ Artif Intell 2:100041. https://doi.org/10.1016/j.caeai.2021.100041
Niessen D, Beierlein C, Rammstedt B, Lechner CM (2021) Interpersonal trust short scale (KUSIV3). ZIS Collect Items Scales Soc Sci. https://doi.org/10.6102/ZIS292_EXZ
Nissenbaum H (2001) How computer systems embody values. Computer 34(3):120–190. https://doi.org/10.1109/2.910905
Novozhilova E, Mays K, Paik S, Katz JE (2024) More capable, less benevolent: trust perceptions of AI systems across societal contexts. Mach Learn Knowl Extract 6(1):342–66. https://doi.org/10.3390/make6010017
Organisation for Economic Co-operation and Development. (2019) OECD AI principles. https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0449
Peacock WG, Brody SD, Highfield W (2005) Hurricane risk perceptions among florida's single family homeowners. Landsc Urban Plan 73(2–3):120–35. https://doi.org/10.1016/j.landurbplan.2004.11.004

Peterson T, Peterson M, Grant W (2004) Chapter two: social practice and biophysical process. Environ Commun Yearb 1:15–32. https://doi.org/10.1207/s15567362ecy0101_2

Pew Research Center (2023) What Americans know about everyday uses of artificial intelligence. Pew Research Center. https://www.pewresearch.org/

Pidgeon N, Fischhoff B (2011) The role of social and decision sciences in communicating uncertain climate risks. Nat Clim Chang 1(1):35–41. https://doi.org/10.1038/nclimate1080

Protzko J, Schooler JW (2023) What I didn't grow up with is dangerous: personal experience with a new technology or societal change reduces the belief that it corrupts youth. Front Psychol 14. https://doi.org/10.3389/fpsyg.2023.1017313

Puzanova ZV, Tertyshnikova AG, Pavlova UO (2024) Technological discourse in the Russian media: main strategies for representing artificial intelligence. RUDN J Sociol 24:747–63. https://doi.org/10.22363/2313-2272-2024-24-3-747-763

Rammstedt B, Beierlein C (2014) Can't we make it any shorter? The limits of personality assessment and ways to overcome them. J Individ Differ 35(4):212–20. https://doi.org/10.1027/1614-0001/a000141

Recchia G, Freeman ALJ, Spiegelhalter D (2021) How well did experts and laypeople forecast the size of the COVID-19 pandemic? Edited by Martial L. Ndeffo Mbah PLos One 16(5):e0250935. https://doi.org/10.1371/journal.pone.0250935

Robinson-Tay K, Peng W (2024) The role of knowledge and trust in developing risk perceptions of autonomous vehicles: a moderated mediation model. J Risk Res. https://doi.org/10.1080/13669877.2024.2360923

Rumelhart DE, Hinton GE, Williams RJ (1986) Learning representations by back-propagating errors. Nature 323(6088):533–36. https://doi.org/10.1038/323533a0

Russell S, Hauert S, Altman R, Manuela V (2015) Robotics: ethics of artificial intelligence. Nature 521(7553):415–18. https://doi.org/10.1038/521415a

Sadek M, Calvo RA, Mougenot C (2024) Closing the socio-technical gap in AI: the need for measuring practitioners' attitudes and perceptions. IEEE Technol Soc Mag 43(2):88–91. https://doi.org/10.1109/mts.2024.3392280

Sanguinetti P, Palomo B (2024) An alien in the newsroom: AI anxiety in European and American newspapers. Soc Sci 13(11):608. https://doi.org/10.3390/socsci13110608

Schwesig R, Brich I, Buder J, Huff M, Said N (2023) Using artificial intelligence (AI)? Risk and opportunity perception of AI predict people's willingness to use AI. J Risk Res 26:1053–84. https://doi.org/10.1080/13669877.2023.2249927

Shneiderman B (2021) Human-centered AI. Issues Sci Technol 37(2):56–61

Siegrist M, Keller C, Kastenholz H, Frey S, Wiek A (2007) Laypeople's and experts perception of nanotechnology hazards. Risk Anal 27(1):59–69. https://doi.org/10.1111/j.1539-6924.2006.00859.x

Sindermann C, Yang H, Elhai JD, Yang S, Quan L, Li M, Montag C (2022) Acceptance and fear of artificial intelligence: associations with personality in a German and a Chinese sample. Discover Psychol. https://doi.org/10.1007/s44202-022-00020-y

Sjöberg L (1998) Risk perception: experts and the public. Eur Psychol 3(1):1–12. https://doi.org/10.1027/1016-9040.3.1.1

Slovic P, Fischhoff B, Lichtenstein S (1986) The psychometric study of risk perception. In: Risk Evaluation and Management, 3–24. Springer US. https://doi.org/10.1007/978-1-4613-2103-3_1

Slovic P, Flynn J, Mertz CK, Poumadère M, Mays C (2000) nuclear power and the public: a comparative study of risk perception in france and the united states. In: Cross-Cultural Risk Perception, 55–102. Springer US. https://doi.org/10.1007/978-1-4757-4891-8_2

Slovic P, Finucane ML, Peters E, MacGregor DG (2007) The affect heuristic. Eur J Oper Res 177(3):1333–52. https://doi.org/10.1016/j.ejor.2005.04.006

Spiel K, Haimson OL, Lottridge D (2019) How to do better with gender on surveys: a guide for HCI researchers. Interactions 26(4):62–65. https://doi.org/10.1145/3338283

Statista (2022) Artificial Intelligence (AI) Worldwide–Statistics & Facts. https://www.statista.com/topics/3104/artificial-intelligence-ai-worldwide/dossier-chapter1. Last accessed 28 Nov. 2022

Stevens JP (2012) Applied multivariate statistics for the social sciences. Routledge. https://doi.org/10.4324/9780203843130

The Alan Turing Institute and Ada Lovelace Institute (2023) Understanding public attitudes to AI. The Alan Turing Institute. https://www.turing.ac.uk/

Thomson ME, Önkal D, Avcioğlu A, Goodwin P (2004) Aviation risk perception: a comparison between experts and novices. Risk Anal 24(6):1585–95. https://doi.org/10.1111/j.0272-4332.2004.00552.x

Thoreau HD (1854) Walden; or, Life in the Woods. Ticknor; Fields, Boston

Tourangeau R, Rips LJ, Rasinski K (2000) The Psychology of Survey Response. Camb Univ Press. https://doi.org/10.1017/cbo9780511819322

Ventura A, Starke C, Righetti F, Köbis N (2025) Relationships in the age of AI: a review on the opportunities and risks of synthetic relationships to reduce loneliness. https://doi.org/10.31234/osf.io/w7nmz_v2

Verdurme A, Viaene J (2003) Consumer beliefs and attitude towards genetically modified food: basis for segmentation and implications for communication. Agribusiness 19(1):91–113. https://doi.org/10.1002/agr.10045

Wang S (2025) public perceptions of artificial intelligence in 20 countries: assessing individual- and country-level factors. Cross-Cult Res. https://doi.org/10.1177/10693971251336803

Wen C-HR, Chen Y-NK (2024) Understanding public perceptions of revolutionary technology: the role of political ideology, knowledge, and news consumption. J Sci Commun. https://doi.org/10.22323/2.23050207

Witte K, Allen M (2000) A meta-analysis of fear appeals: implications for effective public health campaigns. Health Educ Behav 27(5):591–615. https://doi.org/10.1177/109019810002700506

Wolfers L, Baumgartner S (2025) Short but still valid: validating single-item measures for key communication science constructs for experience sampling research. PsyArXiv. https://doi.org/10.31234/osf.io/zkwxa_v1

Yigitcanlar T, Degirmenci K, Inkinen T (2022) Drivers behind the public perception of artificial intelligence: insights from major Australian cities. AI Soc 39(3):833–53. https://doi.org/10.1007/s00146-022-01566-0

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.