Reduced AI Acceptance After the Generative AI Boom: Evidence From a Two-Wave Survey Study

JOACHIM BAUMANN*, University of Zurich and Bocconi University
ALEKSANDRA URMAN, University of Zurich
ULRICH LEICHT-DEOBALD, University of Dublin
ZACHARY J. ROMAN, University of Zurich
ANIKÓ HANNÁK, University of Zurich
MARKUS CHRISTEN, University of Zurich

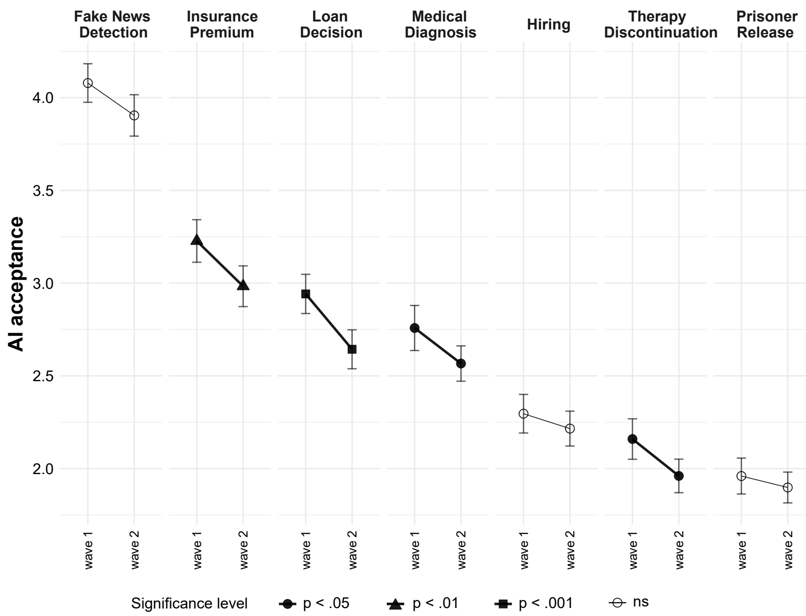

Fig. 1. The generative AI boom that occurred between two survey waves (January-February 2022 and July-August 2023) is associated with a significant decrease in AI acceptance for four out of seven decision-making scenarios among a representative Swiss sample. The error bars represent 95% confidence intervals (CI) and significance levels indicate statistically significant differences across survey waves as determined by a two-sample design-based t-test.

The rapid adoption of generative artificial intelligence (GenAI) technologies has led many organizations to integrate AI into their products and services, often without considering user preferences. Yet, public attitudes toward AI use, especially in impactful decision-making scenarios, are underexplored. Using a large-scale two-wave survey study ($n_{\text{wave 1}} = 1514$, $n_{\text{wave 2}} = 1488$) representative of the Swiss population, we examine shifts in public attitudes toward AI before and after the launch of ChatGPT. We find that the GenAI boom is significantly associated with reduced public acceptance of AI (see Figure 1) and increased demand for human oversight in various decision-making contexts. The proportion of respondents finding AI "not acceptable

*Corresponding author

Authors' Contact Information: Joachim Baumann, joachim.baumann@unibocconi.it, University of Zurich and Bocconi University; Aleksandra Urman, urman@ifi.uzh.ch, University of Zurich; Ulrich Leicht-Deobald, ulrich.leicht-deobald@tcd.ie, University of Dublin; Zachary J. Roman, zacharyjoseph.roman@uzh.ch, University of Zurich; Anikó Hannák, hannak@ifi.uzh.ch, University of Zurich; Markus Christen, markus.christen@dsi.uzh.ch, University of Zurich.

at all" increased from 23% to 30%, while support for human-only decision-making rose from 18% to 26%. These shifts have amplified existing social inequalities in terms of widened educational, linguistic, and gender gaps post-boom. Our findings challenge industry assumptions about public readiness for AI deployment and highlight the critical importance of aligning technological development with evolving public preferences.

Additional Key Words and Phrases: artificial intelligence,generative AI boom,ChatGPT,human-AI collaboration,digital inequality,public opinion

1 Introduction

Artificial intelligence (AI) has become ubiquitous in our daily lives, appearing even in unexpected places such as pillows and toothbrushes [36]. The recent proliferation of generative AI (GenAI) technologies has driven adoption across various domains, from customer service and search engines to health consultations [19, 60, 85, 107]. This rush to incorporate GenAI into products and services is often motivated by a fear of falling behind, leading organizations to hastily adopt AI technologies without fully considering public acceptance. Industry professionals believe in the positive impact of AI [118] and often also believe (e.g., in media [20]) that customers expect the implementation of new, AI-powered tools. However, this perception may diverge from people's preferences, potentially harming organizations' legitimacy due to the widening gap between imagined and real users [113].

Furthermore, with the widespread adoption of AI and its growing capabilities, governments around the world have started issuing regulations to manage its use. A critical element of these regulatory efforts is the clear specification of the required levels of human oversight and control over AI systems. The European Union's Artificial Intelligence Act (EU AI Act), which came into force on August 1, 2024, mandates specific human oversight measures depending on the risk that AI systems may cause [45]. Regulations have also been proposed in other countries, such as the U.S. executive order on AI, published on October 30, 2023 [18].

Therefore, although in today's AI race organizations may try to find more profitable AI use cases [118], they may overlook the opinions and acceptance levels of the general population regarding their use of AI. Moreover, the broader implications of this AI adoption, such as exacerbating existing inequalities – including pronounced distrust of health systems among women and general technological inequality – remain largely underexplored [2, 17, 44, 54]. Moreover, despite ongoing regulation efforts, we still lack a concrete understanding of public opinion on the appropriate levels of human control required for AI systems in different organizational application contexts. This gap underscores the need for research that quantifies laypeople's perceptions and provides actionable insights for policymakers.

1.1 Contributions.

In this study, we aim to investigate the public's acceptance of AI use in different decision-making scenarios and their corresponding preferences for human control through a representative two-wave survey study in Switzerland. Our paper makes three important theoretical contributions.

First, our research substantiates that public opinions shift over time, particularly during a technological boom. Pre-GenAI boom studies have quickly become outdated, as public discourse has shifted tangibly [94]. Our survey design uniquely addresses this by running the same questionnaire in two waves with a temporal gap that captures the shift in public sentiment towards AI in times of technological disruption. Additionally, analyses of detailed individual characteristics reveal increased demographic disparities along the lines of gender, education, and language regions over time.

Second, our paper contributes by examining AI acceptance in Switzerland, a country known for cultural diversity (being at the intersection of German, French, and Italian cultural European spaces) and high levels of innovation and digital literacy [92]. This diversity makes Switzerland an ideal setting to explore how contextual differences may shape public opinion on AI. Previous research has mainly focused on the U.S. context [72], and prior work suggests one cannot simply generalize results from one country to others since AI acceptance varies widely across cultural and national contexts [51, 72]. Furthermore, existing studies predominantly relied on non-representative samples, such as university students, Amazon Mechanical Turk workers, or medical clinic visitors, which may not accurately reflect the general population [31, 64, 73, 133]. To overcome these limitations, we employed a two-wave survey design representative with respect to age, gender, language, education, and political orientation to capture more reliable insights into public sentiment in Switzerland toward different AI use cases.

Finally, we contribute to the literature on human control in AI by offering a fine-grained conceptualization of AI versus human control. Prior research has found that human involvement remains essential and that people are often hesitant to give machines full control [114, 119]. However, European regulation only provides a high-level description of human oversight requirements [45], which does not reflect the complexity of organizational realities in monitoring AI [10, 124, 136]. Building on prior work on algorithm aversion and human-AI collaboration [32, 41, 56], we address this complexity by examining shifts in public preferences for different levels of human control in AI systems over time. In particular, we differentiate between four levels of human control: (a) human-only decisions, (b) human decisions with AI assistance, (c) AI decisions with possible human intervention, and (d) fully autonomous AI decisions. This can help regulators establish clearer recommendations on the appropriate levels of human control across different use cases.

2 Related work on contextual AI acceptance, human control, and the GenAI boom

2.1 Contextual AI acceptance

In this paper, we focus on *contextual* AI acceptance, or the extent to which an individual finds it acceptable to use AI *in a given domain*. Recent reviews of the literature on AI acceptance [72, 73] have shown that most studies to date have focused either on the acceptance of AI in general or on AI in single use cases and/or industries, such as education [75], healthcare [4], or public services [111]. While such studies provide important insights about given usage scenarios, the few studies that examined AI acceptance across usage contexts from a comparative perspective show that AI acceptance varies widely across decision-making scenarios and industries [6, 28, 68, 72, 94]. For example, medicine has been consistently found to be among the domains in which the acceptance of AI tends to be the lowest [68, 72]. This might be related to the fact that medical decisions are highly consequential, and more consequential decision-making scenarios have been linked to lower contextual AI acceptance [28].

It is important to note that acceptance and trust are distinct concepts in the technology adoption literature. While trust can be an antecedent or predictor of acceptance [73], acceptance itself encompasses the behavioral intention or willingness to use, buy, or try a technology [40]. In our study, we build on acceptance models from the literature [40, 58], adapted to our use case, i.e., to investigate users' willingness to accept or object to AI usage in various decision-making scenarios.

While AI acceptance varies across contexts, comparative – that is, cross-industry/cross-usage scenario – evaluations of AI acceptance are rare [73]. This limits our ability to draw generalizable conclusions about AI acceptance across different scenarios. While in principle, one could make such conclusions by, for example, comparing acceptance levels observed in the studies focused on specific single scenarios, in practice, such direct comparisons are often not possible for two main reasons. First, individual studies are conducted at different points in time, and AI acceptance can change from one point to another. Second, the studies rely on different samples, often drawing conclusions not from representative samples of the population but from specific social groups such as students, business professionals, or even Amazon Mechanical Turk workers [72, 73]. Therefore, we focus on *contextual* acceptance of AI rather than more general AI acceptance.

2.2 Human control of AI

The EU AI Act underscores the necessity of *human oversight* for *high-risk* AI systems, stating that: "High-risk AI systems shall be designed and developed in such a way, including with appropriate human-machine interface tools, that they can be effectively overseen by natural persons during the period in which they are in use." [45, Article 14]. This emphasis on human oversight is particularly crucial in high-impact decision-making contexts across various domains, such as healthcare [5, 14, 57, 95, 131, 134]. In clinical decision support systems, for example, AI can provide valuable assistance in diagnosing patients or flagging potential issues, but human involvement remains essential to ensure that the AI's recommendations are correctly interpreted and applied [125].

Human control in AI systems is often conceptualized along a spectrum, ranging from full human control to full AI autonomy [76]. At one extreme, humans make all decisions without AI input; at the other, AI systems operate entirely autonomously without human intervention. Between these two poles, various collaborative forms exist, such as AI-assisted human decisions or AI-generated decisions subject to human oversight [1, 105, 116]. These models are also commonly referred to as *human-in-the-loop* – where the human retains primary decision-making authority – and *human-on-the-loop* – where the human supervises and can intervene to correct AI-driven decisions [50].

The degree of human control over AI significantly shapes how users perceive the role of AI in decision-making, with studies showing a general preference for higher levels of human involvement [8, 93, 114]. Understanding the nuances of this preference, particularly in high-impact scenarios, is crucial for designing AI systems that are trusted and widely accepted. Prior experimental studies have examined various dimensions of human-algorithm interaction, including outcome modification [41], process versus outcome control [32], and the principles that should guide algorithm-in-the-loop decision making [56]. These works highlight the importance of accounting for different levels of human control when designing AI systems. Our study complements this literature by examining how public perceptions have shifted regarding different human-AI collaboration models in decision-making contexts.

2.3 Generative AI boom

Attitudes toward technologies are not static and may change over time, oftentimes nurtured by disruptive changes [70]. In November 2022, the launch of ChatGPT attracted worldwide attention and marked a crucial moment in the public discourse surrounding AI – with then unknown effects on the public's acceptance towards AI [94]. The widespread accessibility of ChatGPT and the capabilities of its underlying large language model (LLM)[1], along with some other highly capable generative AI models, such as Stable Diffusion [108] or DALL-E 2 [104], that were released at a similar time, sparked intense discussions on the potential benefits and challenges of AI [71, 88, 130]. This surge in interest is reflected across various domains, from everyday conversations to extensive media coverage and governmental deliberations, and has been referred to as the *generative AI boom* [78, 127].

As shown in Figure 2 (left), the generative AI boom has led to a drastic increase in the proportion of news articles about AI in Switzerland. The data[2] indicates that coverage has tripled over the period, highlighting growing media attention towards AI.

[1]See OpenAI's technical report on their latest LLM [100].

[2]This data was retrieved from Factiva (c.f. https://global.factiva.com), an international news database produced by Dow Jones that contains 187 different Swiss news outlets. We queried Swiss media outlets using the search terms "Künstliche Intelligenz," "Intelligence artificielle,' and "Intelligenza artificiale," representing the translation of AI into German, French, and Italian – the three official languages in Switzerland in which most news media are published.

Similarly, the AI boom has also led to a parallel growth in political activity related to AI advancements and implications in Switzerland, as shown in Figure 2 (right). This data[3] reflects a major rise in the number of parliamentary interventions involving AI-related terms, underscoring the political response to the evolving AI landscape.

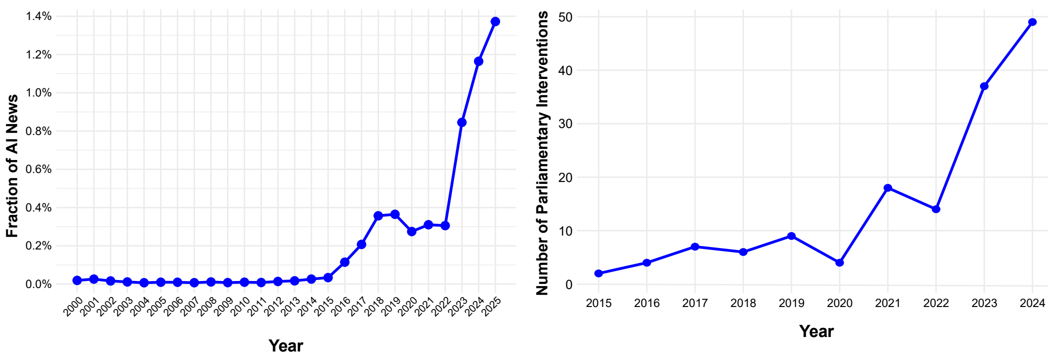

Fig. 2. The generative AI boom in Swiss media and politics. The left panel shows the fraction of AI news in all Swiss news from 2000 to 2025 that mention AI (weighted by the total number of news per year). The left panel shows the number of parliamentary interventions on AI from 2015 to 2024 involving AI-related terms.

Public attitudes towards the use of AI, as well as the need for human oversight, are likely influenced by shifts in public discourse following major technological advancements [52, 115, 135]. Notably, the launch of ChatGPT has been a significant milestone in AI development. With this context in mind, we aim to explore the following research questions:

RQ1a ***How did the AI acceptance change with the generative AI boom?***

RQ1b ***How did the preferences for human control of AI change with the generative AI boom?***

Since AI acceptance and human control are highly contextual [6, 28, 68, 72, 94], we investigate both research questions across several specific usage scenarios.

2.3.1 AI acceptance and individual characteristics. AI usage scenarios and their characteristics are only one factor explaining the differences in AI acceptance rates. However, even within the same usage scenario, AI acceptance can also vary based on people's individual characteristics, including socio-demographic attributes [73, 96].

Gender consistently emerges as a significant predictor of AI acceptance, with studies showing that women tend to be more skeptical of AI, perceive it as less useful, and see fewer personal and societal benefits in its adoption [6, 33, 55, 66, 96, 109]. These findings align with broader research on gender differences in technology acceptance generally, where similar patterns have been observed for digital technologies [27, 122]. Hence, we anticipate that gender will be significantly associated with the respondents' AI acceptance in our study as well, with women exhibiting lower acceptance towards AI use generally. At the same time, we expect the significance and/or the effect size to be different across AI use scenarios, in line with the findings of Araujo et al. [6] who showed that

[3]This data was retrieved from Curia Vista, a database containing all parliamentary proceedings of the Federal Assembly of Switzerland since 1995 – c.f. https://www.parlament.ch/en/ratsbetrieb/curia-vista.

gender differences in the perceptions of AI-based decision-making systems' risk,usefulness, and fairness vary between different decision-making domains.

Age is another demographic factor linked to lower acceptance of AI and technology in general, with older individuals generally holding more negative attitudes towards AI [6, 55, 96]. Additionally, higher levels of *education* and *familiarity with AI* technologies have been shown to correlate positively with AI acceptance and trust in AI decision-making [6].

Given these established group-level differences, we aim to explore potential changes in these dynamics during the rise of generative AI:

> ***RQ2: How have group differences in AI acceptance, based on gender, age, education, language regions, digital skills, and AI knowledge, evolved with the generative AI boom?***

3 Method

3.1 Data collection

We conducted two surveys, the first in January and February 2022, before the GenAI boom, and the second 17 months later in July and August 2023, after the GenAI boom. Both surveys were distributed by a professional survey company[4] considering quotas for age, gender, and language region to ensure the representativeness of the adult Swiss population based on data from Switzerland's Federal Statistical Office [46, 48].

The final samples consist of 1514 1st wave respondents and 1488 2nd wave respondents (more details in Table 1). In total, the 1st survey wave questionnaire was accessed 2318 times. 566 persons (24%) dropped out of the survey; another 238 persons (10%) were excluded due to failure in the attention check. Of the 2056 persons that started the survey wave 2, about 22% dropped out, and about 8% were excluded after failing the attention check.[5]

Table 1. Survey execution and sample sizes.

	Wave 1 (pre GenAI boom)	**Wave 2** (post GenAI boom)
Survey period	Jan 20 – Feb 12, 2022	Jul 26 – Aug 22, 2023
Total participants[a]	2318	2056
Participants dropped out	566	443
Excluded answers[b]	238	125
Final sample size	**1514**	**1488**

[a] This includes all participants who started filling out the survey but excludes those who were not admitted due to unfulfilled demographic quotas.
[b] Both surveys included an attention question, and we only considered data from participants who answered correctly.

We inspected ethical and privacy issues based on the DESAT process of the University of Zurich (for details see https://www.dsi.uzh.ch/en/research/projects/archive/strategic/desat.html). Since we

[4]gfs Bern; c.f. https://www.gfsbern.ch/en/

[5]The attention check consisted of a question about favorite colors with embedded instructions to select both "Red" and "Green" options, regardless of actual preference. Participants who failed to follow these specific directions were excluded from the survey.

collected the data anonymously and none of the questions involved specifically sensitive issues, the study was evaluated as "low risk" and did not require approval by an ethics board. This is also in line with the checklist of the ethical committee of the lead author's institution.

3.2 Questionnaire

3.2.1 Demographics. The questionnaire structure was designed by the authors of this manuscript. Participants could choose to fill out the questionnaire in any of the official languages at the national level, i.e., German, French, or Italian. Furthermore, the beginning of the questionnaire included a set of socio-demographic questions (gender, age, language, education, and political orientation).

Digital skills and AI literacy. Next, we evaluated the respondents' digital skills. Measuring digital skills and AI literacy presents challenges due to the rapidly evolving technological landscape. For this study, we utilized three different scales in the first wave of our survey to assess participants' competencies in these areas. The first scale from Hargittai and Hsieh [61], evaluates respondents' familiarity with advanced digital concepts, including advanced search, PDF, spyware, wiki, cache, and phishing (wave 1: Cronbach's α=0.88, $\mu = 3.29$, $\sigma = 0.94$; wave 2: Cronbach's α=0.87, $\mu = 3.34$, $\sigma = 0.91$). The second scale, which was developed by Thomas Franke and Wessel [117], measures affinity for technology interaction through a 9-item questionnaire, reflecting individuals' propensity to engage with technology (wave 1: $\mu = 3.29$, $\sigma = 0.94$). Additionally, we introduced a scale for AI literacy, asking participants to rate their understanding of the terms "artificial intelligence," "machine learning," "neural networks," and "deep learning" (wave 1: Cronbach's α=0.89, $\mu = 3.65$, $\sigma = 0.98$).

The strong correlations observed among these scales (Spearman rank correlations ranging from 0.45 to 0.56, all with $p < 0.001$) support their reliability and alignment. Therefore, we used only the literacy scale by Hargittai and Hsieh [61], which is well-established in the literature, for the second wave and based all analyses in this manuscript on this measure to ensure consistency and reliability in measuring digital skills.

In the second survey wave, we additionally asked respondents about their familiarity with ChatGPT (see answer options in Table 4).

3.2.2 AI primer. The concept of AI lacks of a universally accepted definition [94, 102, 126]. Thus, it is crucial to appropriately prime study participants before they respond to the questionnaire. By ensuring that participants share a common understanding of AI, we can mitigate variability in interpretation and obtain more consistent responses. Therefore, we primed the participants with the following explanation, ensuring a common definition of AI and its use in decision-making.

AI primer

When we talk about "artificial intelligence" below, we mean the following: The term "artificial intelligence" (AI) refers to computer software that learns from data and is then able to make decisions or predictions. Examples include:

- Digital assistants such as "Alexa" from Amazon or "Siri" from Apple
- Systems that enable self-driving cars
- Systems that make personalized recommendations (e.g. Amazon)
- Systems that control robots
- Systems that moderate content on social media platforms (e.g. detection of hate speech)
- Systems that enable facial recognition (e.g. for access control)

In this survey, "artificial intelligence" (AI) refers to those AI systems that are in use today and their obvious further developments. It does not refer to unrealistic systems that surpass human intelligence, such as "Skynet" in the Terminator movies, robots like in the movie "I Robot" or "HAL 9000" in the movie "2001 A Space Odyssey". Please base the following questions on this understanding of AI.

The primer was specifically designed to orient participants toward existing contemporary AI applications rather than speculative or futuristic conceptions. While alternative definitions of AI exist (see Section 5.5 for a detailed discussion), for the remainder of this section, we proceed on the assumption that questionnaire respondents were successfully primed and thus share a common understanding of AI.

3.2.3 AI acceptance and human control. In the main part of the survey, we assessed the AI acceptance and human control conditions of using AI for automating decision-making, which is motivated by prior work [6]. To capture contextual variations, we consider seven paradigmatic applications, which are listed in Table 2. All scenarios concern already existing or likely real-world AI applications (i.e., fake news detection, insurance premium pricing, loan decisions, medical diagnosis, hiring decisions, therapy discontinuation, and prisoner release decisions).

For AI acceptance, we asked the question *"How acceptable would it be for you if this decision was made by an AI system?"* with standard five-point Likert scales plus an additional option for respondents to indicate that they don't know. Hence, respondents could choose one of the following 6 answer possibilities: (1) *Not acceptable at all*, (2) *rather not acceptable*, (3) *neutral assessment*, (4) *rather acceptable*, (5) *definitely acceptable*, (6) *I don't know.*

For human control, we asked survey participants to "*Please assess what would be the optimal way to make the following decisions,*" by selecting one of the options listed in Table 3.[6]

3.3 Sample description

First, we provide an overview of the survey respondents across both waves of the study. We provide more details in Figure 6 in Appendix A.

Gender. Regarding gender distribution, in wave 1, 740 respondents identified as men, 769 as women, 2 as other, and 3 chose not to answer. In wave 2, the numbers slightly varied, with 739 men and 741 women, 4 identifying as other, and 4 opting not to disclose their gender.

Language. Only a few participants are from the Italian-speaking region, which is also the language region with fewer inhabitants compared to the German- and French-speaking regions.

[6]Notice that for all our numerical analyses, the additional option ("I don't know") was filtered out. As shown in Tables 5 and 6, only very few people have chosen this answer option (1.89% wave 1 and 0.67% in wave 2).

Table 2. Detailed descriptions of the considered scenarios.

Scenario	Description (with relevant references provided in brackets)
Fake news detection	Identification of fake news in a social network. [34, 59, 74]
Insurance premium	Determining the premium for liability insurance based on an individual's claims history. [12, 30, 86]
Loan decision	Deciding whether a person should get a loan based on their credit history. [53, 63, 81]
Medical diagnosis	Making a diagnosis on the basis of a patient's medical data. [7, 39, 103, 112, 128]
Hiring	Selecting a person for an interview based on their CV. [22, 23, 62, 79, 129]
Therapy discontinuation	Decision to discontinue therapy based on a patient's medical condition. [39, 84]
Prisoner release	Deciding whether to release an inmate based on various data (criminal history, behavior in prison, etc.). [3, 16, 21, 120]

Table 3. Answer options for different levels of human control over decisions made. The two answer options *AI decision with human oversight* and *AI-assisted human decision* both fall into the category of **human-AI collaboration**.

Answer option	Description
AI only	An AI system can decide this independently and automatically.
AI decision with human oversight	An AI system can make this decision itself, but a human should always check the decision and intervene if necessary.
AI-assisted human decision	A human should make the decision, but an AI system should give advice on what the optimal decision might be.
Human only	Only a human should make such decisions, without the involvement of AI.
I don't know	I do not know.

Age. The age distribution of respondents remained consistent across survey waves. In wave 1, the average age of participants was 49.73 years, while in wave 2, it was at 50.27 years. For analysis, we categorized the participants into three age groups: under 40, 40-64, and 65 and above.

Education. We used the following ten standard Swiss education levels: (1) Primary school or no degree, (2) compulsory school, (3) transitional education, (4) general education without a baccalaureate, (5) basic vocational training or apprenticeship, (6) baccalaureate or similar (high school, vocational baccalaureate, specialized baccalaureate), (7) post-secondary non-tertiary level (additional or secondary education), (8) higher vocational training with federal certificate, (9) university of applied sciences, university (including bachelor's, master's, and postgraduate studies), (10) doctorate, habilitation, (11) other, (12) no answer. For the analysis, we grouped education levels into three categories: university degree (levels 9-10), advanced (levels 6-8), and general (levels 1-5 & 11-12).

In both waves, the majority of respondents reported having a university degree or higher (levels 9-10), with 740 individuals in wave 1 and 702 in wave 2. About 30% of participants of both survey waves had an advanced degree (levels 6-8). Only 318 (350 in wave 2) participants fell in the general education bucket.

Political orientation. In terms of political orientation, respondents in wave 1 reported an average score of 4.14, whereas wave 2 respondents reported an average of 4.36 on a scale from 0 (far left) to 10 (far right).

ChatGPT awareness. The second wave of our survey indicates a high level of awareness about ChatGPT among participants, as shown in Table 4. The results reveal that a vast majority (96.37%) are aware of this new technology. Specifically, less than 4% of participants have never heard of ChatGPT, and almost half of the participants have even used it themselves. These numbers underscore the widespread awareness of AI among the public.

Table 4. Survey responses on ChatGPT awareness and usage.

Answer option	**Count (in %)**
I have never heard of a program called ChatGPT	54 (3.6%)
I have heard of ChatGPT but have not looked into it	287 (19.3%)
I have read about ChatGPT but have never tried it myself	479 (32.2%)
I have tried ChatGPT myself, but don't use it otherwise	448 (30.1%)
I use ChatGPT regularly for private and/or professional purposes	220 (14.8%)

For the remainder of this manuscript, all results are provided based on weighted survey data.

3.4 Data analysis

We used two mechanisms to ensure representativeness. First, we set quotas of age, gender, and language region when collecting the data to make sure we have a balanced sample according to these demographic characteristics. Second, for both surveys, we weighted the data of the final sample to ensure representativeness with respect to age, gender, language region, education, and political orientation. The weighting was done by the professional survey company and is based on standard survey weighting data from the Swiss Population Census [47]. We used the open-source `survey` package [87] for the survey-weighted analyses in R.

Unless stated otherwise, we present weighted means with 95% CI in the following analysis. We report adjusted p-values [132], following the Benjamini–Hochberg procedure to control the false discovery rate for multiple hypothesis testing [15]. We denote statistical significance with *** if p-value < 0.001, ** if p-value < 0.01, * if p-value < 0.05, and not significant (ns) otherwise.

4 Results

Summary of the main findings

RQ1: **GenAI boom:** The recent surge in generative AI is associated with a **decrease in overall AI acceptance** and an **increased demand for human oversight**. The associations are statistically significant for almost all scenarios (see Figures 1 and 3). Yet, AI acceptance and human control preferences remain highly contextual across both waves.

RQ2: **Amplification of existing inequalities:** Our study reveals that the **GenAI boom is associated with exacerbated pre-existing disparities** in AI acceptance:

- Respondents with university degrees show consistently higher acceptance of AI compared to those with advanced non-university education, with the gap widening after the GenAI boom.
- French-speaking respondents demonstrate lower AI acceptance than German speakers, a difference that remains stable across most scenarios but has increased for medical contexts since the GenAI boom.
- Women express significantly more skepticism toward AI in medical scenarios (medical diagnosis and therapy discontinuation), and this gender gap has grown further post-GenAI boom.

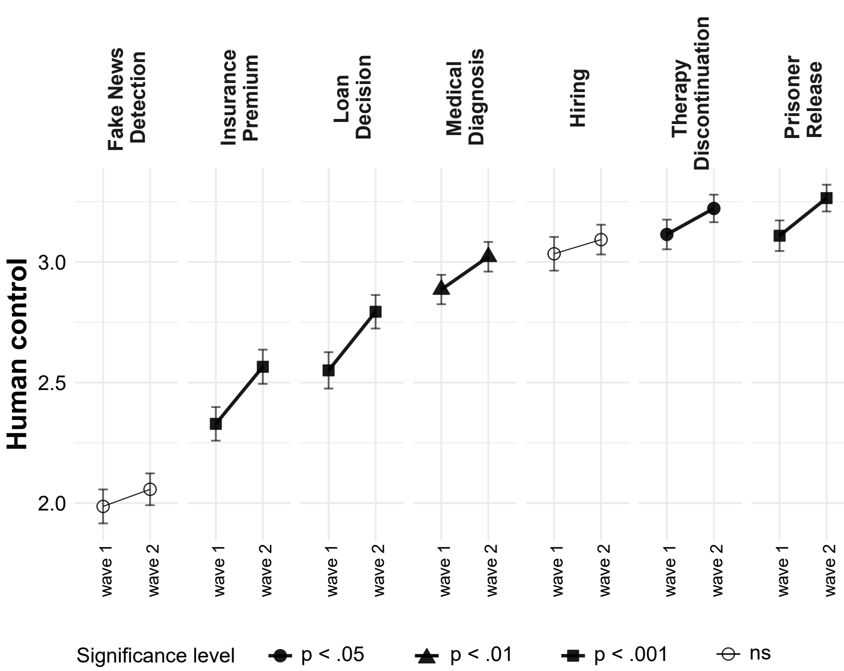

Fig. 3. The generative AI boom is associated with a significant increase in required human control for five out of seven scenarios. In both survey waves, human control preferences are heavily dependent on context. Just as in Figure 1, we show weighted means with 95% CI and significance levels indicate statistically significant differences across survey waves as determined by a two-sample design-based t-test.

4.1 RQ1a: Reduced AI acceptance with the GenAI boom

The average acceptance of AI was significantly lower at wave 2 compared to wave 1 across five out of seven examined scenarios post-GenAI boom, as depicted in Figure 1. Table 5 shows the response distribution for each wave, irrespective of specific scenarios, in more detail. Notably, the fraction of respondents who found AI "not acceptable at all" increased from 23.87% in wave 1 to 30.09% in wave 2, indicating a stark rise in strong opposition to AI applications. Conversely, the percentage of people deeming AI "definitely acceptable" decreased from 13.07% to 10.47%.

Scenarios involving high-impact decisions, such as prisoner release, therapy discontinuation, and hiring decisions, showed the lowest acceptance levels in both waves. But even scenarios with much higher acceptance levels, such as fake news detection and insurance premiums, showed pronounced increases in people finding the use of AI not acceptable at all. With the generative AI boom, all scenarios exhibited consistent declines in the share of people perceiving the use of AI as "definitely acceptable". These effects are consistent across all scenarios, as can be seen in Figure 4.

These results suggest a growing skepticism towards AI in critical decision-making areas, highlighting increased concerns about the reliability and ethics of AI technologies, especially in sensitive contexts.

Table 5. Overview of responses regarding the perceived necessity for human control. Values show weighted averages across all scenarios. The survey-weighted chi-square test for overall distribution change between the two waves is statistically significant ($p < 0.001$). The significance column shows results from individual chi-square tests for each response option, indicating whether the proportion of that response changed significantly between waves.

Response type	Wave 1	Wave 2	Significance
Not at all acceptable	23.87%	30.09%	**
Rather unacceptable	23.38%	22.03%	ns
Neutral assessment	15.10%	15.69%	ns
Rather acceptable	22.70%	21.05%	ns
Definitely acceptable	13.07%	10.47%	*
I don't know	1.89%	0.67%	*

4.2 RQ1b: Increased need for human control with the GenAI boom

The reported necessity for human oversight in AI decisions increased significantly across almost all scenarios, as illustrated in Figure 3. Table 6 shows the response distribution in more detail, and Figure 5 shows the results across the different scenarios. Interestingly, people predominantly favor human-AI collaboration (human in or on the loop) across all scenarios, namely, 72.06% of respondents in wave 1 and 66.68% in wave 2. A considerable fraction of people think AI should not be involved in the decision process at all, 18.04% before and 25.52% after the generative AI boom.

Scenarios with high-impact decisions, such as prisoner release, therapy discontinuation, hiring, or medical diagnosis, showed the highest desire for human control levels across both waves, with about 40% of people requiring only humans to be involved in wave 2. Other scenarios have also seen substantial increases in the reported necessity for human oversight: For example, the average reported human control values rose from 2.33 to 2.58 in the insurance premium scenario and from 2.55 to 2.81 in the banking scenario related to granting loans.

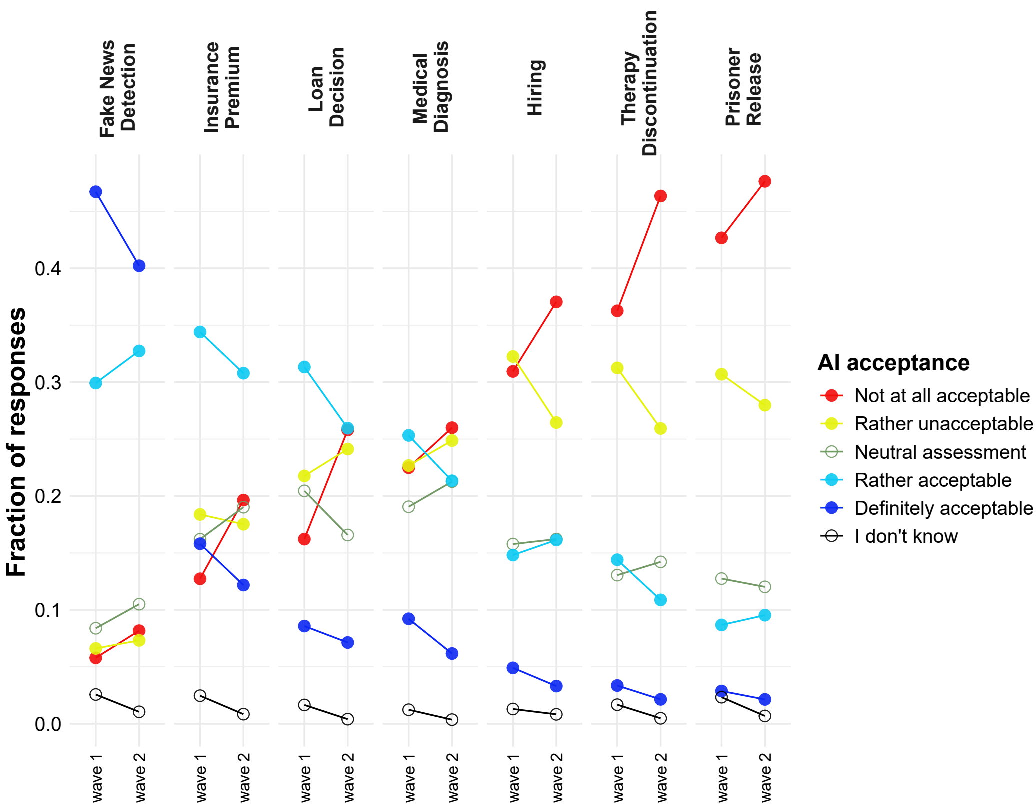

Fig. 4. Response distribution in % for acceptance questions across survey waves and scenarios.

In a similar way to AI acceptance levels, this trend highlights a rising concern about fully automated decision-making in sensitive contexts, and people want humans to be in the loop more often.

Table 6. Overview of responses regarding the perceived necessity for human control. Values show weighted averages across all scenarios. The survey-weighted chi-square test for overall distribution change between the two waves is statistically significant ($p < 0.001$). The significance column shows results from individual chi-square tests for each response option, indicating whether the proportion of that response changed significantly between waves.

Response type	Wave 1	Wave 2	Significance
AI only	7.90%	6.67%	ns
AI decision with human oversight	29.98%	26.07%	**
Human decision with AI assistance	42.08%	40.61%	ns
Human only	18.04%	25.52%	***
I don't know	2.00%	1.14%	ns

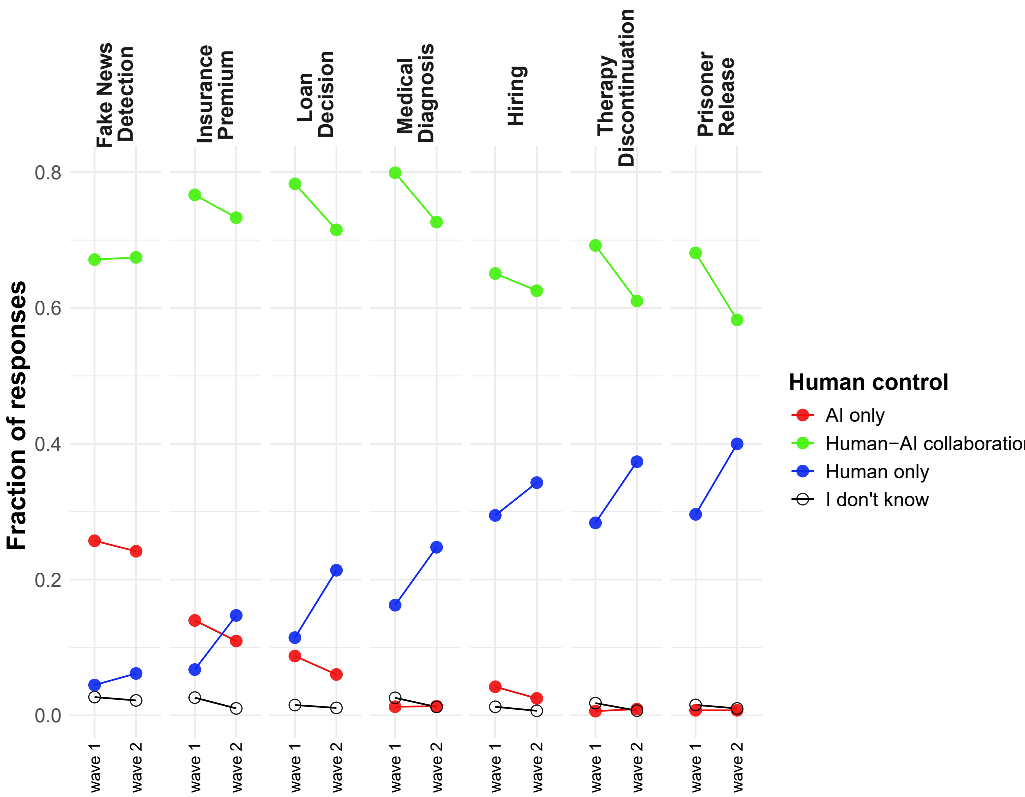

Fig. 5. Response distribution in % for human control questions across survey waves and scenarios. Notice that in this figure, the two answer options (*AI decision with human oversight* and *AI-assisted human decision*) are merged into one single category titled *human-AI collaboration*, as they are conceptually very similar – see Table 3 for more details. In Appendix B.1, we provide the same overview but consider both answer options separately, showing that the overall results remain unaffected. However, notice that the overall reduction of the share of people favoring human-AI collaboration is mainly driven by fewer respondents opting for AI decisions with human oversight.

4.3 RQ2: Amplification of existing inequalities with the GenAI boom

The survey-weighted linear regression analysis in Table 7 reveals several demographic predictors of AI acceptance and preferred human control (across all scenarios), highlighting nuanced trends. The level of education plays a key role in wave 2, with university-educated respondents showing a significant positive association with AI acceptance (compared to people with advanced education) and a tendency toward reduced emphasis on human control, suggesting higher confidence in AI autonomy. On the other hand, general education is associated with lower acceptance in wave 2. Age (40-64) also emerges as a significant predictor: it shows that older people (40-64) find AI use generally less acceptable than young people (<40) pre-GenAI boom. The significant positive predictor of human control in wave 1 for the same age group mirrors this result. Furthermore, German language is negatively associated with human control post-GenAI boom Meanwhile, when looking at the averages across all scenarios, we do not find any significant effects for digital literacy, political orientation, and gender.

Table 7. Survey-weighted linear regression results for mean AI acceptance and human control across all scenarios. Coefficients that are statistically significant are highlighted in bold.

	AI acceptance		**Human control**	
	Wave 1	Wave 2	Wave 1	Wave 2
Gender (f)	−0.037	−0.111	0.037	−0.005
	(0.078)	(0.075)	(0.047)	(0.043)
General education	−0.103	**−0.155***	0.047	0.055
	(0.082)	(0.073)	(0.050)	(0.045)
University degree	0.042	**0.148***	−0.080	**−0.111****
	(0.067)	(0.058)	(0.041)	(0.037)
German language	0.017	0.147	0.065	**−0.239***
	(0.188)	(0.193)	(0.129)	(0.105)
French language	−0.121	−0.221	0.168	0.004
	(0.200)	(0.197)	(0.134)	(0.110)
Digital literacy	0.053	0.070	−0.035	−0.044
	(0.047)	(0.040)	(0.030)	(0.025)
Age (40-64)	**−0.225***	−0.048	**0.136****	**0.104***
	(0.095)	(0.074)	(0.053)	(0.053)
Age (65+)	0.075	0.135	0.022	0.037
	(0.097)	(0.114)	(0.052)	(0.058)
Political orientation	0.020	0.012	0.007	−0.001
	(0.014)	(0.015)	(0.010)	(0.009)
Constant	**2.673*****	**2.380*****	**2.619*****	**3.112*****
	(0.264)	(0.239)	(0.180)	(0.146)
Note:	$^{*}p < 0.05$; $^{**}p < 0.01$; $^{***}p < 0.001$			

The survey results further indicate that the generative AI boom is associated with the reinforcement of existing group disparities for language regions and gender, but only in specific scenarios. In the following, we summarize the amplification of these disparities between wave 1 and 2. Please refer to Appendix B.2 for the associated figures and additional details.

4.3.1 Education gap. The GenAI boom is associated with the widening of the existing education-based disparities in AI acceptance (see Figure 8). Respondents with higher levels of education consistently exhibit greater acceptance of AI. Our findings show a clear trend: individuals with university degrees demonstrate the highest levels of AI acceptance, followed by those with advanced non-university education, while those with only general education express the lowest acceptance levels. This hierarchy is evident across all surveyed scenarios.

Crucially, the gap between these educational groups has further expanded during the the GenAI boom. While university-educated respondents maintained relatively stable acceptance levels, those with lower levels of education became increasingly critical of AI use. For more details, see Appendix B.2.1.

4.3.2 Language region gap. Our results show that exacerbated language-based disparities in AI acceptance have also been associated with the GenAI boom (see Figure 9). French-speaking respondents demonstrate lower AI acceptance compared to German-speaking respondents, and their acceptance significantly decreased across the two survey waves in medical contexts. For more details, see Appendix B.2.2.

4.3.3 Gender gap in health scenarios. The GenAI boom is also associated with the deepening of the gender gap in AI acceptance, especially in health-related scenarios (see Figure 10). Additionally, while women's AI acceptability has significantly decreased in five out of seven scenarios, for men, this effect was only significant in the loan decision scenario (see top panel of Figure 10). Women express significantly more skepticism toward the use of AI for medical diagnosis and therapy discontinuation compared to men, and this gender gap has widened further post-GenAI boom. For more details, see Appendix B.2.3.

5 Discussion

Our study highlights the critical role of contextuality in public acceptance of AI technologies, particularly in light of the recent generative AI boom. In high-impact scenarios like therapy discontinuation, where outcomes could be life-threatening, only about 3% of respondents view AI use without human oversight as acceptable. In contrast, in scenarios such as fake news detection, around 40% of respondents find AI use entirely acceptable. In contrast to Jussupow et al. [68], Kaufmann et al. [72], we find that not only medical scenarios but also hiring decisions and prisoner release decisions exhibit consistently low AI acceptances. Our observed skeptical tendency towards AI tells a very different story than a similar survey conducted in Britain, which found that people's perceived benefit levels outweigh concerns regarding AI use [94]. However, there are two key differences between these two studies: First, while they consider very general AI uses (including use cases like virtual AI assistants or AI-based simulations for advancing knowledge), we focus specifically on high-impact decision-making scenarios. Second, they ran their study before the GenAI boom, which has "already impacted public discourse towards some AI technologies since [their] survey" [94, p. 57]. Furthermore, our findings of varying acceptance in high-impact scenarios align with Castelo et al. [28] observation that people trust algorithms less in more subjective tasks. However, while they studied algorithm trust before the GenAI boom using a simplified definition of algorithms, our study examined AI acceptance after this technological shift with more detailed AI priming, potentially explaining some of the differences in overall sentiment.

In the following, we address several key aspects of our findings. We begin by examining how various AI definitions – including our own approach – shape research outcomes and policy considerations. Next, we explore the implications of public opinion shifts following the generative AI boom and contextualize the socio-demographic variations in AI acceptance we observed against the more general patterns of digital inequalities in Switzerland. We then consider how these changing public perceptions and socio-demographic differences could inform AI governance and regulatory frameworks. Finally, we acknowledge our study's limitations and propose future research directions that could bridge our survey-based methodology with the experimental designs more commonly employed in human-computer interaction (HCI) research on AI acceptance and trust.

5.1 The implications of AI definitions for research and policy

Our study's findings must be interpreted considering how "artificial intelligence" was defined for participants. This reflects a broader challenge in both research and policy: how different conceptions of AI influence public attitudes and regulatory responses.

Public and scholarly discourse around AI encompasses multiple definitions ranging from narrow task-specific systems to hypothetical general intelligence [98]. Our study focused on actual AI applications in use today, yet public attitudes may be shaped by broader cultural narratives about AI from science fiction and media [29, 98]. These definitional ambiguities create challenges for interpreting public opinion and developing governance frameworks.

First, public concerns may target different phenomena than those addressed by current policy initiatives. Despite our priming that aimed to focus participants on contemporary existing AI systems, survey respondents expressing concern about AI-based decision-making may be imagining more autonomous general intelligence rather than the statistical models currently widely deployed. The significant shift in AI acceptance we observed following the generative AI boom may reflect not just reactions to new technologies, but evolving conceptions of what constitutes "AI" in public understanding. Second, technical experts, policymakers, media commentators, and the general public often use the term "AI" differently [82, 98], potentially leading to misaligned expectations between these groups. This definitional fragmentation may partially explain the socio-demographic differences we observed – for example, educational background or gender may correlate with exposure to particular framings of AI [11, 29].

These definitional challenges have several implications for future research and policy development. Research on public attitudes toward AI should explicitly acknowledge how definitional choices may influence findings. Studies might systematically vary AI definitions provided to participants to understand how different framings affect reported attitudes. Such research could help disentangle concerns about current applications from fears about hypothetical future capabilities. For policymakers, these challenges suggest the importance of granular, technology-specific approaches rather than broad regulations of "AI" as a monolithic category. Our findings on context-dependent acceptance patterns provide empirical support for such granularity.

5.2 Generative AI boom and public opinion

One of the key findings of our study is the decline in overall AI acceptance and the increased demand for human control following the generative AI boom. Despite the growing prevalence of AI in everyday applications, respondents showed heightened concerns over AI use, particularly in high-impact contexts such as medical diagnosis and therapy discontinuation. The generative AI boom appears to have raised awareness about the limitations and risks of autonomous decision-making, thereby prompting calls for stricter human control and oversight, which is in line with the expectations that can be derived from related work [8, 93, 114].

Media coverage and public discourse around AI have likely played a major role in shaping these shifting attitudes. The period between our two surveys (2022-2023) saw extensive media reporting on AI in Switzerland, as we show in Figure 2. News stories highlighting hallucinations, biases, and other failures of large language models likely contributed to the increased skepticism we observed, as media framing of AI can significantly influence public perceptions [37, 65, 110]. We suggest that the differential shifts in AI acceptance between Swiss linguistic regions can be partially attributed to differences in how AI was framed across these information environments. While our observations on these differential shifts may seem particularly relevant to officially multilingual countries like Switzerland or Belgium, they also apply to officially monolingual contexts, such as the United States, where a significant portion of the population communicates primarily in languages other than English, such as Spanish. Future work should systematically analyze such media coverage to better understand these mechanisms.

5.3 Socio-demographic differences in AI acceptance

The surge in interest and usage of AI technologies, exemplified by ChatGPT, has not only altered public perceptions generally but also introduced specific challenges related to the potential reinforcement of existing social inequalities. The educational divide in AI acceptance suggests potential knowledge or exposure gaps that could exacerbate existing digital inequalities. Those with higher education levels demonstrated greater acceptance of AI in decision-making contexts, potentially reflecting greater familiarity with technological systems or higher perceptions of AI usefulness for decision-making – a tendency previously found by Araujo et al. [6]. Our findings suggest that educational disparities may create uneven capabilities to understand, evaluate, and potentially benefit from AI systems. This observation mirrors broader patterns of digital inequality in Switzerland, where education strongly predicts internet use, skills and privacy-protective behaviors [25, 49, 69]. Büchi et al. [25] found that lower education is associated with decreased ability to protect one's privacy online, and Kappeler et al. [69] showed that lower-educated people are less likely to be internet users, with this gap persisting between 2011 and 2019.

While a gender gap in AI acceptance has been reported by existing studies [6, 33, 66, 96, 109], we find that it increased further with the GenAI boom, particularly in health-related contexts. This pattern may reflect gendered experiences with technology, different risk assessments with regard to technology [6], or responses to the documented gender biases in many AI systems [101]. These gender differences also align with findings on digital inequality in Switzerland more broadly [25, 49, 69]. Future research should investigate the underlying mechanisms of these disparities and identify ways to mitigate them, focusing on why certain demographic groups exhibit disparate preferences when it comes to AI use in high-impact scenarios.

The demographic disparities in AI acceptance we observed align with existing digital divides in Switzerland [25, 49, 69], suggesting that AI technologies may become another domain where digital inequalities are manifested along the lines of gender and education. The widening of these gaps following the generative AI boom is particularly concerning, as it indicates that rapid technological advancement may exacerbate rather than ameliorate existing inequalities. As AI systems become more integrated into essential services and opportunities, these acceptance disparities could translate into uneven adoption and benefit, further disadvantaging already marginalized groups. Addressing these emerging AI divides requires coordinated efforts from both researchers and policymakers: researchers should investigate mechanisms underlying these disparities and develop inclusive design approaches, while policymakers should consider educational interventions and regulatory frameworks that ensure equitable access to AI literacy and benefits across demographic groups.

5.4 Implications for AI governance and regulation

Our findings on shifting public attitudes following the generative AI boom also highlight a policy-related challenge: how to incorporate public opinion into governance frameworks for rapidly evolving technologies [91]. The significant changes in AI acceptance we observed between 2022 and 2023 demonstrate how quickly public sentiment can evolve in response to technological developments, raising important questions about the appropriate role of public opinion in AI governance. This timing challenge is further complicated by the fact that regulation typically takes years to develop and implement. For example, the EU AI Act was initially proposed in 2021 but only finalized in 2024 [121], during which time public opinion and AI capabilities both evolved significantly.

These challenges suggest the need for more adaptive regulatory frameworks that can evolve with public opinion and technological capabilities. While the specifics of such frameworks should be discussed in-depth by legal and policy scholars, based on relevant research we anticipate the following approaches to be particularly useful for the incorporation of public opinion into AI governance and regulation. First, one could create participatory governance structures that would establish formal roles for diverse public stakeholders in ongoing oversight of AI systems, potentially through citizen advisory boards or similar mechanisms, as discussed by [35, 89]. Second, regulatory sandboxes providing controlled environments where AI applications can be tested with public input before wider deployment could be helpful – similar to approaches used in fintech regulation. See [99] for a detailed discussion of this approach in the context of AI.

Another key insight from our study is that public acceptance varies significantly by context, even among applications that might all be categorized as "high-risk" under frameworks like the EU AI

Act [121]. This suggests that regulatory approaches that differentiate more finely between contexts might better align with public expectations. For example, while the EU AI Act categorizes both healthcare AI and HR AI systems as "high-risk," our findings show that public acceptance differs between medical diagnosis (with higher acceptance) and hiring decisions (with lower acceptance). These nuances could inform more contextually appropriate requirements for human oversight, explainability, and other safeguards.

Additionally, our findings on socio-demographic differences in AI acceptance highlight potential inequities in how AI governance may be perceived. The growing educational divide and gender gap in AI acceptance suggest that regulatory approaches may need to specifically address the concerns of groups showing lower acceptance to build broader public trust.

5.5 Limitations and future work

As with any research, our work is not without limitations that offer important opportunities for future research. While our study highlights issues related to context-dependent AI acceptance and the amplification of socio-demographic differences, it does not clarify the causal mechanisms behind these trends. Future research should investigate these underlying mechanisms further, offering policymakers and industry professionals actionable insights to promote responsible AI adoption and equitable outcomes.

Due to the constraints on questionnaire length, we could not fully describe the nuances of the different high-impact scenarios. Consequently, we cannot be certain that our priming of the AI definition worked so that all participants formed equivalent mental models of AI-based decisions [9, 26, 83]. Elements such as AI prediction performance and its influence on human decisions may also play a role [24, 67, 80], along with the types of biases present [13, 106], especially in high-impact contexts. Given these limitations, our survey-based study is subject to classic issues like social desirability bias, as it does not involve behavioral experiments.

An additional limitation with regards to the AI primer is that our study employed a specific operational definition of AI that emphasized current applications and explicitly distanced itself from more speculative or futuristic conceptions often portrayed in popular media. Public conceptions of AI are diverse and often include these more advanced or general AI systems [98]. Our chosen framing may have influenced participants' responses by potentially narrowing their consideration to more mundane or familiar technologies, which could lead to different attitudes than if participants were considering more advanced AI capabilities. This specific framing may have additionally introduced or strengthened the effects of social desirability bias by implicitly suggesting that concerns about more advanced AI systems are "unrealistic." Future work should consider comparing responses across different AI definitions or allowing participants to respond based on their own conception of AI.

Our methodology faces additional limitations regarding demographic data collection and literacy measurement. As demonstrated by Danaher and Crandall [38], the timing of demographic questions can induce stereotype threat effects that potentially influence results. Although we made deliberate choices to ensure consistency in our self-reported literacy measures between survey waves, these measures lack objective validation and may be subject to reporting biases. Similarly, our ChatGPT awareness assessment lacks comparative baselines against fictional technologies, making it difficult to distinguish between genuine knowledge and mere name recognition. Future research should employ validated measures with objective components and control for these methodological limitations.

Finally, even when based on representative samples, our findings may not be generalizable across countries due to the wide variation in AI acceptance across cultural and national contexts [51, 72].

5.5.1 Methodological considerations for future research on AI acceptance. Our study contributes to research on public perceptions of AI technologies but differs methodologically from much HCI literature, which often relies on experimental designs with specific user groups. This raises questions about how AI acceptance, trust, reliability, and other related constructs should be measured across research contexts.

While our study focused primarily on "acceptance," measured through survey questions rather than experimental designs, the HCI literature has developed a rich ecosystem of related but distinct constructs that capture different aspects of how people perceive and interact with AI systems, measured through experiments. These include, for example, trust [77, 123] or fairness [42, 97]. When it comes to AI acceptance specifically, experimental HCI designs primarily focus on the way different design factors such as various forms of explainability of AI decisions affect user acceptance of AI [43, 80, 90].

These methodological differences reflect distinct research objectives. Our survey approach captures population-level attitudes to inform policy, while experimental HCI research isolates factors influencing user experience to improve design. This generalizability versus specificity tradeoff highlights their complementary nature.

Our context-dependent findings suggest opportunities to bridge these approaches. Experimental designs could investigate why acceptance varies across contexts by manipulating factors like decision importance and domain. For example, the lower acceptance in hiring versus medical contexts could be explored through experiments controlling for perceived reliability, fairness, and explainability.

The demographic differences we observed – particularly in education and gender – suggest experimental studies should use diverse samples and analyze interaction effects between demographic factors and design interventions. Such studies might examine whether different forms of explainability affect acceptance differently across demographic groups.

Both approaches face ecological validity challenges. Surveys rely on hypothetical scenarios that may not reflect actual behavior with AI systems, while experiments often use simplified tasks that do not capture real-world complexity. Field experiments deploying AI systems to diverse populations in natural settings could help address these limitations.

The rapid shift in AI acceptance following the generative AI boom highlights the temporal instability of public perceptions, suggesting both research traditions should incorporate longitudinal designs to capture evolving perceptions. By integrating insights from both approaches, researchers can develop more nuanced understandings of public AI perceptions and more effective system designs.

6 Conclusion

In this study, we leveraged a survey approach to capture public attitudes towards AI. Our study design is unique as we conducted two survey waves, one before and one after the GenAI boom, revealing a decreased AI acceptance coupled with a pronounced preference for humans to be in/on the loop. Our findings indicate a marked decline in public acceptance of AI, coupled with a heightened demand for human oversight, particularly in sensitive applications where the implications of decisions are profound. Notably, the GenAI boom is significantly associated with amplified existing disparities, with acceptance levels diverging along the lines of education, language, and gender. These insights underscore the importance of context-sensitive AI developments and regulation frameworks ensuring appropriate levels of human oversight rather than adopting blanket policies.

Acknowledgements

This work was supported by the National Research Programme "Digital Transformation" (NRP 77) of the Swiss National Science Foundation (SNSF), grant number 187473.

References

[1] Hussein A Abbass. Social integration of artificial intelligence: functions, automation allocation logic and human-autonomy trust. *Cognitive Computation*, 11(2):159–171, 2019.

[2] Daron Acemoglu. Technology and inequality. *NBER Reporter Online*, (Winter 2002/03):12–16, 2002.

[3] Julia Angwin, Jeff Larson, Surya Mattu, and Lauren Kirchner. Machine bias. *ProPublica, May*, 23(2016):139–159, 2016. URL https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing.

[4] Alison L. Antes, Sara Burrous, Bryan A. Sisk, Matthew J. Schuelke, Jason D. Keune, and James M. DuBois. Exploring perceptions of healthcare technologies enabled by artificial intelligence: an online, scenario-based survey. *BMC Medical Informatics and Decision Making*, 21(1):221, July 2021. ISSN 1472-6947. doi: 10.1186/s12911-021-01586-8. URL https://doi.org/10.1186/s12911-021-01586-8.

[5] Naomi Aoki. The importance of the assurance that "humans are still in the decision loop" for public trust in artificial intelligence: Evidence from an online experiment. *Computers in Human Behavior*, 114:106572, 2021.

[6] Theo Araujo, Natali Helberger, Sanne Kruikemeier, and Claes H. de Vreese. In AI we trust? Perceptions about automated decision-making by artificial intelligence. *AI & SOCIETY*, 35(3):611–623, September 2020. ISSN 1435-5655. doi: 10.1007/s00146-019-00931-w. URL https://doi.org/10.1007/s00146-019-00931-w.

[7] Yuri Y M Aung, David C S Wong, and Daniel S W Ting. The promise of artificial intelligence: a review of the opportunities and challenges of artificial intelligence in healthcare. *British Medical Bulletin*, 139(1):4–15, 08 2021. ISSN 0007-1420. doi: 10.1093/bmb/ldab016. URL https://doi.org/10.1093/bmb/ldab016.

[8] Mohammad Babamiri, Rashid Heidarimoghadam, Fakhradin Ghasemi, Leili Tapak, and Alireza Mortezapour. Insights into the relationship between usability and willingness to use a robot in the future workplaces: Studying the mediating role of trust and the moderating roles of age and stara. *PLOS ONE*, 17(6):1–12, 06 2022. doi: 10.1371/journal.pone.0268942. URL https://doi.org/10.1371/journal.pone.0268942.

[9] Gagan Bansal, Besmira Nushi, Ece Kamar, Walter S. Lasecki, Daniel S. Weld, and Eric Horvitz. Beyond accuracy: The role of mental models in human-ai team performance. *Proceedings of the AAAI Conference on Human Computation and Crowdsourcing*, 7(1):2–11, Oct. 2019. URL https://ojs.aaai.org/index.php/HCOMP/article/view/5285.

[10] Gagan Bansal, Tongshuang Wu, Joyce Zhou, Raymond Fok, Besmira Nushi, Ece Kamar, Marco Tulio Ribeiro, and Daniel Weld. Does the whole exceed its parts? the effect of ai explanations on complementary team performance. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, CHI '21, New York, NY, USA, 2021. Association for Computing Machinery. ISBN 9781450380966. doi: 10.1145/3411764.3445717. URL https://doi.org/10.1145/3411764.3445717.

[11] Luye Bao, Nicole M. Krause, Mikhaila N. Calice, Dietram A. Scheufele, Christopher D. Wirz, Dominique Brossard, Todd P. Newman, and Michael A. Xenos. Whose AI? How different publics think about AI and its social impacts. *Computers in Human Behavior*, 130:107182, May 2022. ISSN 0747-5632. doi: 10.1016/j.chb.2022.107182. URL https://www.sciencedirect.com/science/article/pii/S0747563222000048.

[12] Joachim Baumann and Michele Loi. Fairness and risk: an ethical argument for a group fairness definition insurers can use. *Philosophy & Technology*, 36(3):45, 2023.

[13] Joachim Baumann, Alessandro Castelnovo, Riccardo Crupi, Nicole Inverardi, and Daniele Regoli. Bias on demand: A modelling framework that generates synthetic data with bias. In *Proceedings of the 2023 ACM Conference on Fairness, Accountability, and Transparency*, page 1002–1013, 2023. URL https://doi.org/10.1145/3593013.3594058.

[14] Lanthao Benedikt, Chaitanya Joshi, Louisa Nolan, Ruben Henstra-Hill, Luke Shaw, and Sharon Hook. Human-in-the-loop ai in government: a case study. In *Proceedings of the 25th International Conference on Intelligent User Interfaces*, IUI '20, page 488–497, New York, NY, USA, 2020. Association for Computing Machinery. ISBN 9781450371186. doi: 10.1145/3377325.3377489. URL https://doi.org/10.1145/3377325.3377489.

[15] Yoav Benjamini and Yosef Hochberg. Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society: Series B (Methodological)*, 57(1):289–300, 1995. doi: https://doi.org/10.1111/j.2517-6161.1995.tb02031.x. URL https://rss.onlinelibrary.wiley.com/doi/abs/10.1111/j.2517-6161.1995.tb02031.x.

[16] Richard Berk. An impact assessment of machine learning risk forecasts on parole board decisions and recidivism. *Journal of Experimental Criminology*, 13:193–216, 2017.

[17] Aarushi Bhandari. Gender inequality in mobile technology access: the role of economic and social development*. *Information, Communication & Society*, 22(5):678–694, 2019. doi: 10.1080/1369118X.2018.1563206. URL https://doi.org/10.1080/1369118X.2018.1563206.

[18] Joseph R Biden. Executive order on the safe, secure, and trustworthy development and use of artificial intelligence. 2023. URL https://www.whitehouse.gov/briefing-room/presidential-actions/2023/10/30/executive-order-on-the-safe-secure-and-trustworthy-development-and-use-of-artificial-intelligence/.

[19] Microsoft Bing. Introducing bing generative search. 2024. URL https://blogs.bing.com/search/July-2024/generativesearch.

[20] Sina Blassnig, Edina Strikovic, Eliza Mitova, Aleksandra Urman, Anikó Hannák, Claes de Vreese, and Frank Esser. A balancing act: How media professionals perceive the implementation of news recommender systems. *Digital Journalism*, 0(0):1–23, 2024. doi: 10.1080/21670811.2023.2293933. URL https://doi.org/10.1080/21670811.2023.2293933.

[21] Thomas Blomberg, William Bales, Karen Mann, Ryan Meldrum, and Joe Nedelec. Validation of the compas risk assessment classification instrument. *College of Criminology and Criminal Justice, Florida State University, Tallahassee, FL*, 2010.

[22] Miranda Bogen and Aaron Rieke. Help wanted: An examination of hiring algorithms, equity, and bias. *Upturn, December*, 7, 2018.

[23] Brandon M. Booth, Louis Hickman, Shree Krishna Subburaj, Louis Tay, Sang Eun Woo, and Sidney K. D'Mello. Bias and fairness in multimodal machine learning: A case study of automated video interviews. In *Proceedings of the 2021 International Conference on Multimodal Interaction*, ICMI '21, page 268–277, New York, NY, USA, 2021. Association for Computing Machinery. ISBN 9781450384810. doi: 10.1145/3462244.3479897. URL https://doi.org/10.1145/3462244.3479897.

[24] Adrian Bussone, Simone Stumpf, and Dympna O'Sullivan. The role of explanations on trust and reliance in clinical decision support systems. In *2015 International Conference on Healthcare Informatics*, pages 160–169, 2015. doi: 10.1109/ICHI.2015.26.

[25] Moritz Büchi, Noemi Festic, Natascha Just, and Michael Latzer. Chapter 20: Digital inequalities in online privacy protection: effects of age, education and gender. November 2021. ISBN 978-1-78811-657-2. URL https://www.elgaronline.com/edcollchap/edcoll/9781788116565/9781788116565.00029.xml. Section: Handbook of Digital Inequality.

[26] Ángel Alexander Cabrera, Adam Perer, and Jason I. Hong. Improving human-ai collaboration with descriptions of ai behavior. *Proc. ACM Hum.-Comput. Interact.*, 7(CSCW1), April 2023. URL https://doi.org/10.1145/3579612.

[27] Zhihui Cai, Xitao Fan, and Jianxia Du. Gender and attitudes toward technology use: A meta-analysis. *Computers & Education*, 105:1–13, February 2017. ISSN 0360-1315. doi: 10.1016/j.compedu.2016.11.003. URL https://www.sciencedirect.com/science/article/pii/S0360131516302007.

[28] Noah Castelo, Maarten W. Bos, and Donald R. Lehmann. Task-Dependent Algorithm Aversion. *Journal of Marketing Research*, 56(5):809–825, October 2019. ISSN 0022-2437. doi: 10.1177/0022243719851788. URL https://doi.org/10.1177/0022243719851788. Publisher: SAGE Publications Inc.

[29] Stephen Cave, Kanta Dihal, Sarah Dillon, Stephen Cave, Kanta Dihal, and Sarah Dillon, editors. *AI Narratives: A History of Imaginative Thinking about Intelligent Machines.* Oxford University Press, Oxford, New York, March 2020. ISBN 978-0-19-884666-6.

[30] Alberto Cevolini and Elena Esposito. From pool to profile: Social consequences of algorithmic prediction in insurance. *Big Data & Society*, 7(2):2053951720939228, 2020. doi: 10.1177/2053951720939228. URL https://doi.org/10.1177/2053951720939228.

[31] Mingyang Chen, Bo Zhang, Ziting Cai, Samuel Seery, Maria J Gonzalez, Nasra M Ali, Ran Ren, Youlin Qiao, Peng Xue, and Yu Jiang. Acceptance of clinical artificial intelligence among physicians and medical students: a systematic review with cross-sectional survey. *Frontiers in medicine*, 9:990604, 2022.

[32] Lingwei Cheng and Alexandra Chouldechova. Overcoming algorithm aversion: A comparison between process and outcome control. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, CHI '23, New York, NY, USA, 2023. Association for Computing Machinery. ISBN 9781450394215. doi: 10.1145/3544548.3581253. URL https://doi.org/10.1145/3544548.3581253.

[33] Kyong Ah Cho and Yon Hee Seo. Dual mediating effects of anxiety to use and acceptance attitude of artificial intelligence technology on the relationship between nursing students' perception of and intention to use them: a descriptive study. *BMC Nursing*, 23(1):212, March 2024. ISSN 1472-6955. doi: 10.1186/s12912-024-01887-z. URL https://doi.org/10.1186/s12912-024-01887-z.

[34] Murari Choudhary, Shashank Jha, Deepika Saxena, Ashutosh Kumar Singh, et al. A review of fake news detection methods using machine learning. In *2021 2nd international conference for emerging technology (INCET)*, pages 1–5. IEEE, 2021.

[35] Peter Cihon, Matthijs M. Maas, and Luke Kemp. Fragmentation and the Future: Investigating Architectures for International AI Governance. *Global Policy*, 11(5):545–556, 2020. ISSN 1758-5899. doi: 10.1111/1758-5899.12890. URL https://onlinelibrary.wiley.com/doi/abs/10.1111/1758-5899.12890. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/1758-5899.12890.

[36] James Clayton. Ces 2024: Ai pillows and toothbrushes-is it all getting a bit silly. *BBC (British Broadcasting Corporation)*, 2024. URL https://www.bbc.com/news/technology-67959240.

[37] Di Cui and Fang Wu. The influence of media use on public perceptions of artificial intelligence in China: Evidence from an online survey. *Information Development*, 37(1):45–57, March 2021. ISSN 0266-6669. doi: 10.1177/0266666919893411. URL https://doi.org/10.1177/0266666919893411. Publisher: SAGE Publications Ltd.

[38] Kelly Danaher and Christian S Crandall. Stereotype threat in applied settings re-examined. *Journal of Applied Social Psychology*, 38(6):1639–1655, 2008.

[39] Thomas Davenport and Ravi Kalakota. The potential for artificial intelligence in healthcare. *Future Healthcare Journal*, 6(2):94–98, 2019. ISSN 2514-6645. doi: https://doi.org/10.7861/futurehosp.6-2-94. URL https://www.sciencedirect.com/science/article/pii/S2514664524010592.

[40] Fred D. Davis. Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS Quarterly*, 13(3):319–340, 1989. URL http://www.jstor.org/stable/249008.

[41] Berkeley J. Dietvorst, Joseph P. Simmons, and Cade Massey. Overcoming algorithm aversion: People will use imperfect algorithms if they can (even slightly) modify them. *Management Science*, 64(3):1155–1170, 2018. doi: 10.1287/mnsc.2016.2643. URL https://doi.org/10.1287/mnsc.2016.2643.

[42] Jonathan Dodge, Q. Vera Liao, Yunfeng Zhang, Rachel K. E. Bellamy, and Casey Dugan. Explaining models: an empirical study of how explanations impact fairness judgment. In *Proceedings of the 24th International Conference on Intelligent User Interfaces*, IUI '19, pages 275–285, New York, NY, USA, March 2019. Association for Computing Machinery. ISBN 978-1-4503-6272-6. doi: 10.1145/3301275.3302310. URL https://dl.acm.org/doi/10.1145/3301275.3302310.

[43] Carolin Ebermann, Selisky , Matthias, , and Stephan Weibelzahl. Explainable AI: The Effect of Contradictory Decisions and Explanations on Users' Acceptance of AI Systems. *International Journal of Human–Computer Interaction*, 39(9): 1807–1826, May 2023. ISSN 1044-7318. doi: 10.1080/10447318.2022.2126812. URL https://doi.org/10.1080/10447318.2022.2126812. Publisher: Taylor & Francis _eprint: https://doi.org/10.1080/10447318.2022.2126812.

[44] Edelman. 2019 edelman trust barometer (trust in healthcare: Global), 2019. URL https://www.edelman.com/sites/g/files/aatuss191/files/2019-04/2019_Edelman_Trust_Barometer_Global_Health_Report.pdf.

[45] European Parliament. Artificial Intelligence Act, 2024. URL https://www.europarl.europa.eu/doceo/document/TA-9-2024-0138-FNL-COR01_EN.pdf.

[46] Federal Statistical Office Switzerland. Statistical atlas of switzerland: The 4 language areas of switzerland by municipality, 2020. URL https://www.atlas.bfs.admin.ch/maps/13/de/17138_17137_235_227/26599.html.

[47] Federal Statistical Office Switzerland. Population census, 2022. URL https://www.bfs.admin.ch/bfs/en/home/statistics/population/surveys/se.html.

[48] Federal Statistical Office Switzerland. Permanent and non-permanent resident population by institutional units, citizenship (category), sex and age, 2010-2023, 2023. URL https://www.pxweb.bfs.admin.ch/pxweb/en/px-x-0102010000_101/-/px-x-0102010000_101.px/.

[49] Noemi Festic, Moritz Büchi, and Michael Latzer. It's still a thing: digital inequalities and their evolution in the information society. *Studies in Communication and Media*, 10(3):326–361, 2021. ISSN 2192-4007. doi: 10.5771/2192-4007-2021-3-326. URL https://www.nomos-elibrary.de/index.php?doi=10.5771/2192-4007-2021-3-326.

[50] Joel E. Fischer, Chris Greenhalgh, Wenchao Jiang, Sarvapali D. Ramchurn, Feng Wu, and Tom Rodden. In-the-loop or on-the-loop? interactional arrangements to support team coordination with a planning agent. *Concurrency and Computation: Practice and Experience*, 33(8):e4082, 2021. doi: https://doi.org/10.1002/cpe.4082.

[51] Richard Fletcher and Rasmus Kleis Nielsen. What does the public in six countries think of generative AI in news? | Reuters Institute for the Study of Journalism, 2024. URL https://reutersinstitute.politics.ox.ac.uk/what-does-public-six-countries-think-generative-ai-news.

[52] C Funk, A Tyson, B Kennedy, and C Johnson. Publics express a mix of views on ai, childhood vaccines, food and space issues. *Pew Research Center*, 2020.

[53] Andreas Fuster, Paul Goldsmith-Pinkham, Tarun Ramadorai, and Ansgar Walther. Predictably unequal? the effects of machine learning on credit markets. *The Journal of Finance*, 77(1):5–47, 2022. doi: https://doi.org/10.1111/jofi.13090. URL https://onlinelibrary.wiley.com/doi/abs/10.1111/jofi.13090.

[54] David Garcia, Yonas Mitike Kassa, Angel Cuevas, Manuel Cebrian, Esteban Moro, Iyad Rahwan, and Ruben Cuevas. Analyzing gender inequality through large-scale facebook advertising data. *Proceedings of the National Academy of Sciences*, 115(27):6958–6963, 2018. doi: 10.1073/pnas.1717781115. URL https://www.pnas.org/doi/abs/10.1073/pnas.1717781115.

[55] Juan María González-Anleo, Luca Delbello, José Maria Martínez-Gonzálo, and Andres Gómez. Sociodemographic Impact on the Adoption of Emerging Technologies. *Journal of Small Business Strategy*, 34(2):42–50, sep 3 2024. doi: 10.53703/001c.122089.

[56] Ben Green and Yiling Chen. The principles and limits of algorithm-in-the-loop decision making. *Proc. ACM Hum.-Comput. Interact.*, 3(CSCW), November 2019. doi: 10.1145/3359152. URL https://doi.org/10.1145/3359152.

[57] Tor Grønsund and Margunn Aanestad. Augmenting the algorithm: Emerging human-in-the-loop work configurations. *The Journal of Strategic Information Systems*, 29(2):101614, 2020. Strategic Perspectives on Digital Work and Organizational Transformation.

[58] Dogan Gursoy, Oscar Hengxuan Chi, Lu Lu, and Robin Nunkoo. Consumers acceptance of artificially intelligent (ai) device use in service delivery. *International Journal of Information Management*, 49:157–169, 2019. ISSN 0268-4012. doi: https://doi.org/10.1016/j.ijinfomgt.2019.03.008. URL https://www.sciencedirect.com/science/article/pii/S0268401219301690.

[59] Saqib Hakak, Mamoun Alazab, Suleman Khan, Thippa Reddy Gadekallu, Praveen Kumar Reddy Maddikunta, and Wazir Zada Khan. An ensemble machine learning approach through effective feature extraction to classify fake news. *Future Generation Computer Systems*, 117:47–58, 2021. ISSN 0167-739X. doi: https://doi.org/10.1016/j.future.2020.11.022. URL https://www.sciencedirect.com/science/article/pii/S0167739X20330466.

[60] Brian Hall. 185 real-world gen ai use cases from the world's leading organizations. 2024. URL https://cloud.google.com/transform/101-real-world-generative-ai-use-cases-from-industry-leaders.

[61] Eszter Hargittai and Yuli Patrick Hsieh. Succinct survey measures of web-use skills. *Social Science Computer Review*, 30(1):95–107, 2012. doi: 10.1177/0894439310397146. URL https://doi.org/10.1177/0894439310397146.

[62] Drew Harwell. A face-scanning algorithm increasingly decides whether you deserve the job. In *Ethics of Data and Analytics*, pages 206–211. Auerbach Publications, 2022.

[63] Fadi Sakka Hicham Sadok and Mohammed El Hadi El Maknouzi. Artificial intelligence and bank credit analysis: A review. *Cogent Economics & Finance*, 10(1):2023262, 2022. doi: 10.1080/23322039.2021.2023262. URL https://doi.org/10.1080/23322039.2021.2023262.

[64] Manh-Tung Ho, Ngoc-Thang B. Le, Peter Mantello, Manh-Toan Ho, and Nader Ghotbi. Understanding the acceptance of emotional artificial intelligence in japanese healthcare system: A cross-sectional survey of clinic visitors' attitude. *Technology in Society*, 72:102166, 2023. ISSN 0160-791X. doi: https://doi.org/10.1016/j.techsoc.2022.102166.

[65] Shirley S. Ho and Justin C. Cheung. Trust in artificial intelligence, trust in engineers, and news media: Factors shaping public perceptions of autonomous drones through UTAUT2. *Technology in Society*, 77:102533, June 2024. ISSN 0160-791X. doi: 10.1016/j.techsoc.2024.102533. URL https://www.sciencedirect.com/science/article/pii/S0160791X24000812.

[66] Michael C. Horowitz and Lauren Kahn. What influences attitudes about artificial intelligence adoption: Evidence from U.S. local officials. *PLOS ONE*, 16(10):e0257732, October 2021. ISSN 1932-6203. doi: 10.1371/journal.pone.0257732. URL https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0257732. Publisher: Public Library of Science.

[67] Maia Jacobs, Melanie F Pradier, Thomas H McCoy Jr, Roy H Perlis, Finale Doshi-Velez, and Krzysztof Z Gajos. How machine-learning recommendations influence clinician treatment selections: the example of antidepressant selection. *Translational psychiatry*, 11(1):108, 2021.

[68] Ekaterina Jussupow, Izak Benbasat, and Armin Heinzl. WHY ARE WE AVERSE TOWARDS ALGORITHMS? A COMPREHENSIVE LITERATURE REVIEW ON ALGORITHM AVERSION. *ECIS 2020 Research Papers*, June 2020. URL https://aisel.aisnet.org/ecis2020_rp/168.

[69] Kiran Kappeler, Noemi Festic, and Michael Latzer. Left behind in the digital society – Growing social stratification of internet non-use in Switzerland. In Guido Keel and Wibke Weber, editors, *Kappeler, Kiran; Festic, Noemi; Latzer, Michael (2021). Left behind in the digital society – Growing social stratification of internet non-use in Switzerland. In: Keel, Guido; Weber, Wibke. Media Literacy. Baden-Baden: Nomos, 207-224.*, pages 207–224. Nomos, Baden-Baden, 2021. ISBN 978-3-8487-8265-9. doi: 10.5771/9783748920656. URL https://www.zora.uzh.ch/id/eprint/216454/.

[70] Elena Karahanna, Detmar W. Straub, and Norman L. Chervany. Information technology adoption across time: A cross-sectional comparison of pre-adoption and post-adoption beliefs. *MIS Quarterly*, 23(2):183–213, 1999. URL http://www.jstor.org/stable/249751.

[71] Enkelejda Kasneci, Kathrin Sessler, Stefan Küchemann, Maria Bannert, Daryna Dementieva, Frank Fischer, Urs Gasser, Georg Groh, Stephan Günnemann, Eyke Hüllermeier, Stephan Krusche, Gitta Kutyniok, Tilman Michaeli, Claudia Nerdel, Jürgen Pfeffer, Oleksandra Poquet, Michael Sailer, Albrecht Schmidt, Tina Seidel, Matthias Stadler, Jochen Weller, Jochen Kuhn, and Gjergji Kasneci. Chatgpt for good? on opportunities and challenges of large language models for education. *Learning and Individual Differences*, 103:102274, 2023. ISSN 1041-6080. doi: https://doi.org/10.1016/j.lindif.2023.102274. URL https://www.sciencedirect.com/science/article/pii/S1041608023000195.

[72] Esther Kaufmann, Alvaro Chacon, Edgar E. Kausel, Nicolas Herrera, and Tomas Reyes. Task-specific algorithm advice acceptance: A review and directions for future research. *Data and Information Management*, 7(3):100040, September 2023. ISSN 2543-9251. doi: 10.1016/j.dim.2023.100040. URL https://www.sciencedirect.com/science/article/pii/S2543925123000141.

[73] Sage Kelly, Sherrie-Anne Kaye, and Oscar Oviedo-Trespalacios. What factors contribute to the acceptance of artificial intelligence? A systematic review. *Telematics and Informatics*, 77:101925, February 2023. ISSN 07365853. doi: 10.1016/j.tele.2022.101925. URL https://linkinghub.elsevier.com/retrieve/pii/S0736585322001587.

[74] Zeba Khanam, BN Alwasel, H Sirafi, and Mamoon Rashid. Fake news detection using machine learning approaches. *IOP Conference Series: Materials Science and Engineering*, 1099(1):012040, mar 2021. doi: 10.1088/1757-899X/1099/1/012040. URL https://dx.doi.org/10.1088/1757-899X/1099/1/012040.

[75] Jihyun Kim, Kelly Merrill, Kun Xu, and Deanna D. Sellnow. My Teacher Is a Machine: Understanding Students' Perceptions of AI Teaching Assistants in Online Education. *International Journal of Human–Computer Interaction*, 36 (20):1902–1911, December 2020. ISSN 1044-7318. doi: 10.1080/10447318.2020.1801227. URL https://doi.org/10.1080/10447318.2020.1801227.

[76] Taenyun Kim, Maria D. Molina, Minjin (MJ) Rheu, Emily S. Zhan, and Wei Peng. One ai does not fit all: A cluster analysis of the laypeople's perception of ai roles. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, CHI '23, New York, NY, USA, 2023. Association for Computing Machinery. ISBN 9781450394215. doi: 10.1145/3544548.3581340. URL https://doi.org/10.1145/3544548.3581340.

[77] Artur Klingbeil, Cassandra Grützner, and Philipp Schreck. Trust and reliance on AI — An experimental study on the extent and costs of overreliance on AI. *Computers in Human Behavior*, 160:108352, November 2024. ISSN 0747-5632. doi: 10.1016/j.chb.2024.108352. URL https://www.sciencedirect.com/science/article/pii/S0747563224002206.

[78] Will Knight. Google's gemini is the real start of the generative ai boom, December 2023. URL https://www.wired.com/story/google-gemini-generative-ai-boom/.

[79] Alina Köchling and Marius Claus Wehner. Discriminated by an algorithm: a systematic review of discrimination and fairness by algorithmic decision-making in the context of hr recruitment and hr development. *Business Research*, 13 (3):795–848, 2020.

[80] Rafal Kocielnik, Saleema Amershi, and Paul N. Bennett. Will You Accept an Imperfect AI? Exploring Designs for Adjusting End-user Expectations of AI Systems. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, CHI '19, pages 1–14, New York, NY, USA, May 2019. Association for Computing Machinery. ISBN 978-1-4503-5970-2. doi: 10.1145/3290605.3300641. URL https://doi.org/10.1145/3290605.3300641.

[81] Nikita Kozodoi, Johannes Jacob, and Stefan Lessmann. Fairness in credit scoring: Assessment, implementation and profit implications. *European Journal of Operational Research*, 297(3):1083–1094, 2022. ISSN 0377-2217. doi: https://doi.org/10.1016/j.ejor.2021.06.023. URL https://www.sciencedirect.com/science/article/pii/S0377221721005385.

[82] P. M. Krafft, Meg Young, Michael Katell, Karen Huang, and Ghislain Bugingo. Defining AI in Policy versus Practice. In *Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society*, AIES '20, pages 72–78, New York, NY, USA, February 2020. Association for Computing Machinery. ISBN 978-1-4503-7110-0. doi: 10.1145/3375627.3375835. URL https://doi.org/10.1145/3375627.3375835.

[83] Todd Kulesza, Simone Stumpf, Margaret Burnett, and Irwin Kwan. Tell me more? the effects of mental model soundness on personalizing an intelligent agent. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '12, page 1–10, New York, NY, USA, 2012. Association for Computing Machinery. ISBN 9781450310154. URL https://doi.org/10.1145/2207676.2207678.

[84] Hannah Labinsky, Dubravka Ukalovic, Fabian Hartmann, Vanessa Runft, André Wichmann, Jan Jakubcik, Kira Gambel, Katharina Otani, Harriet Morf, Jule Taubmann, Filippo Fagni, Arnd Kleyer, David Simon, Georg Schett, Matthias Reichert, and Johannes Knitza. An ai-powered clinical decision support system to predict flares in rheumatoid arthritis: A pilot study. *Diagnostics*, 13(1), 2023. ISSN 2075-4418. doi: 10.3390/diagnostics13010148. URL https://www.mdpi.com/2075-4418/13/1/148.

[85] Anton Danholt Lautrup, Tobias Hyrup, Anna Schneider-Kamp, Marie Dahl, Jes Sanddal Lindholt, and Peter Schneider-Kamp. Heart-to-heart with chatgpt: the impact of patients consulting ai for cardiovascular health advice. *Open heart*, 10(2), 2023.

[86] Michele Loi and Markus Christen. Choosing how to discriminate: Navigating ethical trade-offs in fair algorithmic design for the insurance sector. *Philosophy & Technology*, 34(4):967–992, 2021.

[87] Thomas Lumley. survey: analysis of complex survey samples, 2024. R package version 4.4.

[88] Brady D Lund and Ting Wang. Chatting about chatgpt: how may ai and gpt impact academia and libraries? *Library hi tech news*, 40(3):26–29, 2023.

[89] Matthijs M. Maas. Aligning AI Regulation to Sociotechnical Change. In Justin B. Bullock, Yu-Che Chen, Johannes Himmelreich, Valerie M. Hudson, Anton Korinek, Matthew M. Young, and Baobao Zhang, editors, *The Oxford Handbook of AI Governance*, pages 358–380. Oxford University Press, 1 edition, March 2022. ISBN 978-0-19-757932-9 978-0-19-757935-0. doi: 10.1093/oxfordhb/9780197579329.013.22. URL https://academic.oup.com/edited-volume/41989/chapter/355438659.

[90] Akihiro Maehigashi, Yosuke Fukuchi, and Seiji Yamada. Experimental Investigation of Human Acceptance of AI Suggestions with Heatmap and Pointing-based XAI. In *Proceedings of the 11th International Conference on Human-Agent Interaction*, HAI '23, pages 291–298, New York, NY, USA, December 2023. Association for Computing Machinery. ISBN 9798400708244. doi: 10.1145/3623809.3623834. URL https://doi.org/10.1145/3623809.3623834.

[91] Gary Elvin Marchant, Braden R. Allenby, and Joseph R. Herkert. *Growing gap between emerging technologies and legal-ethical oversight: the pacing problem.* Number v. 7 in International library of ethics, law and technology. Springer, Dordrecht New York, 2011. ISBN 978-94-007-1356-7.

[92] Christian Marxt and Claudia Brunner. Analyzing and improving the national innovation system of highly developed countries — the case of switzerland. *Technological Forecasting and Social Change*, 80(6):1035–1049, 2013. ISSN 0040-1625. doi: https://doi.org/10.1016/j.techfore.2012.07.008. URL https://www.sciencedirect.com/science/article/pii/S0040162512001771.

[93] Kate K Mays, Yiming Lei, Rebecca Giovanetti, and James E Katz. Ai as a boss? a national us survey of predispositions governing comfort with expanded ai roles in society. *AI & SOCIETY*, pages 1–14, 2022.

[94] Roshni Modhvadia. How do people feel about ai? a nationally representative survey of public attitudes to artificial intelligence in britain. *The Ada Lovelace Institute*, 2023.

[95] Eduardo Mosqueira-Rey, Elena Hernández-Pereira, David Alonso-Ríos, José Bobes-Bascarán, and Ángel Fernández-Leal. Human-in-the-loop machine learning: a state of the art. *Artificial Intelligence Review*, 56(4):3005–3054, 2023.

[96] Mariano Méndez-Suárez, Luca Delbello, Alejandro de Vega de Unceta, and Ana Lucia Ortega Larrea. Factors Affecting Consumers' Attitudes Towards Artificial Intelligence. *Journal of Promotion Management*, 0(0):1–18, 2024. ISSN 1049-6491. doi: 10.1080/10496491.2024.2367203. URL https://doi.org/10.1080/10496491.2024.2367203.

[97] Yuri Nakao, Simone Stumpf, Subeida Ahmed, Aisha Naseer, and Lorenzo Strappelli. Toward Involving End-users in Interactive Human-in-the-loop AI Fairness. *ACM Trans. Interact. Intell. Syst.*, 12(3):18:1–18:30, July 2022. ISSN 2160-6455. doi: 10.1145/3514258. URL https://doi.org/10.1145/3514258.

[98] Arvind Narayanan and Sayash Kapoor. *AI Snake Oil: What Artificial Intelligence Can Do, What It Can't, and How to Tell the Difference.* Princeton University Press, Princeton Oxford, September 2024. ISBN 978-0-691-24913-1.

[99] OECD. Regulatory sandboxes in artificial intelligence, July 2023. URL https://www.oecd.org/en/publications/regulatory-sandboxes-in-artificial-intelligence_8f80a0e6-en.html.

[100] OpenAI, Josh Achiam, Steven Adler, Sandhini Agarwal, Lama Ahmad, Ilge Akkaya, Florencia Leoni Aleman, Diogo Almeida, Janko Altenschmidt, Sam Altman, Shyamal Anadkat, Red Avila, Igor Babuschkin, Suchir Balaji, Valerie Balcom, Paul Baltescu, Haiming Bao, Mohammad Bavarian, Jeff Belgum, Irwan Bello, Jake Berdine, Gabriel Bernadett-Shapiro, Christopher Berner, Lenny Bogdonoff, Oleg Boiko, Madelaine Boyd, Anna-Luisa Brakman, Greg Brockman, Tim Brooks, Miles Brundage, Kevin Button, Trevor Cai, Rosie Campbell, Andrew Cann, Brittany Carey, Chelsea Carlson, Rory Carmichael, Brooke Chan, Che Chang, Fotis Chantzis, Derek Chen, Sully Chen, Ruby Chen, Jason Chen, Mark Chen, Ben Chess, Chester Cho, Casey Chu, Hyung Won Chung, Dave Cummings, Jeremiah Currier, Yunxing Dai, Cory Decareaux, Thomas Degry, Noah Deutsch, Damien Deville, Arka Dhar, David Dohan, Steve Dowling, Sheila Dunning, Adrien Ecoffet, Atty Eleti, Tyna Eloundou, David Farhi, Liam Fedus, Niko Felix, Simón Posada Fishman, Juston Forte, Isabella Fulford, Leo Gao, Elie Georges, Christian Gibson, Vik Goel, Tarun Gogineni, Gabriel Goh, Rapha Gontijo-Lopes, Jonathan Gordon, Morgan Grafstein, Scott Gray, Ryan Greene, Joshua Gross, Shixiang Shane Gu, Yufei Guo, Chris Hallacy, Jesse Han, Jeff Harris, Yuchen He, Mike Heaton, Johannes Heidecke, Chris Hesse, Alan Hickey, Wade Hickey, Peter Hoeschele, Brandon Houghton, Kenny Hsu, Shengli Hu, Xin Hu, Joost Huizinga, Shantanu Jain, Shawn Jain, Joanne Jang, Angela Jiang, Roger Jiang, Haozhun Jin, Denny Jin, Shino Jomoto, Billie Jonn, Heewoo Jun, Tomer Kaftan, Łukasz Kaiser, Ali Kamali, Ingmar Kanitscheider, Nitish Shirish Keskar, Tabarak Khan, Logan Kilpatrick, Jong Wook Kim, Christina Kim, Yongjik Kim, Jan Hendrik Kirchner, Jamie Kiros, Matt Knight, Daniel Kokotajlo, Łukasz Kondraciuk, Andrew Kondrich, Aris Konstantinidis, Kyle Kosic, Gretchen Krueger, Vishal Kuo, Michael Lampe, Ikai Lan, Teddy Lee, Jan Leike, Jade Leung, Daniel Levy, Chak Ming Li, Rachel Lim, Molly Lin, Stephanie Lin, Mateusz Litwin, Theresa Lopez, Ryan Lowe, Patricia Lue, Anna Makanju, Kim Malfacini, Sam Manning, Todor Markov, Yaniv Markovski, Bianca Martin, Katie Mayer, Andrew Mayne, Bob McGrew, Scott Mayer McKinney, Christine McLeavey, Paul McMillan, Jake McNeil, David Medina, Aalok Mehta, Jacob Menick, Luke Metz, Andrey Mishchenko, Pamela Mishkin, Vinnie Monaco, Evan Morikawa, Daniel Mossing, Tong Mu, Mira Murati, Oleg Murk, David Mély, Ashvin Nair, Reiichiro Nakano, Rajeev Nayak, Arvind Neelakantan, Richard Ngo, Hyeonwoo Noh, Long Ouyang, Cullen O'Keefe, Jakub Pachocki, Alex Paino, Joe Palermo, Ashley Pantuliano, Giambattista Parascandolo, Joel Parish, Emy Parparita, Alex Passos, Mikhail Pavlov, Andrew Peng, Adam Perelman, Filipe de Avila Belbute Peres, Michael Petrov, Henrique Ponde de Oliveira Pinto, Michael, Pokorny, Michelle Pokrass, Vitchyr H. Pong, Tolly Powell, Alethea Power, Boris Power, Elizabeth Proehl, Raul Puri, Alec Radford, Jack Rae, Aditya Ramesh, Cameron Raymond, Francis Real, Kendra Rimbach, Carl Ross, Bob Rotsted, Henri Roussez, Nick Ryder, Mario Saltarelli, Ted Sanders, Shibani Santurkar, Girish Sastry, Heather Schmidt, David Schnurr, John Schulman, Daniel Selsam, Kyla Sheppard, Toki Sherbakov, Jessica Shieh, Sarah Shoker, Pranav Shyam, Szymon Sidor, Eric Sigler, Maddie Simens, Jordan Sitkin, Katarina Slama, Ian Sohl, Benjamin Sokolowsky, Yang Song, Natalie Staudacher, Felipe Petroski Such, Natalie Summers, Ilya Sutskever, Jie Tang, Nikolas Tezak, Madeleine B. Thompson, Phil Tillet, Amin Tootoonchian, Elizabeth Tseng, Preston Tuggle, Nick Turley, Jerry Tworek, Juan Felipe Cerón Uribe, Andrea Vallone, Arun Vijayvergiya, Chelsea Voss, Carroll Wainwright, Justin Jay Wang, Alvin Wang, Ben Wang, Jonathan Ward, Jason Wei, CJ Weinmann, Akila Welihinda,

Peter Welinder, Jiayi Weng, Lilian Weng, Matt Wiethoff, Dave Willner, Clemens Winter, Samuel Wolrich, Hannah Wong, Lauren Workman, Sherwin Wu, Jeff Wu, Michael Wu, Kai Xiao, Tao Xu, Sarah Yoo, Kevin Yu, Qiming Yuan, Wojciech Zaremba, Rowan Zellers, Chong Zhang, Marvin Zhang, Shengjia Zhao, Tianhao Zheng, Juntang Zhuang, William Zhuk, and Barret Zoph. Gpt-4 technical report, 2024. URL https://arxiv.org/abs/2303.08774.

[101] Sinead O'Connor and Helen Liu. Gender bias perpetuation and mitigation in AI technologies: challenges and opportunities. *AI & SOCIETY*, 39(4):2045–2057, August 2024. ISSN 1435-5655. doi: 10.1007/s00146-023-01675-4. URL https://doi.org/10.1007/s00146-023-01675-4.

[102] Pat Pataranutaporn, Ruby Liu, Ed Finn, and Pattie Maes. Influencing human–ai interaction by priming beliefs about ai can increase perceived trustworthiness, empathy and effectiveness. *Nature Machine Intelligence*, 5(10):1076–1086, 2023.

[103] Pranav Rajpurkar, Emma Chen, Oishi Banerjee, and Eric J Topol. Ai in health and medicine. *Nature medicine*, 28(1): 31–38, 2022.

[104] Aditya Ramesh, Prafulla Dhariwal, Alex Nichol, Casey Chu, and Mark Chen. Hierarchical text-conditional image generation with clip latents, 2022. URL https://arxiv.org/abs/2204.06125.

[105] Werner Rammert. *Where the Action is. Distributed Agency between Humans, Machines, and Programs.* 2008.

[106] Charvi Rastogi, Yunfeng Zhang, Dennis Wei, Kush R. Varshney, Amit Dhurandhar, and Richard Tomsett. Deciding fast and slow: The role of cognitive biases in ai-assisted decision-making. *Proc. ACM Hum.-Comput. Interact.*, 6, April 2022. URL https://doi.org/10.1145/3512930.

[107] Liz Reid. Generative ai in search: Let google do the searching for you. 2024. URL https://blog.google/products/search/generative-ai-google-search-may-2024/.

[108] Robin Rombach, Andreas Blattmann, Dominik Lorenz, Patrick Esser, and Björn Ommer. High-resolution image synthesis with latent diffusion models, 2022. URL https://arxiv.org/abs/2112.10752.

[109] Laura Sartori and Giulia Bocca. Minding the gap(s): public perceptions of AI and socio-technical imaginaries. *AI & SOCIETY*, 38(2):443–458, April 2023. ISSN 1435-5655. doi: 10.1007/s00146-022-01422-1. URL https://doi.org/10.1007/s00146-022-01422-1.

[110] Dietram A. Scheufele and Bruce V. Lewenstein. The Public and Nanotechnology: How Citizens Make Sense of Emerging Technologies. *Journal of Nanoparticle Research*, 7(6):659–667, December 2005. ISSN 1572-896X. doi: 10.1007/s11051-005-7526-2. URL https://doi.org/10.1007/s11051-005-7526-2.

[111] Stefan Schmager, Charlotte Husom Grøder, Elena Parmiggiani, Ilias Pappas, and Polyxeni Vassilakopoulou. What Do Citizens Think of AI Adoption in Public Services? Exploratory Research on Citizen Attitudes through a Social Contract Lens. 2023. doi: 10.24251/HICSS.2023.544. URL http://hdl.handle.net/10125/103176.

[112] Venkatesh Sivaraman, Leigh A Bukowski, Joel Levin, Jeremy M. Kahn, and Adam Perer. Ignore, trust, or negotiate: Understanding clinician acceptance of ai-based treatment recommendations in health care. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, CHI '23. Association for Computing Machinery, 2023. doi: 10.1145/3544548.3581075. URL https://doi.org/10.1145/3544548.3581075.

[113] Edina Strikovic, Sina Blassnig, Eliza Mitova, Aleksandra Urman, Frank Esser, and Claes de Vreese. Opportunity structures for user acceptance of news recommender systems (nrs): A multi-country survey study of relationships between individual-level factors and evaluations of nrs. *New Media & Society*, 0(0):14614448241263765, 2024. doi: 10.1177/14614448241263765. URL https://doi.org/10.1177/14614448241263765.

[114] S Shyam Sundar. Rise of Machine Agency: A Framework for Studying the Psychology of Human–AI Interaction (HAII). *Journal of Computer-Mediated Communication*, 25(1):74–88, 01 2020. ISSN 1083-6101. doi: 10.1093/jcmc/zmz026. URL https://doi.org/10.1093/jcmc/zmz026.

[115] Viriya Taecharungroj. "what can chatgpt do?" analyzing early reactions to the innovative ai chatbot on twitter. *Big Data and Cognitive Computing*, 7(1), 2023. ISSN 2504-2289. doi: 10.3390/bdcc7010035. URL https://www.mdpi.com/2504-2289/7/1/35.

[116] Leila Takayama. *Telepresence and Apparent Agency in Human–Robot Interaction*, chapter 7, pages 160–175. John Wiley & Sons, Ltd, 2015. ISBN 9781118426456. doi: https://doi.org/10.1002/9781118426456.ch7. URL https://onlinelibrary.wiley.com/doi/abs/10.1002/9781118426456.ch7.

[117] Christiane Attig Thomas Franke and Daniel Wessel. A personal resource for technology interaction: Development and validation of the affinity for technology interaction (ati) scale. *International Journal of Human–Computer Interaction*, 35(6):456–467, 2019. doi: 10.1080/10447318.2018.1456150. URL https://doi.org/10.1080/10447318.2018.1456150.

[118] Thomson Reuters. Future of professionals report 2024, 2024. URL https://www.thomsonreuters.com/en/c/future-of-professionals.html.

[119] Suzanne Tolmeijer, Markus Christen, Serhiy Kandul, Markus Kneer, and Abraham Bernstein. Capable but amoral? comparing ai and human expert collaboration in ethical decision making. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*, CHI '22, New York, NY, USA, 2022. Association for Computing Machinery. ISBN 9781450391573. doi: 10.1145/3491102.3517732. URL https://doi.org/10.1145/3491102.3517732.

[120] Guido Vittorio Travaini, Federico Pacchioni, Silvia Bellumore, Marta Bosia, and Francesco De Micco. Machine learning and criminal justice: A systematic review of advanced methodology for recidivism risk prediction. *International Journal of Environmental Research and Public Health*, 19(17), 2022. ISSN 1660-4601. doi: 10.3390/ijerph191710594. URL https://www.mdpi.com/1660-4601/19/17/10594.

[121] European Union. Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence and amending Regulations (EC) No 300/2008, (EU) No 167/2013, (EU) No 168/2013, (EU) 2018/858, (EU) 2018/1139 and (EU) 2019/2144 and Directives 2014/90/EU, (EU) 2016/797 and (EU) 2020/1828 (Artificial Intelligence Act) (Text with EEA relevance), June 2024. URL http://data.europa.eu/eli/reg/2024/1689/oj/eng. Legislative Body: CONSIL, EP.

[122] Viswanath Venkatesh and Michael G. Morris. Why Don't Men Ever Stop to Ask for Directions? Gender, Social Influence, and Their Role in Technology Acceptance and Usage Behavior. *MIS Quarterly*, 24(1):115–139, 2000. ISSN 0276-7783. doi: 10.2307/3250981. URL https://www.jstor.org/stable/3250981. Publisher: Management Information Systems Research Center, University of Minnesota.

[123] Oleksandra Vereschak, Gilles Bailly, and Baptiste Caramiaux. How to Evaluate Trust in AI-Assisted Decision Making? A Survey of Empirical Methodologies. *Proc. ACM Hum.-Comput. Interact.*, 5(CSCW2):327:1–327:39, October 2021. doi: 10.1145/3476068. URL https://doi.org/10.1145/3476068.

[124] Dakuo Wang, Pattie Maes, Xiangshi Ren, Ben Shneiderman, Yuanchun Shi, and Qianying Wang. Designing ai to work with or for people? In *Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems*, CHI EA '21, New York, NY, USA, 2021. Association for Computing Machinery. ISBN 9781450380959. doi: 10.1145/3411763.3450394. URL https://doi.org/10.1145/3411763.3450394.

[125] Dakuo Wang, Liuping Wang, Zhan Zhang, Ding Wang, Haiyi Zhu, Yvonne Gao, Xiangmin Fan, and Feng Tian. "brilliant ai doctor" in rural clinics: Challenges in ai-powered clinical decision support system deployment. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, CHI '21, New York, NY, USA, 2021. Association for Computing Machinery. ISBN 9781450380966. doi: 10.1145/3411764.3445432. URL https://doi.org/10.1145/3411764.3445432.

[126] Pei Wang. On defining artificial intelligence. *Journal of Artificial General Intelligence*, 10(2):1–37, 2019. doi: doi:10.2478/jagi-2019-0002. URL https://doi.org/10.2478/jagi-2019-0002.

[127] Wikipedia. AI boom — Wikipedia, the free encyclopedia, 2024. URL https://en.wikipedia.org/wiki/AI_boom.

[128] Robin Williams, Stuart Anderson, Kathrin Cresswell, Mari Serine Kannelønning, Hajar Mozaffar, and Xiao Yang. Domesticating ai in medical diagnosis. *Technology in Society*, 76:102469, 2024. ISSN 0160-791X. doi: https://doi.org/10.1016/j.techsoc.2024.102469. URL https://www.sciencedirect.com/science/article/pii/S0160791X24000174.

[129] Christo Wilson, Avijit Ghosh, Shan Jiang, Alan Mislove, Lewis Baker, Janelle Szary, Kelly Trindel, and Frida Polli. Building and auditing fair algorithms: A case study in candidate screening. In *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*, FAccT '21, page 666–677, New York, NY, USA, 2021. Association for Computing Machinery. ISBN 9781450383097. doi: 10.1145/3442188.3445928. URL https://doi.org/10.1145/3442188.3445928.

[130] Tianyu Wu, Shizhu He, Jingping Liu, Siqi Sun, Kang Liu, Qing-Long Han, and Yang Tang. A brief overview of chatgpt: The history, status quo and potential future development. *IEEE/CAA Journal of Automatica Sinica*, 10(5):1122–1136, 2023. doi: 10.1109/JAS.2023.123618.

[131] Xingjiao Wu, Luwei Xiao, Yixuan Sun, Junhang Zhang, Tianlong Ma, and Liang He. A survey of human-in-the-loop for machine learning. *Future Generation Computer Systems*, 135:364–381, 2022. ISSN 0167-739X. doi: https://doi.org/10.1016/j.future.2022.05.014. URL https://www.sciencedirect.com/science/article/pii/S0167739X22001790.

[132] Daniel Yekutieli and Yoav Benjamini. Resampling-based false discovery rate controlling multiple test procedures for correlated test statistics. *Journal of Statistical Planning and Inference*, 82(1):171–196, 1999. ISSN 0378-3758. doi: https://doi.org/10.1016/S0378-3758(99)00041-5. URL https://www.sciencedirect.com/science/article/pii/S0378375899000415.

[133] Halit Yilmaz, Samat Maxutov, Azatzhan Baitekov, and Nuri Balta. Student attitudes towards chat gpt: A technology acceptance model survey. *International Educational Review*, 1(1):57–83, 2023.

[134] Fabio Massimo Zanzotto. Viewpoint: Human-in-the-loop artificial intelligence. *Journal of Artificial Intelligence Research*, 64:243–252, 2019.

[135] Baobao Zhang and Allan Dafoe. Artificial intelligence: American attitudes and trends. *Available at SSRN 3312874*, 2019.

[136] Qiaoning Zhang, Matthew L Lee, and Scott Carter. You complete me: Human-ai teams and complementary expertise. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*, CHI '22, New York, NY, USA, 2022. Association for Computing Machinery. ISBN 9781450391573. doi: 10.1145/3491102.3517791. URL https://doi.org/10.1145/3491102.3517791.

A Survey samples

Figure 6 provides detailed information about the samples from both survey waves.

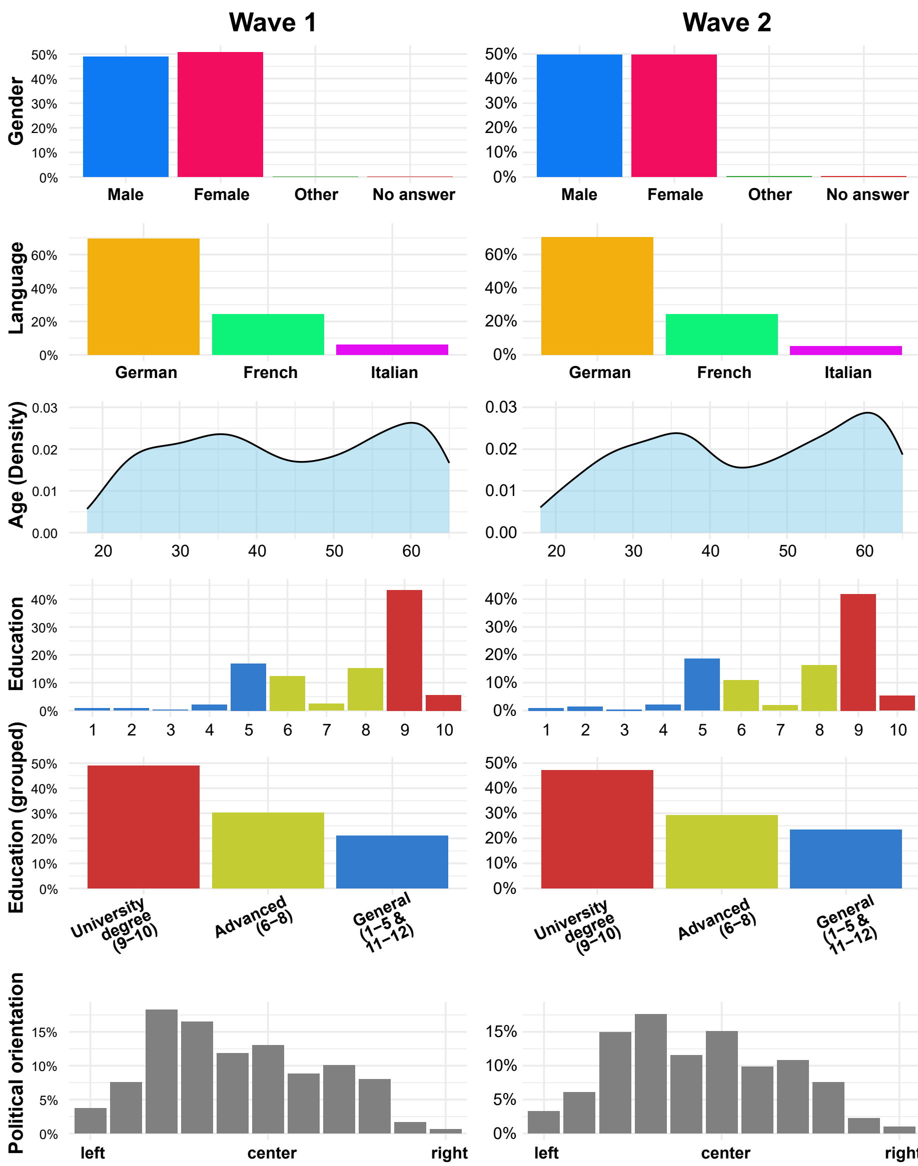

Fig. 6. Demographic overview of survey respondents in both waves.

B Further empirical results

B.1 Perceived need for human control without aggregated answer options

In Section 4.2, we combined the two answer options (*AI decision with human oversight* and *AI-assisted human decision*) into one single category titled *human-AI collaboration*, as they are conceptually very similar. The resulting fraction of selected answer responses is shown in Figure 5. In Figure 7, we show the same overview but without aggregating any of the answer options. The trend of more human oversight after the generative AI boom is clearly visible. Much fewer people want AI to make the decision, with humans merely taking the role of intervening. The only exception is the hiring scenario, where the fraction of people answering "AI decision with human oversight" slightly increased from 17.67 to 20.12. The share of people favoring human decisions with AI assistance is more stable and even decreases with the generative AI boom for scenarios with higher potential impacts (medical diagnosis, hiring, therapy discontinuation, and prisoner release).

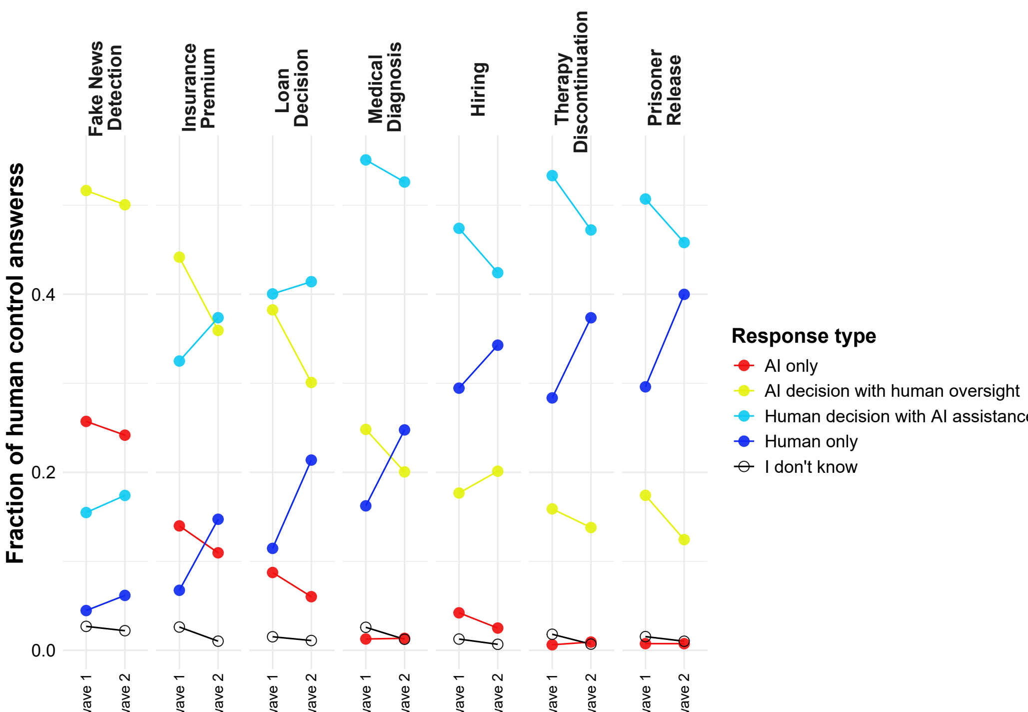

Fig. 7. Response distribution in % for human control questions across survey waves and scenarios.

B.2 Amplification of existing inequalities with the GenAI boom

B.2.1 Amplified education gap. Figure 8 demonstrates that higher levels of education correlate with greater AI acceptance and a greater acceptance towards human-AI collaboration instead of human-only decision-making without the help of AI.

People with university degrees consistently show higher AI acceptance than those with advanced non-university education, who, in turn, exhibit more acceptance than those with only general education. This trend is evident across all surveyed scenarios. Importantly, the gap in AI acceptance between these educational groups has widened post-GenAI boom. In particular, people with a general education level have become more critical. This is visible in the top panel of Figure 8 with the steeper slopes of the green lines, indicating significantly reduced acceptance levels for several scenarios such as fake news detection, insurance premiums, loan decisions, medical diagnosis, and therapy discontinuation. For example, in the context of medical diagnosis, the acceptance for individuals with general education dropped from 2.68 to 2.40, while university-educated maintained relatively stable acceptance levels (2.97 to 2.92).

The negative regression estimates for advanced and general education in Table 8 show that, within both waves, a lower educational level can be predictive of lower AI acceptance. In wave 2, general education is a significant predictor in all scenarios except therapy discontinuation).

The preference for human-AI collaboration declined across all education levels but most sharply among those with general education. The bottom panel of Figure 8 illustrates that individuals with lower education levels increasingly favor human-only decision-making. These differences were amplified with the generative AI boom, especially for loan decisions and medical diagnosis, as well as for insurance premiums. Before the AI boom, 17.6% of respondents with general education preferred human-only involvement in medical diagnosis, rising to 30.9% post-boom. The regression coefficients in Table 9 additionally show that lower educational level is a predictive factor for more human control after the generative AI boom, with significant positive estimates for general education across most scenarios in wave 2.

These results highlight that educational disparities have been amplified in the wake of the generative AI boom, with lower-educated individuals showing increased skepticism and a stronger preference for human oversight.

B.2.2 Amplified language region gap. Figure 9 illustrates a pronounced disparity in AI acceptance and the necessity for human control between the German-speaking and French-speaking populations in Switzerland. French speakers consistently exhibit lower AI acceptance compared to German speakers, with the gap being most pronounced in scenarios like insurance premiums, loan decisions, and the two medical scenarios.

While the disparity in AI acceptance has remained constant for most scenarios, it has widened post-GenAI boom for the two medical scenarios. For example, in the context of medical diagnosis, average AI acceptance among French speakers dropped significantly (from 2.75 to 2.33, $p < 0.05$), while for German speakers, it decreased non-significantly (from 2.76 to 2.63, $p > 0.05$). Similar trends are observed in therapy discontinuation, where the acceptance for French speakers fell significantly from 2.11 to 1.68 ($p < 0.05$), compared to a non-significant decline from 2.17 to 2.03 among German speakers (see top panel of Figure 9. These scenarios show the same trend regarding the reported necessity of human control, as visible in the bottom panel of Figure 9.

Additionally, the gap between the two language regions in terms of human control of AI-based applications also widened for the two scenarios on insurance premiums and prisoner release: French speakers reported much less willingness of human-AI collaborations and instead started favoring more human-only settings with the generative AI boom (as visible with the dashed lines showing substantial differences between waves 1 and 2). In contrast, German speakers only slightly increased their preference for human-only decisions post-boom. For instance, the preference for human-only decisions in therapy discontinuation among French speakers increased from 30.17% in wave 1 to 56.45% in wave 2 (compared to a small increase from 28.37% to 31.55% for German speakers). One exception is the hiring scenario, where the preference for human control among French speakers remained stable, unlike other scenarios where significant shifts were observed.

Tables 8 and 9 further highlights these trends. The regression estimates indicate significant negative coefficients for AI acceptance (positive coefficients for human control, respectively) for French speakers in wave 2 across almost all scenarios.

The trend for reported human control preferences among Italian speakers evolved similarly across both survey waves, mirroring the pattern observed among French speakers. Therefore, for clarity, we omit it from bottom panel of Figure 9.

B.2.3 Amplified gender gap in health scenarios. Figure 10 highlights the disparities in AI acceptance and the need for human control between genders, particularly in health-related scenarios. Women exhibit significantly more skepticism regarding the use of AI for medical diagnosis and therapy discontinuation, and this gender gap has widened post-GenAI boom. For instance, the acceptance of AI in medical diagnosis for women decreased from 2.57 in wave 1 to 2.32 in wave 2, whereas for men, it showed a smaller decline from 2.94 to 2.80 (top panel of Figure 10). Similarly, in the context of therapy discontinuation, the acceptance among women dropped from 2.13 to 1.83, while for men, it decreased slightly from 2.20 to 2.07. In the therapy discontinuation scenario, men and women had similarly low acceptance on average in wave 1 (2.20 for men and 2.13 for women). With the generative AI boom, this decreased slightly for men (to 2.07), but women became significantly less accepting of using AI for therapy discontinuation decisions (to 1.83). This trend is further supported by the regression estimates in Table 8, which show significant negative coefficients for women in scenarios like medical diagnosis and therapy discontinuation in wave 2. Furthermore, women also showed a lower acceptance for using AI in prisoner release scenarios in both waves – see significant negative regression estimate for Gender (f) in Table 8.

The bottom panel of Figure 10 tells a similar story, showing larger increases in human-only answers among women compared to men, for the aforementioned scenarios of medical diagnosis, therapy discontinuation, and prisoner release. Table 9 supports these observations, showing significant positive coefficients for women in wave 2 for the scenarios of medical diagnosis and therapy discontinuation.

Table 8. Predictors of AI acceptance in different scenarios.

Dependent variable: **AI acceptance**

Scenario →	**Medical Diagnosis**	**Therapy Discontinuation**	**Fake News Detection**	**Hiring**	**Loan Decision**	**Insurance Premium**	**Prisoner Release**
By education level							
Wave 1							
Advanced education	-0.17 (0.11)	0.03 (0.1)	-0.11 (0.09)	-0.16 (0.11)	0.01 (0.1)	-0.05 (0.11)	0.03 (0.09)
General education	**-0.28 (0.11)***	0.04 (0.1)	-0.13 (0.09)	**-0.26 (0.09)***	**-0.25 (0.1)***	-0.21 (0.1)	-0.09 (0.09)
Wave 2							
Advanced education	-0.18 (0.08)	0.04 (0.07)	-0.06 (0.09)	-0.18 (0.08)	**-0.24 (0.09)***	-0.12 (0.09)	-0.02 (0.07)
General education	**-0.49 (0.09)*****	-0.08 (0.08)	-0.2 (0.1)	**-0.24 (0.09)***	**-0.53 (0.1)*****	**-0.46 (0.1)*****	-0.14 (0.08)
By language region							
Wave 1							
French language	-0.01 (0.16)	-0.06 (0.15)	-0.12 (0.12)	-0.03 (0.14)	**-0.3 (0.13)***	**-0.42 (0.14)****	-0.06 (0.15)
Italian language	-0.08 (0.21)	0.05 (0.19)	-0.22 (0.23)	0.1 (0.23)	-0.07 (0.23)	0.1 (0.24)	0.25 (0.19)
Wave 2							
French language	**-0.33 (0.11)****	**-0.38 (0.09)*****	**-0.29 (0.12)***	-0.21 (0.11)	**-0.31 (0.12)***	**-0.57 (0.12)*****	**-0.29 (0.09)****
Italian language	-0.19 (0.26)	-0.08 (0.23)	-0.35 (0.18)	-0.35 (0.17)	-0.29 (0.26)	-0.1 (0.26)	-0.12 (0.19)
By gender							
Wave 1							
Gender (f)	**-0.37 (0.12)****	-0.07 (0.11)	0.05 (0.11)	-0.12 (0.11)	-0.07 (0.11)	0.04 (0.12)	-0.19 (0.1)
Wave 2							
Gender (f)	**-0.51 (0.09)*****	**-0.23 (0.09)***	0.08 (0.11)	-0.05 (0.1)	-0.06 (0.11)	-0.03 (0.11)	**-0.26 (0.09)****

Note: This table summarizes the results of 42 separate survey-weighted linear regression models, which cover 7 scenarios, 3 demographic attributes, and 2 survey waves. The dependent variable indicates the acceptance of using AI (1=not acceptable at all, ..., 5=definitely acceptable). Each cell presents regression estimates with standard errors in parentheses. Significant predictors are indicated in bold, with significance levels as follows: *** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$.

Table 9. Predictors of the need for human control in different scenarios.

Dependent variable: **human control**

Scenario →	**Medical Diagnosis**	**Therapy Discontinuation**	**Fake News Detection**	**Hiring**	**Loan Decision**	**Insurance Premium**	**Prisoner Release**
By education level							
Wave 1							
Advanced education	0.05 (0.06)	0.06 (0.05)	0.14 (0.07)	**0.19 (0.07)***	0.08 (0.06)	0.1 (0.07)	0.06 (0.06)
General education	0.11 (0.05)	0.1 (0.06)	**0.23 (0.06)*****	**0.21 (0.06)****	**0.22 (0.07)****	**0.21 (0.06)****	0.04 (0.06)
Wave 2							
Advanced education	0.09 (0.05)	0.01 (0.05)	0.07 (0.06)	0.11 (0.06)	**0.17 (0.05)****	0.11 (0.06)	0.06 (0.05)
General education	**0.23 (0.06)*****	0.07 (0.05)	0.12 (0.06)	0.09 (0.06)	**0.36 (0.06)*****	**0.32 (0.06)*****	0.1 (0.05)
By language region							
Wave 1							
French language	0 (0.08)	-0.01 (0.08)	0.04 (0.1)	0.14 (0.09)	0.21 (0.11)	0.18 (0.09)	0.06 (0.09)
Italian language	-0.1 (0.12)	-0.14 (0.1)	0.01 (0.15)	-0.2 (0.19)	-0.02 (0.16)	0.05 (0.14)	-0.17 (0.14)
Wave 2							
French language	**0.3 (0.07)*****	**0.33 (0.06)*****	0.11 (0.08)	0.13 (0.07)	**0.2 (0.08)***	**0.37 (0.09)*****	**0.23 (0.06)*****
Italian language	0.25 (0.15)	0.21 (0.16)	**0.37 (0.12)****	**0.24 (0.09)***	**0.34 (0.13)***	**0.5 (0.15)****	0.1 (0.17)
By gender							
Wave 1							
Gender (f)	0.12 (0.06)	0.1 (0.06)	0.08 (0.07)	0.07 (0.07)	0.09 (0.08)	0 (0.07)	-0.01 (0.06)
Wave 2							
Gender (f)	**0.18 (0.06)****	0.13 (0.06)	-0.12 (0.07)	0.02 (0.06)	-0.09 (0.07)	0 (0.07)	0.02 (0.06)

Note: This table summarizes the results of 42 separate survey-weighted linear regression models, encompassing 7 scenarios, 3 demographic attributes, and 2 survey waves. The dependent variable indicates the preference for decision-making between humans, AI, or both together (1=human only, ..., 4=AI only). Each cell presents regression estimates with standard errors in parentheses. Significance levels are denoted as follows: *** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$.

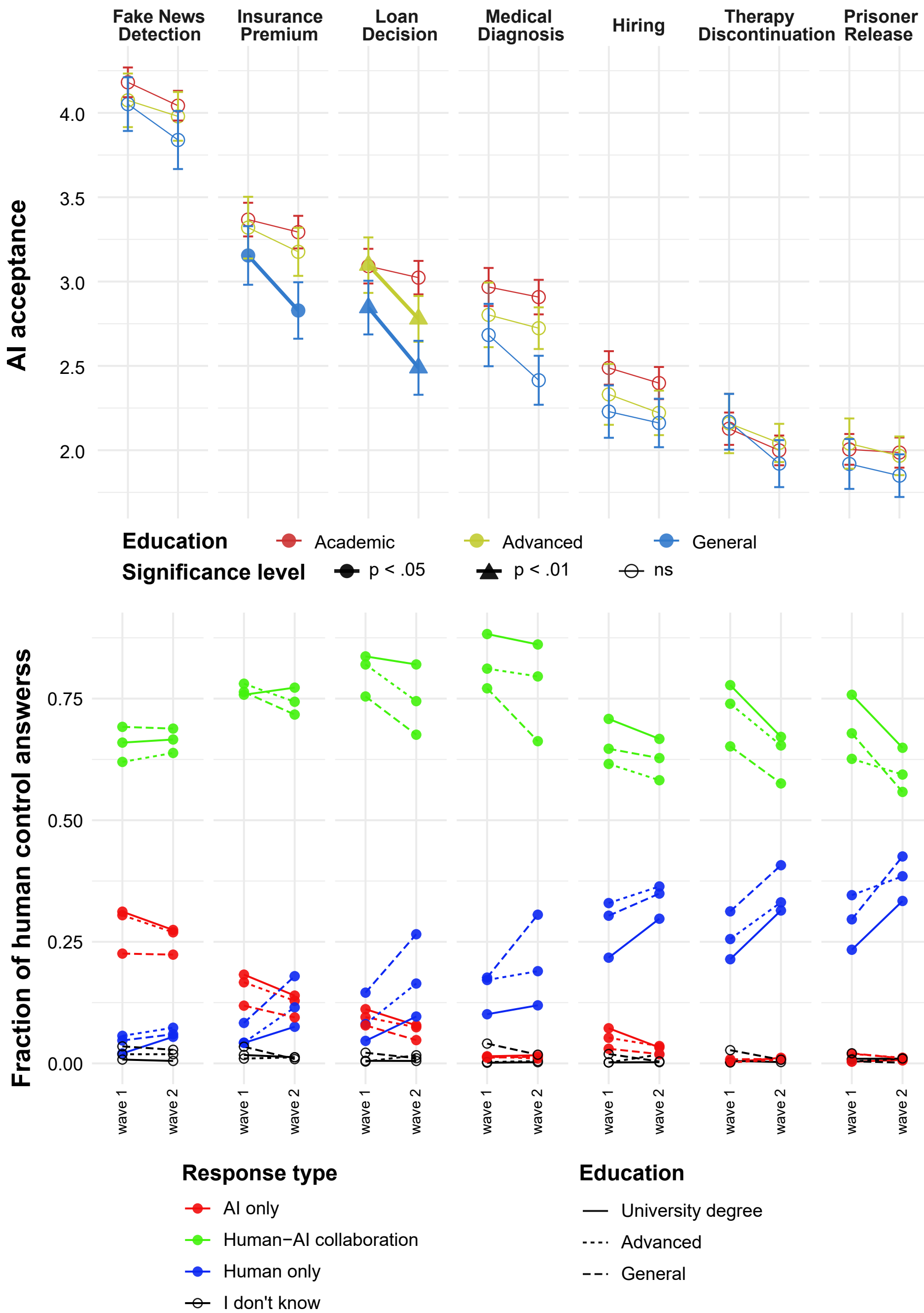

Fig. 8. AI acceptance (top) and human control (bottom) across different education levels, grouped by scenarios and survey waves. Individuals with lower education levels exhibit a larger skepticism toward AI involvement. This gap has increased with the generative AI boom: AI acceptance increased, and the perceived necessity of human involvement decreased for people with general education while remaining more stable for those with a university degree.

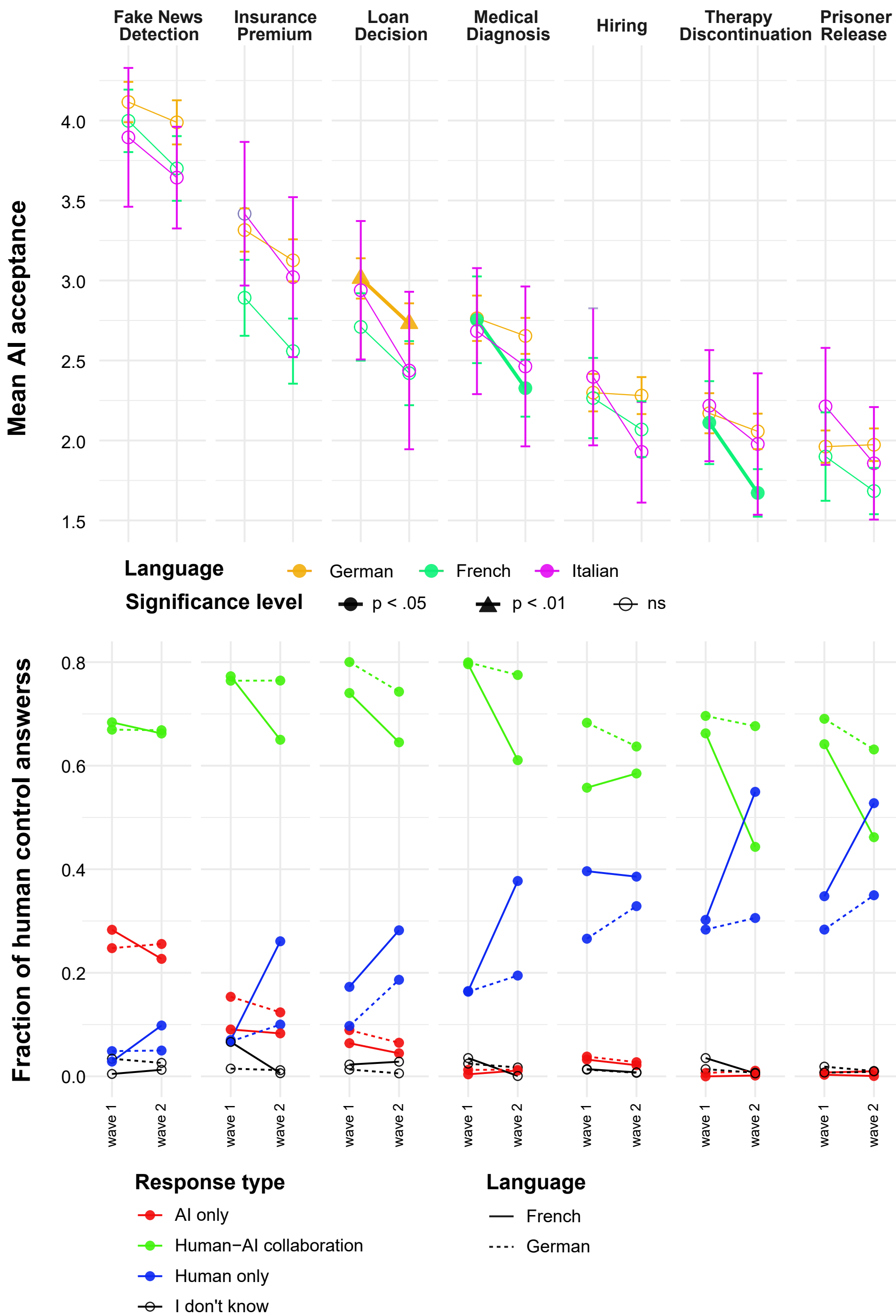

Fig. 9. AI acceptance (top) and human control (bottom) across different language regions, grouped by scenarios and survey waves. The gap between German- and French-speaking regions has increased: more skepticism toward AI among French speakers after the generative AI boom (i.e., lower AI acceptance and a shift from human-AI collaboration to human-only decisions).

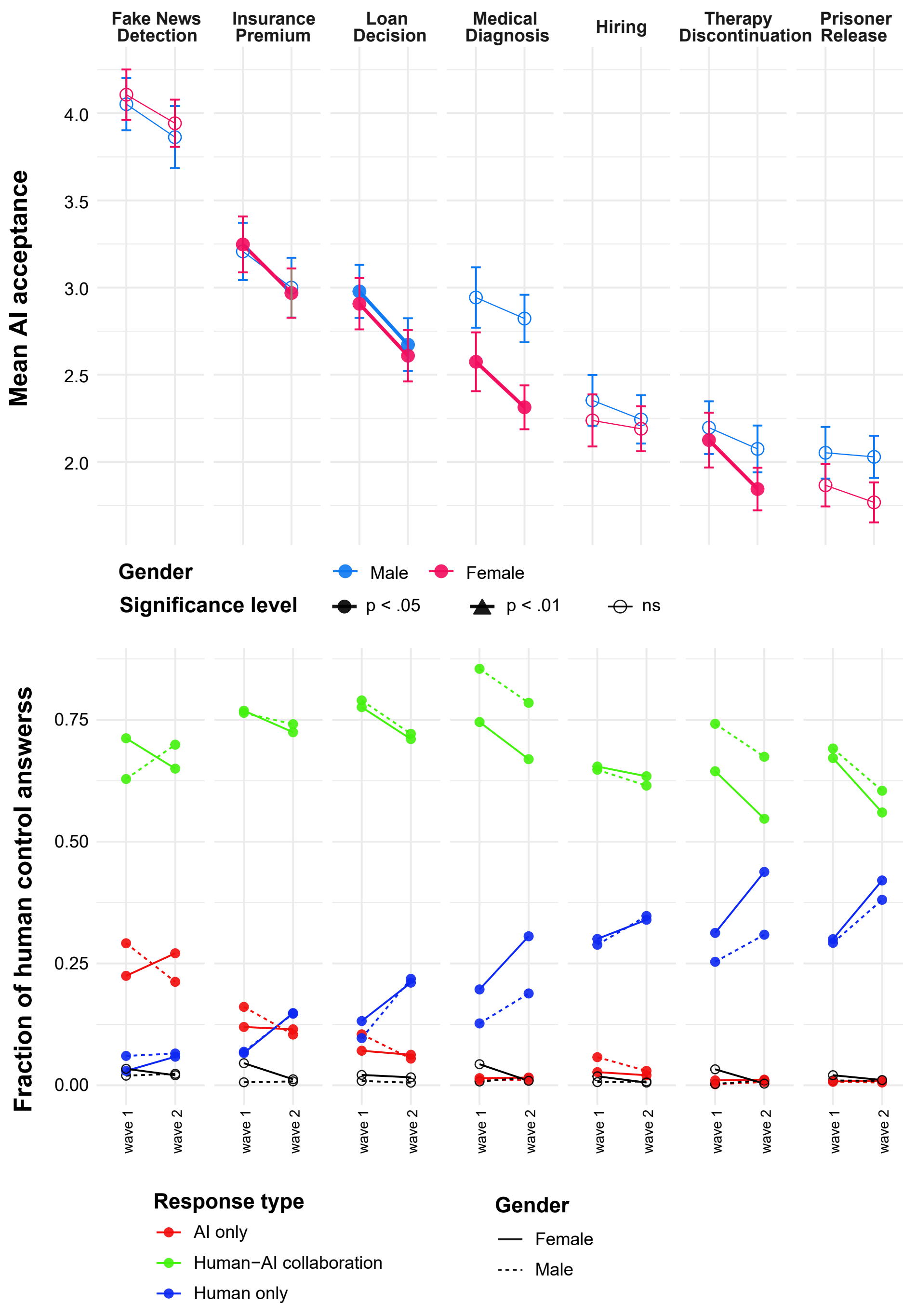

Fig. 10. AI acceptance (top) and human control (bottom) across genders, grouped by scenarios and survey waves. Women exhibit significantly lower AI acceptance compared to men, with the gap widening post-GenAI boom for health-related scenarios. Compared to survey wave 1, women reported human-only decisions as the preferred option more often after the generative AI boom, on average.