"Because AI is 100% right and safe": User Attitudes and Sources of AI Authority in India

Shivani Kapania
kapania@google.com
Google Research
Bengaluru, India

Oliver Siy
siyj@google.com
Google Research
Cambridge, Massachusetts, United States

Gabe Clapper
gabeclapper@google.com
Google Research
Seattle, Washington, United States

Azhagu Meena SP
azhagu@google.com
Google Research
Bengaluru, India

Nithya Sambasivan
nithyasamba@google.com
Google Research
Bengaluru, India

ABSTRACT

Most prior work on human-AI interaction is set in communities that indicate skepticism towards AI, but we know less about contexts where AI is viewed as aspirational. We investigated the perceptions around AI systems by drawing upon 32 interviews and 459 survey respondents in India. Not only do Indian users accept AI decisions (79.2% respondents indicate acceptance), we find a case of *AI authority*—AI has a legitimized power to influence human actions, without requiring adequate evidence about the capabilities of the system. AI authority manifested into four user attitudes of vulnerability: faith, forgiveness, self-blame, and gratitude, pointing to higher tolerance for system misfires, and introducing potential for irreversible individual and societal harm. We urgently call for calibrating AI authority, reconsidering success metrics and responsible AI approaches and present methodological suggestions for research and deployments in India.

CCS CONCEPTS

• **Human-centered computing** → **Empirical studies in HCI**; • **Social and professional topics** → Cultural characteristics.

KEYWORDS

artificial intelligence, perceptions of AI, algorithmic decision-making, India

1 INTRODUCTION

A growing body of research on public attitudes toward algorithmic systems indicates skepticism and moderate acceptance towards these technologies in contexts such as the US, UK and Germany [66], where studies report that individuals express concerns about the fairness and usefulness of these systems [60, 117]. The research on improving human-AI interactions is thus often set in communities where studies indicate user mistrust towards algorithmic systems[1] [19, 74, 80], with a goal to mitigate harm to users, while many are also explicitly motivated to increase the user acceptance of technologies through various approaches like explainability [26], privacy [22] and transparency [112]. However, the acceptance among users may be shaped by specific online trajectories and exposure levels, possibly not generalizing to contexts with newer Internet citizens from under-researched socio-cultural settings.

In particular, AI deployments in India are emerging in several niche, high-stakes areas (*e.g.,* healthcare [100, 128], finance [116], agriculture [27]). Marda *et al.* [82] describe how AI is also emerging as a focus for policy development in India. Prior research, however, presents a case for techno-optimism among technology users that envision technology with a socio-economic promise [95] for India. AI is viewed aspirationally, as a means to development and modernity [14], potentially leading to adoption of high-stakes solutions, often before sufficient ethics testing takes place [107]. As the world's second largest Internet user population [59], it is important to understand how Indian users perceive AI systems.

In this research, we draw upon, and extend the concept of algorithmic authority proposed by Lustig and Nardi [81] to AI applications. We define AI authority as *the legitimized power of AI to influence human actions, without requiring adequate evidence about the capabilities of the given system.* We center the term 'Artificial Intelligence' to investigate public perceptions and contextual factors, which for instance, Kozyreva and Herzog *et al.* [65] have also reported to be significantly more familiar (86% *vs.* 58%) among the German public as opposed to more technical terms such as 'Computer algorithms'.

In this paper, we describe the characteristics of authority of AI, the sources through which AI systems derive authority, and the user

---

[1]the results on perceptions of trust and fairness are domain-dependent and mixed, but predominantly negative in contexts such as the US and EU, which we highlight in the following section

attitudes that AI authority manifests into—from 32 interviews and 459 survey responses across various domains and conditions, set in India. In our study, **79.2%** of survey respondents were willing to *accept* AI decisions. A high acceptance for AI held true even in high-stakes[2] scenarios, with 73% respondents indicating acceptance of AI decisions for medical diagnosis, loan approval and hiring. AI authority is legitimized in our study, with participants demonstrating both acceptance and willingness to act upon AI-based decisions.

The sources through which an AI derived authority were often extraneous to the system, and not indicative of the reliability, usefulness, or effectiveness of the given AI system. AI authority was heavily influenced by institutional, infrastructural and societal factors that lay outside the boundaries of the AI—users' interactions with ineffectual institutions, polarized narratives and misleading terminologies, users' prior experiences with technology, and the availability of human systems around them. For example, many participants were left exasperated by the corrupt or discriminatory practices in their interactions with financial institutions, which is why they perceived AI as a better alternative to avoid those forms of exploitation. AI authority manifests into four user attitudes towards AI: faith in AI capabilities, blaming themselves for adverse and biased outcomes, forgiveness, and gratitude towards AI. Taken together, our results indicate how AI authority led to a greater tolerance for AI harms, and lower recognition of AI biases. An authority of AI in our study meant that participants under-recognized and even tolerated AI harms, posing serious questions on adverse impacts of system errors, bias, abuse, and misfires on nascent Internet users.

Our work has implications for the design of responsible AI systems, by highlighting that users from under-studied settings could have different attitudes towards and behaviors with AI due to the contextual or non-technological factors. Our results indicate that there is already a high acceptance towards adopting AI systems among Indian technology users, which means we must approach design and research differently (*e.g.,* by supporting people in maintaining a healthy distrust and critical awareness of AI). If AI authority remains over-calibrated, even high-risk and under-tested AI applications may receive user acceptance relatively quickly, and thus might be easily adopted for public use. We call for (i) efforts to calibrate authority, and greater attention towards the trajectory of users that begin from a place of authority for AI and a cultural mistrust for human institutions, (ii) embracing the variability in AI understanding, (iii) redefining success metrics and (iv) developing and disseminating alternative narratives around AI. Our paper makes three main contributions, we:

— present *evidence for AI authority* in India, and describe its characteristics
— empirically *document the sources influencing AI authority* beyond system development procedures
— *provide implications* for researchers and designers in introducing AI technologies in India or other contexts with authority towards AI

[2]We define high-stakes as a situation with possibly far-reaching consequences for the future of an individual, but acknowledge that stakes involved in a decision are subjective and personal.

2 RELATED WORK

Researchers have studied and documented the ways in which the design of AI-infused systems diverges from traditional user interface designs [29, 34]. Most of these differences are attributed to the probabilistic nature of AI, that heavily relies on nuances of various task design and system settings, and often manifests as inconsistent behavior over time or across users [5, 72]. Below, we situate our work in a body of related research on trust and algorithmic authority, perceptions about algorithmic systems and decisions, algorithmic folk theories, and techno-optimism in India.

2.1 Trust and algorithmic authority

Understanding the role and measurement of trust continues to be an active topic of research for several decades, across various disciplines such as interpersonal relationships [104], organizational studies [89], consumer relations [87], and technology [61]. Several definitions and frameworks have emerged from the widespread and varied interest in trust [58, 83, 104]. Lee and See [73] define trust in automation as *"an attitude that an agent will help achieve an individual's goals in a situation characterized by uncertainty and vulnerability."*, and used the framework by Fishbein and Ajzen [42] to characterize trust as an attitude of the individual, rather than an intention or a behavior. Beyond defining what it means to trust, researchers have devised frameworks for the factors that contribute to trust. Mayer, Davis and Schoorman [83] present an integrative model of trustworthiness with three components: ability, integrity, and benevolence. Ability refers to the competencies that are required for a trustee to have influence in a specific domain. Integrity is about following a set of principles that are acceptable to the trustor. Benevolence refers to the extent to which the intentions of the trustee align with the trustor. We utilize these frameworks in the following sections to reflect on our results and present implications.

Extensive prior research has explored constructs to understand the emergent dynamics between human and algorithmic interactions (*e.g.,* AI utopia [111], algorithmic appreciation [79], algorithm aversion [32]). Santow [111] described AI utopia as an *imagined* future with improvements in every aspect of life. In this research, we draw inspiration from Lustig and Nardi's work [81] that introduced the concept of algorithmic authority grounded in their mixed-methods research of the Bitcoin community. Users perceived Bitcoin as an 'apolitical currency', as more reliable and trustworthy than banks or governments, and the algorithmic was seen as a form of explicit resistance against traditional financial institutions. The concept of trust is of particular importance to algorithmic authority. Bitcoin users place trust in the algorithmic authority: trust is an attitude that mediates relationships between humans and the algorithmic authority. Lustig and Nardi discuss the ways in which Bitcoin was perceived to possess a predictability which led to greater trust in the algorithmic authority. However, the algorithmic authority of Bitcoin was mediated by human judgment, that it required human oversight even if the algorithmic was perceived as 'self-contained'. In this research, we extend the work of Lustig and Nardi to examine the perceptions and acceptance of AI-based applications among Indian users, and the factors which contribute to these perceptions.

2.2 Perceptions of algorithmic systems and decisions

The communities of HCI and CSCW have investigated the perceptions of algorithmic decision-making in comparison with human decisions (*e.g.,* [13, 19]) with conflicting results. Most studies find that individuals are more likely to trust human decisions over algorithmic decisions for tasks that require human-like skills (*e.g.,* subjective judgment, emotion) (*e.g.,* [80]). For instance, Lee *et al.* [74] explore perceptions of fairness and trust for four managerial decisions and find that human decisions are considered fairer and more trustworthy than algorithmic decisions for tasks requiring human skills (hiring and work evaluation). Dietvorst *et al.* [32] found that human forecasters are resistant to using algorithms, especially after seeing them perform (algorithm aversion). Logg, Minson, and Moore [79], however, present contradicting results: they find that people rely on algorithmic advice over human advice for forecasting (algorithmic appreciation), but this effect waned when they had to choose between their own judgment *vs.* an algorithm. Others in the community have studied human perceptions of AI through a wide-ranging set of methods [39, 63]. Cave, Couglan and Dihal [20] conducted a survey to examine public responses to AI in the UK, and find that 25% participants associated AI with robots. Prior research has also investigated worker perceptions of technology in organizations [76, 93]. Höddinghaus *et al.* [48] compared workers' trustworthiness perceptions of automated *vs.* human leadership. Their results indicate workers perceived automated leadership agents as being higher on integrity, but lower on benevolence than human leadership agents. Nagtegaal [88] demonstrate the role of task quantifiability/complexity in perceptions of procedural justice: decision automation was perceived as less fair in tasks with higher complexity (and lower quantifiability [71], such as hiring).

A related area of research examines the factors influencing perceptions of trust, algorithmic fairness and justice (*e.g.,* [75, 136]). People's perception of fairness can be complicated and nuanced, and go beyond formal algorithmic constraints. Araujo *et al.* [7] explored the role of individual characteristics in perceptions about automated decision-making, and report that people with more knowledge about AI are more optimistic about algorithmic decision-making. Ashoori and Weisz [10] report that model interpretability and transparency about training/testing datasets played a critical role in establishing trust. However, recent research by Wang *et al.* [130] argues that for non-expert stakeholders, the effect of development procedures on perceived fairness is much smaller as compared to algorithmic outcomes and biases.

Relatively less work has been done to uncover the perceptions and acceptance of algorithmic systems in India (*e.g.,* [52, 84]). Okolo *et al.* [92] studied the perceptions of community health workers (CHWs) in rural India towards AI, and find that CHWs considered AI applications trustworthy with expertise greater than their own. Thakkar *et al.* [122] examined the perceptions and practices of vocational workers towards automated systems. We extend this body of work by focusing on the ways in which perceptions and intentions are shaped through factors beyond system design. We deliberately focus on 'AI systems' as opposed to 'algorithmic systems', given

the specific associations that people make through terminologies and conceptual metaphors [41].

2.3 Algorithmic folk theories

Algorithmic folk theories are conceptions to explain algorithmic behavior and outcomes that are informally shared among similar members of a culture, but are not necessarily accurate [28]. Prior CSCW and HCI literature has explored people's understanding of social media algorithms [31, 45], and find that users develop these folk models through their experience with technologies and social interactions. Their results indicate that a majority of people have high-level folk theories about how these algorithms work. Rader and Gray [102] studied user beliefs about news feed curation on Facebook, the algorithmic behaviors that users notice and the behaviors to which they respond. Taking it further, Eslami *et al.* [36] uncovered and codified the folk theories of the Facebook News Feed curation. Most people become aware of algorithmic involvement through unexpected behavior, which causes them to adjust their future actions.

Prior research has also made headway in examining user awareness of algorithms, and the role of that awareness in their interactions. Bucher [17] developed the concept of the algorithmic imaginary as the *"ways of thinking about what algorithms are, what they should be and how they function [...] plays a generative role in moulding the Facebook algorithm itself."* Hargittai *et al.* [45] study people's algorithmic skills, and report on the methodological challenges. Algorithmic skills involve the ability to identify that a dynamic system is in place (awareness), and reflecting on how those systems work (understanding) [45]. Another recent thread of work focuses on the ways in which users adapt behaviors around algorithms [53, 55]. Eslami *et al.* investigate the ways in which users perceive and manage bias in online rating platforms [38], and propose the concept of 'bias aware design' to harness the power of users to reveal algorithmic biases.

Most of this research has been conducted with US or EU-based respondents, largely with social media, and/or MTurk users who are known to be tech-savvier than the average Internet user [46]. US and EU tend to have critical media discourse around the use of algorithmic systems [64, 119], scrutiny about algorithmic biases from activists and civil society [3], and existing/emerging laws and regulations around the use of AI [25, 54, 105]. This is in contrast with India, which currently lacks substantial research on responsible AI development [107]. Algorithmic folk theories are shaped through the above elements, which differ greatly between the contexts previously studied and our study site. In addition, our goal is not to provide explicit folk theories, instead, we connect folk understandings with users' intentions to engage with, and act upon AI decisions or recommendations.

2.4 Techno-optimism and modernity in India

Prior work in HCI and ICTD has studied the discourses around technology in India, and how it has been frequently tied with notions of development [95]. The last two decades, in particular, have been crucial in shaping technology as the means for prosperity in India. Digital technologies are viewed as a vehicle for progress, as a solution to societal problems in developing countries [16]. The

following is an excerpt from Prime Minister Modi's speech in 2018 [86]: *"Can Artificial Intelligence help us detect serious health conditions before they manifest physically? Can Artificial Intelligence help our farmers make the right decisions regarding weather, crop and sowing cycle? Friends, our Government is of the firm belief, that we can use this power of twenty-first century technology [AI] to eradicate poverty and disease. In doing so, we can bring prosperity to our poor and under-privileged sections. We are committed to achieving this vision."*

The Indian government's vision for technological development manifested through two major initiatives over the recent years. First, the introduction of Aadhaar— a biometric identification system for 1.3 billion citizens, which was legitimized through a promise for poverty reduction and financial inclusion [120]. Secondly BHIM, an application for digital payments introduced soon after demonetization as the future for cashless payments [97]. Technology played a symbolic and functional role in enabling the 'leapfrogging' into the modern era. Bozarth and Pal [14] examine the social media (Twitter) discourse, and find that politicians have an inclination to discuss technology in connection with development as part of their political messaging. Public discourse around technology and AI has been hyper-optimistic from the general public, the tech industry, and the government [96], with several deployments underway [90]. Sambasivan *et al.* [107] describe how due to the aspirational role played by AI, high risk AI solutions are often adopted too quickly in India, including in predictive policing and facial recognition, without sufficient checks and balances. Our research aligns with this perspective on effects of techno-optimism and extends this body of literature by presenting findings of the ways in which this aspirational value is shaping AI authority among Indian technology users.

3 METHODOLOGY

We use a mixed-methods research approach with an exploratory design [50] in which the qualitative data serves as the foundation of the inquiry. In this design, the qualitative data is collected and analyzed first, followed by the collection and analysis of the quantitative data which typically uses constructs based on the emergent phenomenon from the qualitative study. First we conducted semi-structured interviews to investigate the perceptions about AI systems, users' intentions, and the ways in which those relate to their conceptions of AI. We then implemented a survey as a follow up study to measure the acceptance of AI-based outcomes with our target population. The quantitative research complements the qualitative by presenting results on acceptability with a larger sample size, and establish baseline data for Indian technology users. We describe implementation details for each method in the following subsections.

3.1 Interviews

We conducted interviews with 32 adult Internet users based in different regions within India to understand perceptions of AI and their acceptance of AI-based decisions. Participants were located in various tier 1 and tier 2 cities of India, belonging to diverse age groups and occupations. Our sample also consisted of a mix of internet experience, with 16 nascent internet users that first accessed the internet only in the last 2 years, and the remaining 16 more 'experienced' that have been online for more than 2 years. We interviewed 17 male and 15 female participants. Refer to table 1 for details on participant demographics.

Participant recruitment and moderation. We recruited participants through a combination of a UXR database internal to our organization, and a market research company Dowell. We conducted 'screener' conversations with potential participants to ask if *"[they] had heard of the words AI or Artificial Intelligence?"*, and if yes, we asked them to describe what it means. We selected people who were aware of these terminologies and did not filter any participant based on their descriptions of AI. The sessions were conducted using video conferencing, due to COVID-19 travel limitations, and each interview lasted 75-90 minutes. We conducted interviews in three languages (Hindi, Tamil and English) by allowing participants to select the language with which they are most comfortable. The interview notes were recorded through field notes and video recording, and transcribed within 24 hours of each interview by the corresponding moderator. Each participant received a thank-you gift in the form of a gift card worth 27 USD or 2000 INR.

Interview structure. In line with our goal, each interview had structured sub-sections beginning with the participants' general perceptions and understanding of AI. We asked participants to share *"in [their] opinion, what do [they] think Artificial Intelligence means?"*, and, *"overall, do [they] think AI is doing more good or more harm for us? Why? In what ways?"* To explore whether participants are able to recognize AI, they were then shown a series of images each containing commonly used AI applications in India across domains [131], and requested to identify if and how those products use AI or not. In the main part of the interview, we used a scenario-based approach (details in the following section to probe experiences, attitudes and intentions to act. For each scenario, we began by presenting a scenario description accompanied with visuals, and then invited participants to share their initial reactions. We explicitly clarified that all applications in the following scenarios are built using AI. Following questions aligned with our goal of understanding participants' acceptance of AI technologies, how it relates to their conceptions of AI, and the factors that influence these perceptions. We asked questions around their beliefs and intentions (*e.g., if AI made a decision on your loan application, would you believe it to be correct?*), their preferences for human *vs.* AI decision-maker (*e.g., What are the differences between a human making a decision on your loan application and an AI making that decision?*, the kinds of information they would like to know about the system, and the level of control they believed the AI systems should have in the given scenario. Within each scenario, we also elicited reactions for a negative outcome for the individual (*e.g.,* loan rejection, wrong medical diagnosis). The final part of the interview focused on reflecting on qualities of an ideal, responsible AI, the kinds of tasks that an AI system should or should not do.

Scenario selection. We draw inspiration from prior research on examining various perceptions of trust [10, 74], fairness [7, 136], justice [13] and explanations [33] in using scenarios[3] to investigate

---

[3]Drawing from Lee *et al.* [74], *"studies have suggested consistency between people's behaviors in scenario-based experiments and their behaviors in real life [137]."*

Type	
Gender	Female **(17)**, Male **(15)**
Location	Delhi **(7)**, Uttar Pradesh **(5)**, Haryana **(4)**, Gujarat **(3)**, Karnataka **(3)**, Maharashtra **(3)**, Tamil Nadu **(2)**, Goa **(1)**, Kerala **(1)**, Odissa **(1)**, Rajasthan **(1)**, Telangana **(1)**
Age	Average **(32.9)**, Standard deviation **(8.6)**, Minimum value **(21)**, Maximum value **(54)**
Highest Education	Higher secondary **(10)**, Bachelor's degree (or equivalent) **(19)**, Master's degree (or higher) **(3)**
Occupation	Small-medium business owner **(6)** (*e.g.,* jewellery shop, electricity shop, tea shop) Full-time consultant **(3)** (*e.g.,* account resolution, legal manager) Full-time logistics employee **(3)** (*e.g.,* shipping, export, manufacturing) Teaching **(2)** (*e.g.,* school teacher, home tuitions) Freelance **(2)** (*e.g.,* content writer) Miscellaneous **(5)** (*e.g.,* non-profit, sales, recruiting, bank manager, executive assistant) Homemaker **(6)**, Farming **(2)**, Unemployed **(3)**
Internet experience	0-2 years **(16)**, 2+ years **(16)**

Table 1: Summary of participant demographics for the interviews, n = 32. We aim for regional diversity in our sample, but do not have enough participants from each city to cross-tabulate the differences across regions.

perceptions. We used four scenarios in the study: (1) loan assessment, (2) financial advisor, (3) medical diagnosis, and (4) hospital finder (details in appendix A). The scenarios were a part of a vignette experiment with 2 (decision-support *vs.* decision-making) x 2 (health *vs.* finance) within-subjects design. We randomized the four scenarios across these participants using Latin Square Design [15] to control for ordering effects, if any.

Healthcare and finance are a visible, explicit policy focus (see NITI Aayog Strategy of 2018 [1], and the latest responsible AI strategy of India [2]), and key areas for private sector deployments [44, 100]. Participants were the subject of the decision-making in these scenarios, and thus, we selected relatively common interaction use cases [13] within which most participants could situate themselves (at present or in the future), even if they had not previously encountered those specific algorithmic systems. We developed the scenarios to represent decision-making with different kinds of agency to affect decision-making (*e.g.,* decision support where the AI makes a recommendation and the user is the deciding agent, and decision-making where the AI makes a decision about or on the behalf of the user). Each scenario was based on real-world examples of AI systems. We used hypothetical scenarios (instead of commonly-used consumer applications) to control for differences in actual prior experiences across participants, so that these effects do not manipulate the results of our study.

Analysis and coding. After transcribing, we translated all the Hindi and Tamil interviews into English, our primary language of analysis. We followed the qualitative data analysis approach by Thomas [123] to conduct multiple rounds of coding at the sentence and paragraph level. We began by open coding instantiations of perceptions, assumptions and expectations of/around AI. Two members of the research team independently read all units multiple times, and identified categories (unit of analysis) together with a description and examples of each category, until a saturation point was reached. Our upper level categories were guided by the evaluation aims, comprising (1) general perceptions about AI, (2)

perceived harms and benefits, (3) sources of AI perceptions, (4) acceptability of AI interventions, (5) willingness to give authority, and (6) human *vs.* AI *vs.* human and AI. These codes were discussed and iteratively refined through meeting, diverging and synthesizing. We consolidated our codes into two top-level categories on defining AI authority, and the sources of AI authority.

Research ethics and anonymization. During recruitment, participants were informed of the purpose of the study, the question categories, and researcher affiliations. Participants signed informed consent documents acknowledging their awareness of the study purpose and researcher affiliation before the interview. At the beginning of each interview, the moderator additionally obtained verbal informed consent. We stored all data in a private Google Drive folder, with access limited to the research team. To protect participant identities, we deleted all personally identifiable information in research files. We redact all identifiable details when quoting participants. Members with research backgrounds in HCI, AI, psychology, and interaction design constitute our research team. Authors located in India moderated the interviews. All researchers were involved in the research framing, data analysis, and synthesis.

3.2 Survey

The interviews revealed that AI authority manifested, in part, as acceptance of AI-based decisions. We further confirm this observation by examining the acceptability of AI outcomes through a survey with a larger audience. Our qualitative research informed the operationalization process, and we use acceptability as the construct for measurement. AI authority is a multi-dimensional concept of which a likelihood to accept AI decisions is only one element. The survey was completed online and received a total of 459 respondents. We used a scenario-based approach for the survey, similar to prior research in this space [7, 10]. The scenarios were a mix of high- and low-stakes situations, which we describe below.

Sample. Members of an online market research panel, Cint were invited to participate in our survey. All participants were based in India, and above 18 years of age (minimum age of consent). We

Type	
Gender	Female **(207)**, Male **(249)**, Non-binary/ third-gender **(0)**, Prefer not to answer **(3)**
Age	18-24 **(110)**, 25-34 **(238)**, 35-44 **(84)**, 45-54 **(22)**, 55 or older **(4)**, Prefer not to say **(1)**
Highest Education	Higher secondary **(96)**, General Graduate **(176)**, Professional Graduate **(187)**
Internet experience	0-5 years **(107)**, 5+ years **(352)**

Table 2: Summary of participant demographics for the survey with 459 respondents

began the survey by eliciting informed consent from respondents. The inclusion criteria for our survey was similar to the interviews. The screener question asked, *"how much do you know about Artificial Intelligence?"* We included respondents that had at least heard of the words Artificial Intelligence. Those who answered, *"Never heard of AI"* were not asked any further questions. After the screening and attention check questions, we were left with 459 respondents that completed the entire survey. Each respondent received $1.70 (INR 150) as compensation for their participation. We describe their demographic details in table 2.

Questionnaire. The survey questionnaire was implemented in Qualtrics. Our questionnaire consisted of 14 questions in total, with a mix of multiple-choice (11) and open-text questions (3). It was divided in four sections: understanding of AI (2), AI acceptability in general (1), scenario-specific AI acceptability (7) and demographics (4). On average, it took respondents 6.5 minutes to complete the online survey. After completing the screener, the first question focused on their understanding of AI. The open-ended question asked, *"how would you explain Artificial Intelligence to a friend?"* We then requested respondents to provide upto three examples of how AI is used today. The seven questions about measuring AI acceptability ranged on a Likert scale of 1 being *"Not at all accepting"* to 5 being *"Extremely accepting"*. The questions were worded with the respondents as the subject of decision-making. Each question on AI acceptability was phrased with a variation of, *"How accepting would you be of a decision made by an AI for you?"*, with the scenarios embedded at the end of the question and highlighted for readability.

After understanding their general willingness to rely on AI for decision-making, we used six scenarios to understand the level of AI acceptability in varying situations/domains: three high-stakes, and three low-stakes. The six scenarios are (1) medical diagnosis, (2) loan approval, (3) hiring decision, (4) song recommendation, (5) route recommendation, and (6) auto rickshaw pricing. We use a within-subjects design, where all respondents were shown all 6 scenarios, presented independent of each other in a randomized fashion to account for any ordering effects. In addition to the Likert scale items, participants were asked to describe *why* they were willing to accept decisions made by an AI, if at all. Finally, we also included demographic questions (optional to answer) about respondent's age, gender, education, and internet exposure.

Analysis. We performed several kinds of analysis to determine the prevalence of AI acceptability across various scenarios. We present two key measures from our analysis in table 2(a), (1) the mean acceptability in our sample, (2) *'significant acceptability'* which represents the percentage of respondents that selected midpoint (>=3 *i.e.,* *"Moderately accepting"*) or above on the Likert scale for AI acceptability. The questions on levels of AI acceptability were

analyzed by calculating percentages and cross-tabulated to view demographic-specific percentages for certain questions. We performed an independent samples t-test to compare AI acceptability between males and females. We conducted a one-way ANOVA to study the influence of age and internet exposure on acceptability. Finally, we performed a qualitative analysis to the open-ended questions using an approach similar to the interviews (see section 3.1). We conducted multiple rounds of coding at the response level in conjunction with participants' survey ratings to surface high-level themes. We include direct quotes from our survey respondents in the Findings with the prefix 'S#', to differentiate from our interview participants prefixed 'P#'.

4 FINDINGS

We describe the concept of AI authority and the various high-risk actions that participants indicated they would be willing to take in section 4.1. Our study identified four user attitudes to AI authority (section 4.2). We then describe the sources through which an AI acquires authority, which lie outside the boundaries of the system (section 4.3).

4.1 What is AI authority?

We draw upon Weber's definition of authority, as a form of power whose use is legitimized by those on whom the power is exercised [132], and extend the work of Lustig and Nardi [81] to define AI authority as the *legitimized power of AI to influence human actions, without requiring adequate evidence about the capabilities of a given system.* AI was seen as reliable and infallible, and thus considered worthy of authority. Participants were willing to trust and accept AI-based outcomes, and perceived AI as a *better alternative* to the human institutions with which they currently interacted. Thus, AI authority emerged as a form of 'voluntary compliance' [127] that was seen as justified by users.

The conceptual metaphors and terminologies can play an important role in our interactions with technology [41]. We broaden the conceptualization of authority from 'algorithmic' [81] to 'Artificial Intelligence' for the unique disposition of 'AI' to gain authority due to the hype, optimism and narratives surrounding its use. The techno-optimistic narratives, among other sources, played a key role in generating legitimacy for AI authority among the participants in our study. The acceptance of AI decisions was often due to reasons beyond system design characteristics and performance. Participants perceived AI as a *panacea*— a tool to prevent discrimination and injustice that they had previously encountered in interactions with traditional institutions, as opposed to the ways in which Bitcoin users in Lustig and Nardi's research [81] viewed the algorithm as a technology of resistance.

Case/Measure		Acceptability (mean)	Significant acceptability
General		3.60 (SD=1.15)	79.2%
High-Stakes			
	Medical Diagnosis	3.37 (SD=1.23)	72.8%
	Loan Approval	3.46 (SD=1.21)	76.2%
	Hiring Decision	3.36 (SD=1.27)	72.3%
Low-Stakes			
	Song Recommendation	3.81 (SD=1.08)	85.0%
	Route Recommendation	3.87 (SD=1.03)	87.3%
	Auto Rickshaw Pricing	3.63 (SD=1.16)	80.2%

Table 3: The level of acceptability of AI-based decisions for the six scenarios from the survey, with a total of 459 respondents. Significant acceptability for a scenario represents the percentage of respondents that selected mid-point (>=3 *i.e.,* 'Moderately accepting') or above on the Likert scale.

We find that over 79% survey respondents reported significantly high acceptance (*i.e.,* selected midpoint (>=3 *"Moderately accepting"*) or above) of decisions/recommendations made by an AI system, averaged across the six scenarios. A willingness to accept AI-based outcomes for respondents in both studies was often guided by the consequences associated with a decision. On average, 84.1% survey respondents were willing to accept decisions in 'low-stakes' AI scenarios, whereas only 73.8% respondents were willing to accept decisions in high-stakes AI scenarios (see table 3). We combined the three high-stakes scenarios (Cronbach's α = 0.83), and three low-stakes scenarios (Cronbach's α = 0.77) into two distinct scales to examine the effects of personal characteristics on acceptability. Gender was relevant for AI acceptability across the three measurements, with females seeing AI systems as significantly more acceptable than males. Finally, we found no significant association between internet exposure and AI acceptability. We present the responses across the acceptability scale for each scenario in Appendix A.

Interview participants also indicated a willingness to take high-risk actions based on outcomes. Several participants noted an acceptance of the medical diagnosis by AI, with an openness to undergo medical treatment, for instance, *"the AI has given me a diagnosis. I will go to the hospital and get the necessary treatment. [...] I have more confidence on an AI instead of a doctor."* (P21) Participants were willing to accept loan assessments, and in fact, preferred that AI evaluated their application over a human loan officer. For the decision-support scenarios, most participants reported an inclination to follow AI recommendations (*e.g.,* visit a hospital, budget income, invest in financial schemes), often mediated by their own judgment. As P11 expressed, *"I would definitely invest in businesses suggested by an AI, because I am 100% sure that it would be the right recommendation."* During each interview, we explored situations in which users might receive opposite recommendations from a human expert *vs.* an AI. Several participants reported not only an intention to act on AI outcomes, but also an inclination to follow AI-based recommendations over those by a doctor or financial advisor. As P7 described, *"I would definitely go with AI if there is a contradiction. There is a clear logic behind an AI."* Overall, participants demonstrated a tendency to accept AI decisions as correct, and believed AI was worthy of authority unless they had evidence to the contrary.

4.2 User attitudes to AI Authority

Participants' trust in the AI applications of the scenarios was underpinned by their beliefs about AI's high competence and benevolence, where different levels of trust could result in diverging intentions or behaviors. AI was considered *reliable*, that it is high-performing and capable of making correct decisions or recommendations. Participants also perceived AI as *infallible* and emotionless— that systems are based on facts, logic, rules and conditions or parameters, for example, a loan assessment AI looks at whether an individual's salary is above a certain threshold to determine whether they are eligible for a loan.

Participants in our study ascribed a positive or neutral intent to AI in their responses (more in 4.2.1), and displayed a *motive-based trust* in the systems presented in the scenarios, which was also observed by Kramer and Tyler [67]. The social conception of trust suggests that people are influenced by their attributions of the motives of authorities [67]— *"attributions of positive intent lead members to trust the authority, and accept the decision onto themselves."* Next, we present the four user attitudes of faith, forgiveness, self-blame, and gratitude, that accompanied the authority of AI, and the ways in which these attitudes affect people's interactions with AI applications. The following attitudes towards AI made participants more likely to accept AI-based outcomes, even in situations where they receive a wrong or unfavorable outcome.

4.2.1 Faith. A commonly held attitude towards AI was a faith in its capabilities and intent, which persisted even when interviewees received an unfavorable outcome. Participants maintained that AI is between 90% to 99% accurate[4], with a recurrent estimate of 95% accurate. For example, P28 suggested that they *"might face an issue 1 out of 100 times. It is very rare."* for a medical diagnosis application. A normative perception across our interviews and survey was that *"computers do not make mistakes like humans do"* (S265). Many participants had the faith that AI would only provide an outcome if it is confident in its abilities. As P9 suggested, *"I will believe an AI because it will not tell me such a big thing that I have a [blood] clot, unless it is confident in itself. "*

[4]Participants did not specifically use technical jargon like accuracy, instead, they expressed the percentage of times an AI system would give the 'right/correct' outcome, which we loosely translate to accuracy.

Participants ascribed neutral or good intentions to AI applications in the scenarios, in contrast to their perceptions of humans with malicious intentions to gain out of their circumstances. Participants suggested that an AI system would have no incentive to cheat or fraud— which P19 described as, *"AI has nothing to gain, not a fixed salary or a commission or incentive. AI is neutral, it is a machine, it is a robot. So the suggestions from an AI will also be good."* Prior work has documented the ways in which 'information' and its use is considered inherently objective, free of personal opinion [6, 99]. Our research confirmed this perspective— AI is seen as fair because it is 'driven by data'. In participants' rationale for accepting AI decisions, the data driven nature of AI was limited to their own data, as opposed to curated training datasets that can be fraught with biases [101]. P2 reported, *"if the system/AI is doing that, it will not take into account that factor of human judgment. In a way, it is good that **everyone will be treated impartially or fairly**. The one who deserves the loan and meets the full requirements, gets the loan. The system will take into account every aspect of your ability to pay the loan."*

In addition, the AI systems in the scenarios were seen as more capable of fair decision-making than human institutions. Where human processes were riddled with inconsistencies and manipulation, AI was seen as simply a rule-following, clearly specifiable system that took into consideration every parameter that should be a part of the decision making process, and *thus fair*. P12 expressed their intention to accept a loan assessment AI over a human officer because, *"officers make you go through so many procedures for a single approval. It will be easy if you have connections at the office. Otherwise, they will ask you to visit one counter after another, and make you wait in long queues. It is simply exhausting."* Colquitt presented four types of organizational justice [24]: distributive, procedural, interactional, informational. Participants consider the procedures, workflows, practices of institutions as a frequent source of unfairness (procedural and interactional injustice). These findings point to a need to reorient our research from an exclusive focus on outcomes towards a approach which takes into consideration the entire procedure of receiving an outcome and the various interactions it involves.

4.2.2 Forgiveness. Several participants demonstrated an inclination to forgive or justify 'wrong' decisions or incorrect recommendations, and give the system multiple opportunities before reducing or terminating use. P19 expressed that it would not be fair to blame the hospital finder AI if they did not have a great experience at the hospital: *"if the AI gave a recommendation, I visited the hospital and did not like the doctor's treatment, then that is not the AI's fault. That is my personal judgment. AI cannot tell me that the doctor will behave well or give the right treatment. "* The 'forgiveness' towards AI was greater in the decision-support scenarios because participants had a 'choice' among the recommendations. For example, if the user selected between the five schemes recommended by the AI, and the recommendation did not work well for them, then they did not blame the AI because they believed they made a choice among those options.

Several participants did not view AI itself as a frequent source of errors or biases. The forgiveness towards AI partly arose out of participants' under-recognition of the range of system errors and biases embedded in AI systems. Instead, participants humanized the AI development process by recognizing the high-level system design decisions (*e.g., "what conditions are used" (P24)*), the collection and input of data— for which institutions or other individuals are responsible. If an AI made the wrong decision, participants hypothesized that it could be the fault of the system developer (*e.g., "low-level mistakes by the programmer" (P25)*), or the institution engaging the system (*e.g., "hospital did not update the information"*). A few participants acknowledged the ways in which the data could be manipulated (*e.g.,* fake reviews or institutions intentionally sharing the wrong information). *"I would be disappointed, but I would not blame it completely. It is a machine after all. It is trained to do things in a certain way. [whom would you blame then?] Maybe the programmer. [laughs] They have not trained the computers to do the hospital search in a better way."*

4.2.3 Self-blame. Participants believed that AI is infallible, that it makes the right decisions and gives the correct outcomes. Users perceived that AI is built by specifying conditions (like a rule-based system, if-else), and there was not much scope of questioning the AI outcomes. This notion of being based on clear conditions manifested into self-blame. Participants consistently blamed themselves for receiving an adverse outcome, especially in the loan assessment scenario. As P32 described, *"If we give proper documents, it [AI] will give loan, else there must be some problem with our documents."* Users conjectured that an unfavorable outcome was their error, because an AI rarely ever makes mistakes. For example, participants had a tendency to believe a wrong outcome by AI meant they did not correctly input their medical history into the health application, or entered the wrong location for finding a hospital, or did not upload the necessary documents for financial advising.

Participants viewed AI systems as emotionless, logical entities. As P18 mentioned, *"the decision is right if we keep our emotions aside and think logically. If they [institutions/developers] have certain rules, then they will decide based on those rules. If we did not fulfill those requirements, then we did not receive the loan."* Moreover, in some cases, participants believed that they deserved to receive an unfavorable outcome. For instance, users speculated that a loan rejection was *"based on [their] transactions"* (P4), and meant they had ineligible finances or collateral to receive a loan. Overall, participants had a low recognition of the potential for AI errors and biases, and instead held themselves or institutions accountable for unfavorable outcomes. Self-blame was a recurring attitude which we observed among our participants, but it frequently had no basis from the scenario. There is an urgent need to combat self-blame (some potential approaches in section 5), or else users might not recognize when an outcome is biased, unfair. Users might not seek recourse or alternative opportunities if they believe that they received an unfavorable outcome, but one that is deserved.

4.2.4 Gratitude and benevolence. Participants felt a sense of loyalty and gratitude towards AI— that they were indebted to AI for the convenience it afforded and reducing dependence on others. 41 survey respondents mentioned 'ease' and convenience within their rationale for accepting AI-based decisions. AI, and technology in general, helped several participants feel independent, for they did not need to rely on others for 'trivial' tasks such as navigating to a location or finding some information. This indebtedness

and gratitude was more noticeable with newer internet users, who shared instances of feeling liberated through their use of AI. Participants' prior interactions with general-purpose AI systems (*e.g.,* maps, search, voice to text) generated this attitude of indebtedness. In this way, the authority of transferred from a general conception of AI as a technology to more specific AI use cases that we discussed in our interviews.

Participants often equated Artificial intelligence with information-finding tools or voice technologies, for instance, P9 expressed, *"I bought this smartphone recently. The age at which I am, if I had to ask something, I used to go to 10-12th class students. Now AI has made it easier for me, as I do not feel any awkwardness in asking anything— it is artificial afterall. "* Applications built using AI had showed them promise, and transformed their lives in many ways. As P11 described, *"I fully believe AI. Till now I have used it for many online tasks and it has never given me the wrong result. Whatever it says will be the right thing."* Participants were willing to act upon AI decisions because their prior interactions with AI were reliable, and used those instances as justification for conferring authority to AI, even if these interactions were with completely different systems. P11 recounted their experience with a Maps application while discussing the medical diagnosis scenario and how that solidified their trust in AI applications: *"I like all things created with AI. While doing field work, on some nights I had to stay outside till 12 am. It [AI] helped me reach home safely. I can trust it with my eyes closed. "*

Beyond these sentiments of indebtedness, several participants reported a tendency for benevolence in sharing their data. While interviewees wanted more control over the use of their data, including knowing what data is collected/used, most of them felt comfortable sharing the data in exchange for reliable decisions or recommendations. In participants' perspectives, good outcomes necessitated sharing various kinds of information. P22 discussed their current use of a health app for managing PCOD, and that they were comfortable sharing their health information because *"the information that they give is valuable to me."* Users viewed AI systems as a 'safe space' free of judgment, and several interviewees were more comfortable sharing their personal data with an AI, 'a machine' over humans. As P24 discussed, *"Most people are not comfortable to discuss their financial needs for loan. There are certain personal questions that a bank might ask me. I am answering questions at my home for an AI. A machine is asking me a question, and I am not being encroached or judged for my choices even though my information is going out."*

4.3 Sources of AI authority

In this section, we present four sources through which AI acquired authority. A propensity to confer authority to AI was influenced by: (1) interactions with ineffectual institutions, (2) techno-optimistic narratives and misleading terminologies around AI, (3) users' prior interactions with applications, and (4) unavailability of alternative systems. AI authority was often derived through one or all of the above sources, often a combination (*e.g.,* a good experience with a different healthcare mobile application, and the unavailability of a medical doctor in their village). We emphasize that AI authority could be misplaced: if AI derives legitimacy and influence over people's actions through sources which lie outside the AI system
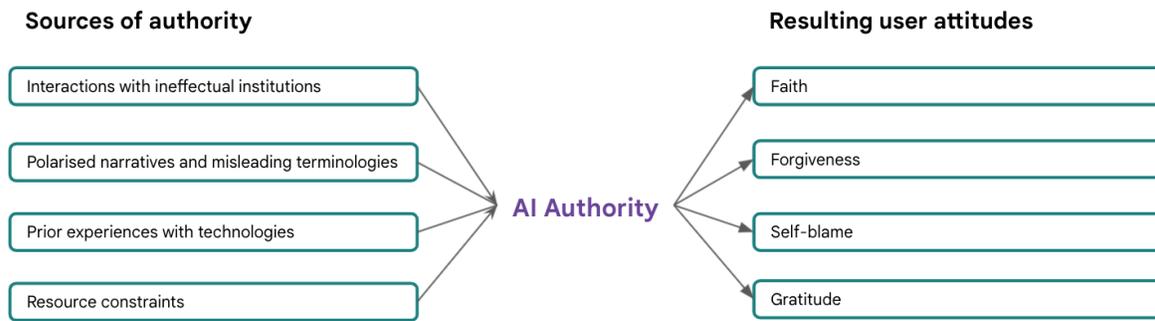
and are *unindicative* of the efficacy of the given system, then an intention to act upon AI decisions introduces a potential for individual harm. These sources of AI authority are extraneous to the system— they provide no evidence that the system will be effective, and moreover, unbiased to the user. In reality, an AI system could be completely dysfunctional (poor performance, unfair), and still gain authority among users.

4.3.1 Interactions with ineffectual institutions. Many participants were exasperated by the discriminatory practices and unjust interactions with human systems. While users appreciated the perceived qualities of AI, they often projected AI as a better alternative to human systems. This contrasts with prior work by Lee and Rich [77] which reported that users in the US with cultural mistrust in doctors consider healthcare AI equally untrustworthy. We find that that authority represented a balancing act: AI was *more authority-worthy*, a better alternative because human institutions were perceived as ineffectual. For example, participants found it unfair when business representatives directed patients to specific hospitals to receive a small fee, or financial planners recommended certain products to meet their quarterly targets. Several interviewees narrated their experiences with bureaucratic decision-making institutions (not just banks), with corruptible officers demanding a bribe as the only way to approve their application. AI was consistently seen as a mechanism to avoid encountering corrupt or prejudiced practices in their interactions with institutions. As P26 expressed,

> *"if a lower middle class individual visits a bank, they ask for so much documentation which you cannot fulfill. The biggest reason for this is that bribes work in many banks for sanctioning loans in India. I have faced this myself. If you go to a bank, the mediators will get a hold of you and ask for 15-20% of the loan amount. Then they will clear all your documents. If there is an AI in this place, then there would not be any issue."*

Participants' mental models of how institutions functioned also influenced their willingness to give authority to an AI. Human systems were perceived as flexible and open to negotiation— *"they can bend a few rules and make it work"* (P2). Human institutions often used appearance and identities as proxies for decision-making, which was perceived as unfair, and an aspect that could be detached from AI systems. P6 described how, *"AI works on logical thinking, not sentimental thinking. Unlike an officer, AI does not have a predetermined image of anyone— it would not judge them by their appearance or standard of living. "* Participants conferred authority to AI as they believed that an AI would not consider their appearance/identity but only their documents to arrive at a decision.

Building and maintaining interpersonal relationships within institutions helped individuals navigate decision-making situations with ease. AI was a *better* alternative than humans with a perceived 'partiality' towards known acquaintances, friends or family. Overall, participants considered human institutions as prone to biases, and thus, a human-AI collaborative decision would be equivalent to a human-only decision susceptible to manipulation. Several interviewees described the concept of human-in-the-loop as 'interference' (here, humans supervising AI decisions). Users preferred that *only* AI made the decision to avoid any form of exploitation, and a human

Sources of authority

Interactions with ineffectual institutions

Polarised narratives and misleading terminologies

Prior experiences with technologies

Resource constraints

AI Authority

Resulting user attitudes

Faith

Forgiveness

Self-blame

Gratitude

Figure 1: AI derives authority through sources that lay outside the system: users' interactions with ineffectual institutions, techno-optimistic narratives and misleading terminologies, users' prior experiences with technology, and the resource constraints with human systems. Finally, AI authority resulted in user attitudes of faith, forgiveness, self-blame and gratitude that made them vulnerable to harm.

can be involved before or after the decision is finalized (*e.g.,* guiding/supporting through the procedure or an unfavorable outcome, receiving feedback on the recommendation or decision).

4.3.2 Techno-optimistic narratives and misleading terminologies. The narratives about AI that surfaced in our research often originated from media (*e.g.,* Sci-Fi movies like the Terminator), news articles, or government perspectives and initiatives. Most descriptions of AI were polarized and extreme: mostly optimistic, and rarely very pessimistic. The pessimistic accounts (*e.g.,* killer robots) were often too far-fetched for participants to imagine or consider a possible reality. The realistic, shades-of-grey narratives were often missing in participants' descriptions of AI, its benefits and harms. People carried an optimism about AI owing to the breadth of coverage about AI's potential, applicability, and planned use by the government. As P16 described, *"the government has made everything an online system nowadays. That cannot go wrong at all."* Several participants brought up the promise of various technological deployments (*e.g., "Modi is launching a driverless train"* (P28)) as a way to demonstrate their support and interest in adopting, accepting and acting upon AI decisions.

We observed that many interview participants and survey respondents saw AI as futuristic (similar to Kelley *et al.* [60]). AI was marketed as a for-good initiative and perceived as *"very progressive"* (S11). Participants noted the ways in which they believed society could benefit from AI, especially in social good domains (*e.g.,* agriculture, healthcare). Ultimately, AI was considered a course towards modernity, for instance, S58 described AI as the *"the modern technology that will change the world."* The words Artificial Intelligence invoked a peculiar imagination in our participants, of a machine that imitated human intelligence with superior performance and without any biases. Overall, the portrayals and AI narratives lend authority to an AI system.

Research participants frequently associated AI with 'machines' (*e.g.,* ATM, washing machine, ECG machine), which brought unwarranted credibility to its capabilities. Several interview participants compared an AI system with an ECG machine in the medical diagnosis scenario. Participants suggested that if doctors use ECG

machines to conduct medical tests then it is acceptable to use a 'machine' (AI) to make a diagnosis as well. This belief often stemmed from the notion that AI is a computer/machine derivative, and thus, similarly reliable. AI was portrayed a tool of convenience: an AI application would make their lives easier by completing tasks and decision-making faster. AI was perceived as an unbiased tool that would get the job done without the complications and emotions that come with human interactions. Participants that believed that AI lacked emotions, considered that AI would take good, correct, and accurate decisions, even if it was unfavorable to them. As S394 noted, *"because an AI system is more logical and calculative compared to a human trying to make a decision. It is heartless yet accurate in making the decision with no emotions meddling the decision making."*

4.3.3 Prior experiences with technologies. Participants supposed AI is highly reliable often through their prior experiences with AI/non-AI technological systems. Respondents frequently referenced other AI systems (*e.g.,* maps tech, voice assistants) as they discussed the capabilities of the given AI system. As a result, interviewees often port over the performance and safety aspects of relatively reliable non-AI systems on to AI systems, for example, as P2 suggested, *"nowadays, we are using net-banking, and the transactions are pretty safe. So I think this [financial advisor] AI will also work in the same way."* The capabilities of the given AI system were conflated by analogizing technologies in the same domain that are built with different underlying mechanisms (*e.g.,* money transfer *vs.* financial advisor). They believed that the given AI system deserved authority, if other perceived technologies (which they considered often considered AI) had given them good, reliable decisions or recommendations. As P28 described, *"I do not think an AI will make a mistake. Till now whatever we have encountered, it has never done the wrong thing."* Non-AI technologies (*e.g.,* GPS, Internet, ATMs) were often misconstrued as AI systems. The boundaries for what constitutes as AI are often blurry. We find that of importance is which tools are perceived as 'AI-based' and the confidence, if any, such associations instills among users.

We observed that the acceptance of AI transferred through social interactions with friends and family, and through systems released previously by the institution that created this AI system. Several

participants mentioned the ways in which their initial acceptance of systems often relied on experiences of their friends or family. If people they knew had a positive encounter with an AI application, then it would help the participant calibrate authority as well. As P3 suggested, *"it helps me if I hear from colleagues, friends that by using this application they are able to manage finances well."* Another factor that built initial acceptance was if the system was created by an organization that participants' had trusted over the years. Both institutional or social transfer of authority do not determine or indicate how a system will respond to an individual due to the nuances through which AI systems are built [5], or if the current system is built with the same standards/care as earlier systems by the same organization.

4.3.4 Unavailability of alternative systems. AI acquired authority through the unavailability of human institutions to meet users' basic needs. Participants relied on an AI system if their other alternatives (*i.e.,* human systems) are infeasible and/or unavailable due to economic constraints, their location, or the timing. For example, receiving medical treatment or finding financial advice was often cost-prohibitive for several participants. On the other hand, an AI-based medical diagnosis application presented an opportunity to save on the consultation fees of a doctor. As P26 noted, *"doctors charge a minimum fee of 300-400 INR, which we can save if we use AI."* Individuals were, thus, inclined to accept AI-based decisions or recommendations, especially if it was not an emergency or high-stakes situation. A few participants shared a hesitation for utilizing AI in emergency, high-stakes situations (*e.g.,* those requiring first aid).

Traditional public services (*e.g.,* medical, financial) have been historically scarce and challenging to access in remote areas in India [57]. As P22 noted, *"it also depends on whether you are in a rural or urban area, and if this is a small or big money lender."* P21 mentioned that there are no MBBS doctors in their village, and they only have alternative medicine practitioners (*e.g.,* Ayurvedic). Usually, respondents' (and their family members') chronic life-threatening disorders remained undiagnosed from a lack of multi-specialty hospitals in their rural area of residence. P25 described how their friend's Lupus was finally detected after they relocated to a major city in Odisha. For P25, *"AI has tons of information to aggregate and make the result available to people, so they can diagnose any diseases."* AI systems were conferred authority as a result of inadequate availability and a glaring divide in access to services across geographies.

AI systems were believed to be available round-the-clock as per participant's convenience, whereas human systems were available only during certain hours. Even more importantly, AI was associated with faster decision-making. In addition, with the lower turn-around time, participants could explore other options if they did not receive a favorable outcome. For instance, if an AI decides to reject the participant's loan within a few minutes, they could visit another bank. As P4 described, *"there is no delay in using the machine. If it is rejected, it gives a direct answer. If we go to loan officers, they say come tomorrow, and this keeps going on."*

5 DISCUSSION

Our results indicate the high acceptance towards AI where **79%** respondents had intentions to accept AI-based outcomes. AI acquired an authority to influence actions through sources that lay outside the AI system. AI decisions were considered reliable, with a performance close to 95%, when most real-world deployments currently struggle to reach such high performance, especially in high-stakes [109]. AI systems were also seen as infallible and fairer than human decision-making. These conceptions hold true in high-stakes scenarios (*e.g.,* healthcare, finance) as well. Owing to their abundant faith in AI capabilities, participants engaged in self-blame or blaming other individuals for an unfavorable outcome. Acute and over-calibrated authority has negative implications. Participants in our research demonstrated faith and gratitude towards AI for 'improving their lives'. They indicated a willingness to adopt and accept new systems, as they believed AI would provide the right outcome. These attitudes could potentially make users vulnerable to algorithmic abuse.

It is extensively documented that if AI systems are not built with care, they perpetuate various forms of bias and inequalities (*e.g.,* through datafication, feedback loops [12, 23, 35]). The impacts of algorithmic biases can be exacerbated in contexts where people are willing to accept systems and algorithmic decisions deployed without adequate accountability. The fact that many users confer authority to AI could be easily exploited to introduce a system into a large but optimistic user base, to utilize harmful approaches (*e.g.,* data maximization, collecting non-consensual data) without oversight. Even without engaging in overtly malicious practices, creators of these systems might ignore the nuances of the context in which they are deploying, use approaches that worked in the West but are inappropriate in India [8, 108]. In the following subsections, we present implications of AI authority for researchers and designers.

5.1 Calibrate AI authority towards an appropriate level

Participants in our study indicated a propensity to actively rely on AI for decision-making or recommendations. This might seem desirable or align with business goals, indeed, many organizations utilize consumer trust as a metric for success [125, 129]. Greater trust might seem to indicate a well-performing product for users. However, our findings indicate that users' experiences with alternative, human systems could easily confound measurements of trust. Is it possible that a high trust is simply an indication that users place authority in AI instead of existing human institutions? For instance, an application might be dysfunctional, unsafe, or unfair to certain users, but they would still rely on it because AI is perceived as a better alternative. Overall, high perceived trust might not be an indicator about the system performance. Efforts to reduce acceptance of AI outcomes might seem antithetical to business goals: regardless, if authority is well-calibrated over time, then it might mitigate harm, retain users, and increase overall satisfaction with the application, leading to success. Overall, unwarranted authority might initially seem desirable, but it can negatively harm product experience in the long-term, as people continually receive outcomes that do not align with their expectations.

We emphasize the need to calibrate authority towards an appropriate level aligned with the actual trustworthiness of AI [51, 124], instead of a trust built through the proxy, confounding sources that we document in our findings. The gratitude shared by our participants was not isolated to the applications that they were currently using (*e.g.,* social media, navigation, voice assistants), but often extended to all of AI systems, including those in our scenarios. Designers can consider introducing features to communicate the actual capabilities of the system while optimizing for understanding [5]. One can set the right expectations from early-use about the system's capabilities. In particular, designers can consider making the limitations of the system explicit before and during early use, by describing the ways in which a system operates and arrives at a decision, offering examples of situations in which the system is likely to provide unreliable results (see the PAIR Guidebook [94], and the Microsoft Human-AI Toolkit [4]). More research is needed to discover the nuances of the ways in which users might adjust these components of trustworthiness with continued interactions with a system. Future work can consider designing alternative metrics for measuring success, beyond user trust in a product or feature, that explicitly takes into account the contextual factors and their prior experiences which contribute to acceptance and AI authority.

5.2 Embrace users' qualitatively different meanings

Individuals' beliefs and understanding was a major factor that contributed to AI authority. Participants conceptualized AI systems as rule-based, as embedded with clear, specifiable conditions. Users' understanding of the ways in which AI works heavily deviates from non-deterministic, deep learning models which are primarily used in current product experiences. Indeed, the communities of HCI and CSCW have long been interested in understanding the ways in which people perceive algorithmic systems (particularly social media applications) [31, 37, 62]. We find that AI is not understood in a standardized manner across cultures, contexts, internet exposures and age groups. In fact, respondents do not come with a 'common denominator', a shared understanding about AI systems when researchers are measuring various aspects of AI perceptions. AI carries a qualitatively different meaning, especially owing to the narratives surrounding its use.

We argue for embracing the variability in research participants' perceptions about AI which heavily influence their intentions and behaviors around algorithmic systems. Perhaps, instead of trying to create a shared meaning of AI or algorithmic systems, could we elicit user conceptions as a basis to guide our analysis of trust, fairness or usefulness? How might we foreground these conceptions or use them to contextualize our research findings? Comparisons of human *vs.* algorithmic decision-making must be foregrounded in participants' beliefs about AI's functioning, which in turn influences their intentions and behaviors. Creating a standardized understanding across users (*e.g.,* by defining AI for the participant) might not reflect real-world situations, or even be effective in eroding/replacing users' own beliefs. In our study, several participants suspected the presence of AI in certain applications, but their reason for believing that those applications are created using AI were frequently distorted. For instance, many people alluded that an

application which requires Internet or GPS is AI-infused. Overall, people might be able to accurately indicate which systems contain AI by conflating non-AI technologies with AI. AI literacy tests may be unable to reveal users' actual competencies or understandings about AI. Combining quantitative tests for AI literacy or awareness with a qualitative approach could probe these beliefs and expectations in a holistic way.

Lateral comparisons of the various constructs of AI perceptions (*e.g.,* acceptability, trust, fairness) across countries must be conducted with extreme caution. Prior research documents the various kinds of survey biases (*e.g.,* acquiescence bias [133], social desirability [121]). People in different cultures have different response styles. Subjective likert scales are often compromised due to response artifacts [91] or the reference group effect; *"the more cultures differ in their norms (i.e., the more cultures are really different on the dimension), the more the cultural comparisons are confounded."* [47]. A straightforward comparison of beliefs and attitudes across cultures with widely differing norms might lead to measurement errors, and thus requires a careful wording of questionnaire design that proactively accounts for variations in response styles and regional differences. For instance, in the EU, perceptions might be very different considering the attention to GDPR [126], and the resulting awareness about algorithmic systems and its biases [60]. This applies to interpretations as well: there is a need to cautiously interpret pre-reported data from different countries, especially if there is a chance that the contextual factors might be lending authority to an AI.

5.3 Build competencies on AI use

Though the research on trust and fairness perceptions offers mixed results (often domain- and task-dependent), several prior studies find that respondents in the EU, US indicate a tendency to trust humans more than algorithmic systems [19, 32, 70], especially for tasks requiring human skills [74] (with some exceptions such as Logg *et al.* [79] and Lee and Rich [77]). In our technological context, we find a case of acceptance of AI outcomes with a confidence greater than a system might deserve. Misplaced authority is consequential: AI systems have potential to cause harm through incorrect or biased outcomes, especially because users might not actively seek out information about the capabilities of the given system. The effects of misplaced authority are exacerbated and far-reaching in high-stakes scenarios (*e.g.,* hiring, finance, social benefits). Several participants considered a lack of human involvement in making the decision as desirable. Even the use of general-purpose products in critical situations (*e.g.,* virtual assistants for job interview reminders) could lead to adverse outcomes for users.

Mayer, Davis, and Schoorman [83] present an integrative model of trustworthiness with three components: ability, integrity, and benevolence. We use this framework to reflect on our findings and present a path forward. Prior work (based in US or EU) reports that users rate technology companies (and products) with low trust (benevolence) [11, 85], but high ability. Therefore, when a technology system goes wrong, it is seen as a benevolence issue. Our results suggest that both—ability and benevolence—components of AI trustworthiness are viewed as high in India. Users are willing to give authority to AI because they have faith in the competence

and benevolence of AI systems. When a system makes a mistake, neither its capabilities (ability) nor its intent (benevolence) were readily questioned. This resulted in self-blame or placing blame on other actors, without a recognition that AI could have made that error. Especially for the loan assessment scenario, many participants in our study believed that either they deserved an unfavorable outcome or could be their own fault. The self-blame was exacerbated when participants indicated a faith in the capabilities of AI. Low or non-contestation of AI outcomes has potential to cause individual harm if people accept decisions which might be wrong. In addition, users reporting decisions or behaviors to the system represents important opportunities for model feedback. This scope for mitigating harm and improving models is lost if users believe the outcomes to be correct, with a possibility of causing harm to other users.

There is a well-established body of work on explainable AI [9, 49], and how explanations impact user trust and reliance [98, 118]. The focus of this research is to explore when to explain (high *vs.* low stakes, *e.g.,* [18]), what to explain (global *vs.* local [40, 69]), and how to explain [138]. To make the system more transparent, there's increasing interest within the HCI, XAI communities to find human-centered approaches for explainability [78]. There are broadly three levels of building competencies on AI use: (1) outcome/in-the-moment explanation of a certain outcome, (2) an understanding of how a given AI application works, (3) general, accessible education about what AI is, how it works, and what are its strengths and limitations. Most explainability approaches have been context-agnostic, however, there needs to be particular attention to emerging internet users that often have less familiarity with technical jargon, often coupled with lower literacy [92]. Additionally, in-the-moment explanations are not always feasible due to legal or usability constraints. Consider a credit card approval application. Designers might be unable to 'explain' the reason behind a particular decision due to anticipated legal issues. Even without legal constraints, explanations are most useful when actionable [56]. A credit card rejection for an applicant with AI authority might mean that they believe that the system made the correct decision. As a result, they might not seek alternative platforms when it could easily be the case that the AI made an error or gave a biased outcome. The onus is on the designers/developers to add safeguards, ensure that users do not share a utopian view of the systems with which they interact [111], and acknowledge the range of system errors that are possible. Another approach to consider is to leverage existing capacities [134] by educating the user about how the system works in general (*e.g.,* [43]) that could be a valuable starting point for users to calibrate their responses based off their general understanding of how an application provided an outcome.

5.4 Imagine and disseminate narratives of contemporary risks

The narratives shared by our respondents were often too fantastical and polarized. They described science fiction narratives (*e.g.,* movies like Terminator), and several respondents exclusively associated AI with robot applications. Participants dismissed the more pessimistic narratives (*e.g.,* of killer robots, see also [20]) which were hard to imagine or relate with. On the other hand, the optimistic narratives

were entertained and adopted. AI was perceived as an innocuous, neutral tool to *'make [their] life easier'*. Our work reinforces the sentiment shared by Cave *et al.* [21] that *"narratives might influence the development and adoption of AI technologies."* Then how might we develop alternative narratives about the ways in which AI technologies are currently used? An important emerging research area is to create and disseminate AI narratives that align with contemporary issues of algorithmic bias. Realistic accounts could gradually help people calibrate their authority by painting a less skewed, lopsided picture of the capabilities of AI.

People generally have a tendency to engage with mainstream media (articles, movies), and our results indicate that the polarized narratives to which people were exposed greatly shaped their acceptance and the authority of AI technologies. Overall, an optimistic portrayal of AI, especially through state-sponsored initiatives, could result in less-than-rigorous technologies making their way into usage by the general public, causing irreversible individual and societal harm. Several participants actively acknowledged the human involvement in AI development, however, we note a faith in the intentions of AI applications— that these systems are built to support users and improve current workflows, with neutral/good intentions, partly owing to the narratives about the benefits of emerging technologies.Several state-sponsored bodies are pushing for AI curriculum in schools and dedicated programs in universities [30, 106], strategy documents are reporting credible information [2, 114], however, it is equally important to disseminate these to a broader audience. For instance, "The Social Dilemma" [68] is a film that gained notoriety in the US, which documents the dangers of AI, specifically, social networking. Mainstream media modalities that educate people about the complexities and contextual concerns about AI, but more importantly initiate dialogue about these topics, might play a crucial role in calibrating AI authority.

5.5 Responsible AI approaches do not generalize across contexts

Traditional auditing methods (*e.g.,* involving researchers, activists) [103, 110] can be inadequate in surfacing harmful behaviors when the auditors lack the cultural backgrounds or lived experiences to recognize if something is problematic [113]. Shen, Devos *et al.* [115] propose the concept of everyday auditing through which users detect and interrogate problematic biases/behaviors through their everyday interactions with algorithmic systems. However, auditing can be premised on informed users with the capacities, and even more, the *intentions* to recognize and report bias. What if users are not likely to seek out instances of biased behaviors? What if, beyond that, users are defensive about AI systems, or unquestioning of its abilities to provide good results? Eslami *et al.* [38] report that users were able to notice bias in the rating platforms. User auditing approaches predicated on skeptical users might not generalize well to contexts that demonstrate techno-optimism and faith towards AI and its capabilities. People with high acceptability of AI might not see the shortcomings of using these systems. Sambasivan *et al.* [107] extensively report the ways in which a straightforward porting of responsible AI tenets can be inadequate and often harmful. How might we involve users with optimistic views about AI into algorithmic audits? Certainly, their

perspectives would be valuable contributions in surfacing biased outcomes. Empowering users to interrogate these systems might be an approach to calibrate AI authority towards an appropriate level. Future work might investigate how we could leverage existing capacities in users that have high confidence in AI to acknowledge bias. This could benefit platforms in two ways: visibilising bias and mitigating harm, but also through building alternative, realistic narratives about AI that are better aligned with the capabilities of the system.

6 LIMITATIONS AND FUTURE WORK

Although our interviews included a diverse sample, it may be subject to common limitations of qualitative studies, including recall and observer bias, and participant self-censorship. Our findings represent the perceptions of our sample, and may not generalize across countries, to groups with different levels of exposures to AI systems or the discourses around AI. Within India itself, there is a wide range of cultural and social associations with AI that could yield varying results about the attitudes towards these systems. In addition, our screening criteria included people that had previously heard of AI, however, this might exclude perspectives from people who had developed conceptions about algorithmic behaviors through their interactions with technological systems, but had not heard of those exact words. Researchers could consider expanding the inclusion criteria to elicit perspectives from users aware of AI decisions but not familiar with the terminologies. We acknowledge that participants may have had different responses if these scenarios had been more commonplace, for example, through the mobile products with which they already interact. Finally, we used a hypothetical scenario-based approach (similar to prior research [7, 75, 135]), so the results might not generalize to newer domains or use cases of AI. Future work could consider investigating acceptance of users of existing high-stakes AI applications (*e.g.,* instant loan applications) or extend the research to other domains of study.

7 CONCLUSION

Our everyday life experiences are increasingly mediated through AI systems. It is, thus, of crucial importance to investigate the ways in which technology users across diverse contexts would respond to or adopt AI-based decisions. We presented a mixed-methods study of perceptions about AI systems (both decision-making and decision-support) in various domains and settings by drawing upon 32 interviews and 459 survey respondents in India. We observed that acceptance of AI is high, and find that attitudes towards these systems indicate a form of authority of AI. AI was afforded a legitimized power by our participants, and it manifested as four user attitudes that could lead the participants vulnerable to abuse. AI is conferred authority through extraneous factors which do not provide evidence for the capabilities of the system. More notably, users give AI an authority because they perceive it as a better alternative to the human institutions with which they currently interact. We urgently call for calibrating AI authority by reconsidering methodological norms success metrics which might be confounding, and through appropriate, alternative narratives for deploying AI systems in India.

REFERENCES

[1] NITI Aayog. 2018. NationalStrategy-for-AI-Discussion-Paper.pdf. https://indiaai.gov.in/documents/pdf/NationalStrategy-for-AI-Discussion-Paper.pdf. (Accessed on 12/03/2021).

[2] NITI Aayog. 2021. Responsible-AI-22022021.pdf. https://www.niti.gov.in/sites/default/files/2021-02/Responsible-AI-22022021.pdf. (Accessed on 12/03/2021).

[3] ACLU. 2016. Community Control Over Police Surveillance.

[4] Microsoft Aether. 2021. Microsoft HAX Toolkit. https://www.microsoft.com/en-us/haxtoolkit/. (Accessed on 09/08/2021).

[5] Saleema Amershi, Dan Weld, Mihaela Vorvoreanu, Adam Fourney, Besmira Nushi, Penny Collisson, Jina Suh, Shamsi Iqbal, Paul N. Bennett, Kori Inkpen, Jaime Teevan, Ruth Kikin-Gil, and Eric Horvitz. 2019. Guidelines for Human-AI Interaction. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (Glasgow, Scotland Uk) *(CHI '19)*. Association for Computing Machinery, New York, NY, USA, 1–13. https://doi.org/10.1145/3290605.3300233

[6] C.W. Anderson. 2018. *Apostles of Certainty: Data Journalism and the Politics of Doubt.* Oxford University Press, New York, USA. https://books.google.co.in/books?id=53ZoDwAAQBAJ

[7] Theo Araujo, Natali Helberger, Sanne Kruikemeier, and Claes H De Vreese. 2020. In AI we trust? Perceptions about automated decision-making by artificial intelligence. *AI & SOCIETY* 35, 3 (2020), 611–623.

[8] Payal Arora. 2016. Bottom of the data pyramid: Big data and the global south. *International Journal of Communication* 10 (2016), 19.

[9] Vijay Arya, Rachel K. E. Bellamy, Pin-Yu Chen, Amit Dhurandhar, Michael Hind, Samuel C. Hoffman, Stephanie Houde, Q. Vera Liao, Ronny Luss, Aleksandra Mojsilovic, Sami Mourad, Pablo Pedemonte, Ramya Raghavendra, John T. Richards, Prasanna Sattigeri, Karthikeyan Shanmugam, Moninder Singh, Kush R. Varshney, Dennis Wei, and Yunfeng Zhang. 2019. One Explanation Does Not Fit All: A Toolkit and Taxonomy of AI Explainability Techniques. *CoRR* abs/1909.03012 (2019), 1–18. arXiv:1909.03012 http://arxiv.org/abs/1909.03012

[10] Maryam Ashoori and Justin D. Weisz. 2019. In AI We Trust? Factors That Influence Trustworthiness of AI-infused Decision-Making Processes. arXiv:1912.02675 http://arxiv.org/abs/1912.02675

[11] Brooke Auxier. 2020. How Americans view U.S. tech companies in 2020 | Pew Research Center. https://www.pewresearch.org/fact-tank/2020/10/27/how-americans-see-u-s-tech-companies-as-government-scrutiny-increases/. (Accessed on 09/06/2021).

[12] Solon Barocas and Andrew D Selbst. 2016. Big data's disparate impact. *Calif. L. Rev.* 104 (2016), 671.

[13] Reuben Binns, Max Van Kleek, Michael Veale, Ulrik Lyngs, Jun Zhao, and Nigel Shadbolt. 2018. 'It's Reducing a Human Being to a Percentage': Perceptions of Justice in Algorithmic Decisions. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems.* Association for Computing Machinery, New York, NY, USA, 1–14. https://doi.org/10.1145/3173574.3173951

[14] Lia Bozarth and Joyojeet Pal. 2019. Twitter Discourse as a Lens into Politicians' Interest in Technology and Development. In *Proceedings of the Tenth International Conference on Information and Communication Technologies and Development* (Ahmedabad, India) *(ICTD '19)*. Association for Computing Machinery, New York, NY, USA, Article 33, 5 pages. https://doi.org/10.1145/3287098.3287129

[15] James V Bradley. 1958. Complete counterbalancing of immediate sequential effects in a Latin square design. *J. Amer. Statist. Assoc.* 53, 282 (1958), 525–528.

[16] Eric Brewer, Michael Demmer, Bowei Du, Melissa Ho, Matthew Kam, Sergiu Nedevschi, Joyojeet Pal, Rabin Patra, Sonesh Surana, and Kevin Fall. 2005. The case for technology in developing regions. *Computer* 38, 6 (2005), 25–38.

[17] Taina Bucher. 2017. The algorithmic imaginary: exploring the ordinary affects of Facebook algorithms. *Information, communication & society* 20, 1 (2017), 30–44.

[18] Andrea Bunt, Matthew Lount, and Catherine Lauzon. 2012. Are Explanations Always Important? A Study of Deployed, Low-Cost Intelligent Interactive Systems. In *Proceedings of the 2012 ACM International Conference on Intelligent User Interfaces* (Lisbon, Portugal) *(IUI '12)*. Association for Computing Machinery, New York, NY, USA, 169–178. https://doi.org/10.1145/2166966.2166996

[19] Noah Castelo, Maarten W Bos, and Donald R Lehmann. 2019. Task-dependent algorithm aversion. *Journal of Marketing Research* 56, 5 (2019), 809–825.

[20] Stephen Cave, Kate Coughlan, and Kanta Dihal. 2019. "Scary Robots": Examining Public Responses to AI. In *Proceedings of the 2019 AAAI/ACM Conference on AI, Ethics, and Society* (Honolulu, HI, USA) *(AIES '19)*. Association for Computing Machinery, New York, NY, USA, 331–337. https://doi.org/10.1145/3306618.3314232

[21] Stephen Cave, Claire Craig, Kanta Dihal, Sarah Dillon, Jessica Montgomery, Beth Singler, and Lindsay Taylor. 2018. Portrayals and perceptions of AI and why they matter.

[22] Shuchih Ernest Chang, Anne Yenching Liu, and Wei Cheng Shen. 2017. User trust in social networking services: A comparison of Facebook and LinkedIn. *Computers in Human Behavior* 69 (2017), 207–217.

[23] Alexandra Chouldechova. 2017. Fair prediction with disparate impact: A study of bias in recidivism prediction instruments. *Big data* 5, 2 (2017), 153–163.

[24] Jason A Colquitt. 2001. On the dimensionality of organizational justice: a construct validation of a measure. *Journal of applied psychology* 86, 3 (2001), 386.

[25] European Commission. 2021. Regulation Of The European Parliament And Of The Council - Laying Down Harmonised Rules On Artificial Intelligence (Artificial Intelligence Act) And Amending Certain Union Legislative Acts. https://eur-lex.europa.eu/legal-content/EN/TXT/?qid=1623335154975&uri=CELEX%3A52021PC0206. (Accessed on 09/09/2021).

[26] Henriette Cramer, V. Evers, Satyan Ramlal, M. Someren, L. Rutledge, N. Stash, Lora Aroyo, and B. Wielinga. 2008. The effects of transparency on trust in and acceptance of a content-based art recommender. *User Modeling and User-Adapted Interaction* 18 (2008), 455–496.

[27] Aman Dalmia, Jerome White, Ankit Chaurasia, Vishal Agarwal, Rajesh Jain, Dhruvin Vora, Balasaheb Dhame, Raghu Dharmaraju, and Rahul Panicker. 2020. Pest Management In Cotton Farms: An AI-System Case Study from the Global South. In *Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining* (Virtual Event, CA, USA) *(KDD '20)*. Association for Computing Machinery, New York, NY, USA, 3119–3127. https://doi.org/10.1145/3394486.3403363

[28] Roy G d'Andrade. 1995. *The development of cognitive anthropology.* Cambridge University Press, Cambridge, UK.

[29] Maartje de Graaf, Somaya Ben Allouch, and Jan van Dijk. 2017. Why Do They Refuse to Use My Robot? Reasons for Non-Use Derived from a Long-Term Home Study. In *Proceedings of the 2017 ACM/IEEE International Conference on Human-Robot Interaction* (Vienna, Austria) *(HRI '17)*. Association for Computing Machinery, New York, NY, USA, 224–233. https://doi.org/10.1145/2909824.3020236

[30] Srishti Deoras. 2020. CBSE Integrates AI Curriculum In 200 Indian Schools In Collaboration With IBM. https://analyticsindiamag.com/cbse-integrates-ai-curriculum-in-200-indian-schools-in-collaboration-with-ibm/. (Accessed on 08/22/2021).

[31] Michael A. DeVito, Darren Gergle, and Jeremy Birnholtz. 2017. "Algorithms Ruin Everything": #RIPTwitter, Folk Theories, and Resistance to Algorithmic Change in Social Media. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems* (Denver, Colorado, USA) *(CHI '17)*. Association for Computing Machinery, New York, NY, USA, 3163–3174. https://doi.org/10.1145/3025453.3025659

[32] Berkeley J Dietvorst, Joseph P Simmons, and Cade Massey. 2015. Algorithm aversion: People erroneously avoid algorithms after seeing them err. *Journal of Experimental Psychology: General* 144, 1 (2015), 114.

[33] Jonathan Dodge, Q. Vera Liao, Yunfeng Zhang, Rachel K. E. Bellamy, and Casey Dugan. 2019. Explaining Models: An Empirical Study of How Explanations Impact Fairness Judgment. In *Proceedings of the 24th International Conference on Intelligent User Interfaces* (Marina del Ray, California) *(IUI '19)*. Association for Computing Machinery, New York, NY, USA, 275–285. https://doi.org/10.1145/3301275.3302310

[34] Graham Dove, Kim Halskov, Jodi Forlizzi, and John Zimmerman. 2017. UX Design Innovation: Challenges for Working with Machine Learning as a Design Material. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems* (Denver, Colorado, USA) *(CHI '17)*. Association for Computing Machinery, New York, NY, USA, 278–288. https://doi.org/10.1145/3025453.3025739

[35] Danielle Ensign, Sorelle A. Friedler, Scott Neville, Carlos Scheidegger, and Suresh Venkatasubramanian. 2018. Runaway Feedback Loops in Predictive Policing. In *Proceedings of the 1st Conference on Fairness, Accountability and Transparency* *(Proceedings of Machine Learning Research, Vol. 81)*, Sorelle A. Friedler and Christo Wilson (Eds.). PMLR, New York, NY, USA, 160–171. https://proceedings.mlr.press/v81/ensign18a.html

[36] Motahhare Eslami, Karrie Karahalios, Christian Sandvig, Kristen Vaccaro, Aimee Rickman, Kevin Hamilton, and Alex Kirlik. 2016. First I "like" It, Then I Hide It: Folk Theories of Social Feeds. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems* (San Jose, California, USA) *(CHI '16)*. Association for Computing Machinery, New York, NY, USA, 2371–2382. https://doi.org/10.1145/2858036.2858494

[37] Motahhare Eslami, Aimee Rickman, Kristen Vaccaro, Amirhossein Aleyasen, Andy Vuong, Karrie Karahalios, Kevin Hamilton, and Christian Sandvig. 2015. "I Always Assumed That I Wasn't Really That Close to [Her]": Reasoning about Invisible Algorithms in News Feeds. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems* (Seoul, Republic of Korea) *(CHI '15)*. Association for Computing Machinery, New York, NY, USA, 153–162. https://doi.org/10.1145/2702123.2702556

[38] Motahhare Eslami, Kristen Vaccaro, Karrie Karahalios, and Kevin Hamilton. 2017. "Be Careful; Things Can Be Worse than They Appear": Understanding Biased Algorithms and Users' Behavior Around Them in Rating Platforms, In Proceedings of the Eleventh International AAAI Conference on Web and Social Media. *Proceedings of the International AAAI Conference on Web and Social Media* 11, 1, 62–71. https://ojs.aaai.org/index.php/ICWSM/article/view/14898

[39] Ethan Fast and Eric Horvitz. 2017. Long-Term Trends in the Public Perception of Artificial Intelligence. In *Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence (AAAI'17)*. AAAI Press, San Francisco, California, USA, 963–969.

[40] Shi Feng and Jordan Boyd-Graber. 2019. What Can AI Do for Me? Evaluating Machine Learning Interpretations in Cooperative Play. In *Proceedings of the 24th International Conference on Intelligent User Interfaces* (Marina del Ray, California) *(IUI '19)*. Association for Computing Machinery, New York, NY, USA, 229–239. https://doi.org/10.1145/3301275.3302265

[41] Giselle Martins dos Santos Ferreira, Luiz Alexandre da Silva Rosado, Márcio Silveira Lemgruber, and Jaciara de Sá Carvalho. 2020. Metaphors we're colonised by? The case of data-driven educational technologies in Brazil. *Learning, Media and Technology* 45, 1 (2020), 46–60.

[42] Martin Fishbein and Icek Ajzen. 1977. Belief, Attitude, Intention, and Behavior: An Introduction to Theory and Research. *Philosophy and Rhetoric* 10, 2 (1977), 130–132.

[43] Google. 2019. How Google Search Works (in 5 minutes) - YouTube. https://www.youtube.com/watch?v=0eKVizvYSUQ. (Accessed on 09/06/2021).

[44] Saman Goudarzi, Elonnai Hickok, Amber Sinha, S Mohandas, PM Bidare, S Ray, and A Rathi. 2018. AI in Banking and Finance.

[45] Eszter Hargittai, Jonathan Gruber, Teodora Djukaric, Jaelle Fuchs, and Lisa Brombach. 2020. Black box measures? How to study people's algorithm skills. *Information, Communication & Society* 23, 5 (2020), 764–775.

[46] Eszter Hargittai and Aaron Shaw. 2020. Comparing internet experiences and prosociality in amazon mechanical turk and population-based survey samples. *Socius* 6 (2020), 2378023119889834.

[47] Steven J Heine, Darrin R Lehman, Kaiping Peng, and Joe Greenholtz. 2002. What's wrong with cross-cultural comparisons of subjective Likert scales?: The reference-group effect. *Journal of personality and social psychology* 82, 6 (2002), 903.

[48] Miriam Höddinghaus, Dominik Sondern, and Guido Hertel. 2021. The automation of leadership functions: Would people trust decision algorithms? *Computers in Human Behavior* 116 (2021), 106635.

[49] Andreas Holzinger. 2018. Explainable ai (ex-ai). *Informatik-Spektrum* 41, 2 (2018), 138–143.

[50] Nataliya V Ivankova and John W Creswell. 2009. Mixed methods. *Qualitative research in applied linguistics: A practical introduction* 23 (2009), 135–161.

[51] Alon Jacovi, Ana Marasović, Tim Miller, and Yoav Goldberg. 2021. Formalizing Trust in Artificial Intelligence: Prerequisites, Causes and Goals of Human Trust in AI. In *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency* (Virtual Event, Canada) *(FAccT '21)*. Association for Computing Machinery, New York, NY, USA, 624–635. https://doi.org/10.1145/3442188.3445923

[52] Mohit Jain, Pratyush Kumar, Ishita Bhansali, Q Vera Liao, Khai Truong, and Shwetak Patel. 2018. FarmChat: a conversational agent to answer farmer queries. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 2, 4 (2018), 1–22.

[53] Mohammad Hossein Jarrahi and Will Sutherland. 2019. Algorithmic Management and Algorithmic Competencies: Understanding and Appropriating Algorithms in Gig Work. In *Information in Contemporary Society*, Natalie Greene Taylor, Caitlin Christian-Lamb, Michelle H. Martin, and Bonnie Nardi (Eds.). Springer International Publishing, Cham, 578–589.

[54] Pramile Jayapal. 2020. Jayapal Joins Colleagues In Introducing Bicameral Legislation to Ban Government Use of Facial Recognition, Other Biometric Technology - Congresswoman Pramila Jayapal. https://jayapal.house.gov/2020/06/25/jayapal-joins-rep-pressley-and-senators-markey-and-merkley-to-introduce-legislation-to-ban-government-use-of-facial-recognition-other-biometric-technology/. (Accessed on 09/09/2021).

[55] Shagun Jhaver, Yoni Karpfen, and Judd Antin. 2018. Algorithmic Anxiety and Coping Strategies of Airbnb Hosts. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. Association for Computing Machinery, New York, NY, USA, 1–12. https://doi.org/10.1145/3173574.3173995

[56] Shalmali Joshi, Oluwasanmi Koyejo, Warut Vijitbenjaronk, Been Kim, and Joydeep Ghosh. 2019. Towards Realistic Individual Recourse and Actionable Explanations in Black-Box Decision Making Systems. *CoRR* abs/1907.09615 (2019), 1–19. arXiv:1907.09615 http://arxiv.org/abs/1907.09615

[57] Anup Karan, Himanshu Negandhi, Suhaib Hussain, Tomas Zapata, Dilip Mairembam, Hilde De Graeve, James Buchan, and Sanjay Zodpey. 2021. Size, composition and distribution of health workforce in India: why, and where to invest? *Human resources for health* 19, 1 (2021), 1–14.

[58] Herbert W Kee and Robert E Knox. 1970. Conceptual and methodological considerations in the study of trust and suspicion. *Journal of conflict resolution* 14, 3 (1970), 357–366.

[59] Sandhya Keelery. 2021. Internet usage in India - statistics & facts | Statista. https://www.statista.com/topics/2157/internet-usage-in-india/. (Accessed on 09/09/2021).

[60] Patrick Gage Kelley, Yongwei Yang, Courtney Heldreth, Christopher Moessner, Aaron Sedley, Andreas Kramm, David T. Newman, and Allison Woodruff. 2021. *Exciting, Useful, Worrying, Futuristic: Public Perception of Artificial Intelligence in 8 Countries.* Association for Computing Machinery, New York, NY, USA, 627–637. https://doi.org/10.1145/3461702.3462605

[61] David Kipnis. 1996. Trust and technology. *Trust in organizations: Frontiers of theory and research* 39 (1996), 50.

[62] Erin Klawitter and Eszter Hargittai. 2018. "It's like learning a whole other language": The role of algorithmic skills in the curation of creative goods. *International Journal of Communication* 12 (2018), 3490–3510.

[63] Rafal Kocielnik, Saleema Amershi, and Paul N. Bennett. 2019. 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* (Glasgow, Scotland Uk) *(CHI '19).* Association for Computing Machinery, New York, NY, USA, 1–14. https://doi.org/10.1145/3290605.3300641

[64] Ava Kofman. 2016. How facial recognition can ruin your life–intercept.

[65] Anastasia Kozyreva, Stefan Herzog, Philipp Lorenz-Spreen, Ralph Hertwig, and Stephan Lewandowsky. 2020. Artificial intelligence in online environments: Representative survey of public attitudes in germany.

[66] Anastasia Kozyreva, Philipp Lorenz-Spreen, Ralph Hertwig, Stephan Lewandowsky, and Stefan M Herzog. 2021. Public attitudes towards algorithmic personalization and use of personal data online: evidence from Germany, Great Britain, and the United States. *Humanities and Social Sciences Communications* 8, 1 (2021), 1–11.

[67] R.M. Kramer and T.R. Tyler. 1995. *Trust in Organizations: Frontiers of Theory and Research.* SAGE Publications, Thousand Oaks, CA. https://books.google.co.in/books?id=ddpyAwAAQBAJ

[68] Exposure Labs. 2020. The Social Dilemma. https://www.thesocialdilemma.com/. (Accessed on 09/10/2021).

[69] Himabindu Lakkaraju, Ece Kamar, Rich Caruana, and Jure Leskovec. 2019. Faithful and Customizable Explanations of Black Box Models. In *Proceedings of the 2019 AAAI/ACM Conference on AI, Ethics, and Society* (Honolulu, HI, USA) *(AIES '19).* Association for Computing Machinery, New York, NY, USA, 131–138. https://doi.org/10.1145/3306618.3314229

[70] Markus Langer, Cornelius J König, and Maria Papathanasiou. 2019. Highly automated job interviews: Acceptance under the influence of stakes. *International Journal of Selection and Assessment* 27, 3 (2019), 217–234.

[71] Markus Langer and Richard N. Landers. 2021. The future of artificial intelligence at work: A review on effects of decision automation and augmentation on workers targeted by algorithms and third-party observers. *Computers in Human Behavior* 123 (2021), 106878. https://doi.org/10.1016/j.chb.2021.106878

[72] Chia-Jung Lee, Jaime Teevan, and Sebastian de la Chica. 2014. Characterizing Multi-Click Search Behavior and the Risks and Opportunities of Changing Results during Use. In *Proceedings of the 37th International ACM SIGIR Conference on Research & Development in Information Retrieval* (Gold Coast, Queensland, Australia) *(SIGIR '14).* Association for Computing Machinery, New York, NY, USA, 515–524. https://doi.org/10.1145/2600428.2609588

[73] John D Lee and Katrina A See. 2004. Trust in automation: Designing for appropriate reliance. *Human factors* 46, 1 (2004), 50–80.

[74] Min Kyung Lee. 2018. Understanding perception of algorithmic decisions: Fairness, trust, and emotion in response to algorithmic management. *Big Data & Society* 5, 1 (2018), 2053951718756684.

[75] Min Kyung Lee, Ji Tae Kim, and Leah Lizarondo. 2017. A Human-Centered Approach to Algorithmic Services: Considerations for Fair and Motivating Smart Community Service Management That Allocates Donations to Non-Profit Organizations. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems* (Denver, Colorado, USA) *(CHI '17).* Association for Computing Machinery, New York, NY, USA, 3365–3376. https://doi.org/10.1145/3025453.3025884

[76] Min Kyung Lee, Daniel Kusbit, Evan Metsky, and Laura Dabbish. 2015. Working with Machines: The Impact of Algorithmic and Data-Driven Management on Human Workers. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems* (Seoul, Republic of Korea) *(CHI '15).* Association for Computing Machinery, New York, NY, USA, 1603–1612. https://doi.org/10.1145/2702123.2702548

[77] Min Kyung Lee and Katherine Rich. 2021. Who Is Included in Human Perceptions of AI?: Trust and Perceived Fairness around Healthcare AI and Cultural Mistrust. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* (Yokohama, Japan) *(CHI '21).* Association for Computing Machinery, New York, NY, USA, Article 138, 14 pages. https://doi.org/10.1145/3411764.3445570

[78] Q. Vera Liao, Daniel Gruen, and Sarah Miller. 2020. Questioning the AI: Informing Design Practices for Explainable AI User Experiences. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems.* Association for Computing Machinery, New York, NY, USA, 1–15. https://doi.org/10.1145/3313831.3376590

[79] Jennifer M Logg, Julia A Minson, and Don A Moore. 2019. Algorithm appreciation: People prefer algorithmic to human judgment. *Organizational Behavior and Human Decision Processes* 151 (2019), 90–103.

[80] Chiara Longoni, Andrea Bonezzi, and Carey K Morewedge. 2019. Resistance to medical artificial intelligence. *Journal of Consumer Research* 46, 4 (2019), 629–650.

[81] Caitlin Lustig and Bonnie Nardi. 2015. Algorithmic Authority: The Case of Bitcoin. In *2015 48th Hawaii International Conference on System Sciences.* IEEE, Hawaii, USA, 743–752. https://doi.org/10.1109/HICSS.2015.95

[82] Vidushi Marda. 2018. Artificial intelligence policy in India: a framework for engaging the limits of data-driven decision-making. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* 376, 2133 (2018), 20180087.

[83] Roger C Mayer, James H Davis, and F David Schoorman. 1995. An integrative model of organizational trust. *Academy of management review* 20, 3 (1995), 709–734.

[84] Indrani Medhi Thies, Nandita Menon, Sneha Magapu, Manisha Subramony, and Jacki O'Neill. 2017. How Do You Want Your Chatbot? An Exploratory Wizard-of-Oz Study with Young, Urban Indians. In *Human-Computer Interaction - INTERACT 2017*, Regina Bernhaupt, Girish Dalvi, Anirudha Joshi, Devanuj K. Balkrishan, Jacki O'Neill, and Marco Winckler (Eds.). Springer International Publishing, Cham, 441–459.

[85] Catherine Miller, Hannah Kitcher, Kapila Perera, and Alao Abiola. 2020. People, Power and Technology: The 2020 Digital Attitudes Report. https://doteveryone.org.uk/report/peoplepowertech2020/. (Accessed on 09/06/2021).

[86] Narendra Modi. 2018. Make Artificial Intelligence in India, Make Artificial Intelligence Work for India: PM Modi. https://www.narendramodi.in/prime-minister-narendra-modi-inaugurated-wadhwani-institute-of-artificial-intelligence-at-the-university-of-mumbai--538994. (Accessed on 09/09/2021).

[87] Robert M Morgan and Shelby D Hunt. 1994. The commitment-trust theory of relationship marketing. *Journal of marketing* 58, 3 (1994), 20–38.

[88] Rosanna Nagtegaal. 2021. The impact of using algorithms for managerial decisions on public employees' procedural justice. *Government Information Quarterly* 38, 1 (2021), 101536.

[89] Ronald C Nyhan. 2000. Changing the paradigm: Trust and its role in public sector organizations. *The American Review of Public Administration* 30, 1 (2000), 87–109.

[90] Ministry of Electronics & Information Technology. 2020. INDIAai. https://indiaai.gov.in/. (Accessed on 08/24/2021).

[91] Shigehiro Oishi, Jungwon Hahn, Ulrich Schimmack, Phanikiran Radhakrishan, Vivian Dzokoto, and Stephen Ahadi. 2005. The measurement of values across cultures: A pairwise comparison approach. *Journal of research in Personality* 39, 2 (2005), 299–305.

[92] Chinasa T. Okolo, Srujana Kamath, Nicola Dell, and Aditya Vashistha. 2021. "It Cannot Do All of My Work": Community Health Worker Perceptions of AI-Enabled Mobile Health Applications in Rural India. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* (Yokohama, Japan) *(CHI '21).* Association for Computing Machinery, New York, NY, USA, Article 701, 20 pages. https://doi.org/10.1145/3411764.3445420

[93] Sonja K Ötting and Günter W Maier. 2018. The importance of procedural justice in human–machine interactions: Intelligent systems as new decision agents in organizations. *Computers in Human Behavior* 89 (2018), 27–39.

[94] Google Research PAIR Team. 2021. People + AI Research - Guidebok. https://pair.withgoogle.com/guidebook/. (Accessed on 09/08/2021).

[95] Joyojeet Pal. 2012. The machine to aspire to: The computer in rural south India. https://doi.org/10.5210/fm.v17i2.3733

[96] Joyojeet Pal. 2017. The Technological Self in India: From Tech-Savvy Farmers to a Selfie-Tweeting Prime Minister. In *Proceedings of the Ninth International Conference on Information and Communication Technologies and Development* (Lahore, Pakistan) *(ICTD '17).* Association for Computing Machinery, New York, NY, USA, Article 11, 13 pages. https://doi.org/10.1145/3136560.3136583

[97] Joyojeet Pal, Priyank Chandra, Vaishnav Kameswaran, Aakanksha Parameshwar, Sneha Joshi, and Aditya Johri. 2018. Digital Payment and Its Discontents: Street Shops and the Indian Government's Push for Cashless Transactions. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems.* Association for Computing Machinery, New York, NY, USA, 1–13. https://doi.org/10.1145/3173574.3173803

[98] Andrea Papenmeier, Gwenn Englebienne, and Christin Seifert. 2019. How model accuracy and explanation fidelity influence user trust. *CoRR* abs/1907.12652 (2019), 1–7. arXiv:1907.12652 http://arxiv.org/abs/1907.12652

[99] Sylvain Parasie and Eric Dagiral. 2013. Data-driven journalism and the public good:"Computer-assisted-reporters" and "programmer-journalists" in Chicago. *New media & society* 15, 6 (2013), 853–871.

[100] Munoz Claire Parry and Urvashi Aneja. 2020. 3.AI in Healthcare in India: Applications, Challenges and Risks | Chatham House – International Affairs Think Tank. https://www.chathamhouse.org/2020/07/artificial-intelligence-healthcare-insights-india-0/3-ai-healthcare-india-applications. (Accessed on 12/24/2021).

[101] Amandalynne Paullada, Inioluwa Deborah Raji, Emily M Bender, Emily Denton, and Alex Hanna. 2021. Data and its (dis) contents: A survey of dataset development and use in machine learning research. *Patterns* 2, 11 (2021), 100336.

[102] Emilee Rader and Rebecca Gray. 2015. Understanding User Beliefs About Algorithmic Curation in the Facebook News Feed. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems* (Seoul, Republic of Korea) *(CHI '15)*. Association for Computing Machinery, New York, NY, USA, 173–182. https://doi.org/10.1145/2702123.2702174

[103] Inioluwa Deborah Raji, Andrew Smart, Rebecca N. White, Margaret Mitchell, Timnit Gebru, Ben Hutchinson, Jamila Smith-Loud, Daniel Theron, and Parker Barnes. 2020. Closing the AI Accountability Gap: Defining an End-to-End Framework for Internal Algorithmic Auditing. In *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency* (Barcelona, Spain) *(FAT* '20)*. Association for Computing Machinery, New York, NY, USA, 33–44. https://doi.org/10.1145/3351095.3372873

[104] John K Rempel, John G Holmes, and Mark P Zanna. 1985. Trust in close relationships. *Journal of personality and social psychology* 49, 1 (1985), 95.

[105] GlobalData Thematic Research. 2021. Landmark AI legislation could tackle algorithmic bias - Verdict. https://www.verdict.co.uk/ethical-ai-regulation-eu/. (Accessed on 09/09/2021).

[106] Supriya Roy. 2021. Govt pushes for AI curriculum across schools, universities. https://www.techcircle.in/2021/03/16/govt-pushes-for-ai-curriculum-across-schools-universities. (Accessed on 08/22/2021).

[107] Nithya Sambasivan, Erin Arnesen, Ben Hutchinson, Tulsee Doshi, and Vinodkumar Prabhakaran. 2021. Re-Imagining Algorithmic Fairness in India and Beyond. In *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency* (Virtual Event, Canada) *(FAccT '21)*. Association for Computing Machinery, New York, NY, USA, 315–328. https://doi.org/10.1145/3442188.3445896

[108] Nithya Sambasivan and Jess Holbrook. 2018. Toward responsible AI for the next billion users. *Interactions* 26, 1 (2018), 68–71.

[109] Nithya Sambasivan, Shivani Kapania, Hannah Highfill, Diana Akrong, Praveen Paritosh, and Lora M Aroyo. 2021. "Everyone Wants to Do the Model Work, Not the Data Work": Data Cascades in High-Stakes AI. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* (Yokohama, Japan) *(CHI '21)*. Association for Computing Machinery, New York, NY, USA, Article 39, 15 pages. https://doi.org/10.1145/3411764.3445518

[110] Christian Sandvig, Kevin Hamilton, Karrie Karahalios, and Cedric Langbort. 2014. Auditing algorithms: Research methods for detecting discrimination on internet platforms. *Data and discrimination: converting critical concerns into productive inquiry* 22 (2014), 4349–4357.

[111] Edward Santow. 2020. Emerging from AI utopia.

[112] Philipp Schmidt, Felix Biessmann, and Timm Teubner. 2020. Transparency and trust in artificial intelligence systems. *Journal of Decision Systems* 29, 4 (2020), 260–278.

[113] Nick Seaver. 2017. Algorithms as culture: Some tactics for the ethnography of algorithmic systems. *Big data & society* 4, 2 (2017), 2053951717738104.

[114] Yogima Seth Sharma. 2020. NITI Aayog wants dedicated oversight body for use of artificial intelligence. https://economictimes.indiatimes.com/news/economy/policy/niti-aayog-wants-dedicated-oversight-body-for-use-of-artificial-intelligence/articleshow/79260810.cms?from=mdr. (Accessed on 08/22/2021).

[115] Hong Shen, Alicia DeVos, Motahhare Eslami, and Kenneth Holstein. 2021. Everyday Algorithm Auditing: Understanding the Power of Everyday Users in Surfacing Harmful Algorithmic Behaviors. *Proc. ACM Hum.-Comput. Interact.* 5, CSCW2, Article 433 (oct 2021), 29 pages. https://doi.org/10.1145/3479577

[116] Anubhutie Singh and Srikara Prasad. 2020. Dvara Research Blog | Artificial Intelligence in Digital Credit in India. https://dvara.com/blog/2020/04/13/artificial-intelligence-in-digital-credit-in-india/. (Accessed on 09/06/2021).

[117] Aaron Smith. 2018. Public Attitudes Toward Computer Algorithms | Pew Research Center. https://www.pewresearch.org/internet/2018/11/16/public-attitudes-toward-computer-algorithms/. (Accessed on 08/22/2021).

[118] Alison Smith-Renner, Ron Fan, Melissa Birchfield, Tongshuang Wu, Jordan Boyd-Graber, Daniel S. Weld, and Leah Findlater. 2020. No Explainability without Accountability: An Empirical Study of Explanations and Feedback in Interactive ML. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. Association for Computing Machinery, New York, NY, USA, 1–13. https://doi.org/10.1145/3313831.3376624

[119] Cynthias Spiess. 2021. Is That Traffic Light Tracking You? A Case Study on a Municipal Surveillance Technology in Seattle. *IEEE Transactions on Technology and Society* 2, 1 (2021), 15–19.

[120] Janaki Srinivasan and Aditya Johri. 2013. Creating Machine Readable Men: Legitimizing the 'Aadhaar' Mega e-Infrastructure Project in India. In *Proceedings of the Sixth International Conference on Information and Communication Technologies and Development: Full Papers - Volume 1* (Cape Town, South Africa) *(ICTD '13)*. Association for Computing Machinery, New York, NY, USA, 101–112. https://doi.org/10.1145/2516604.2516625

[121] Gerard J Tellis and Deepa Chandrasekaran. 2010. Extent and impact of response biases in cross-national survey research. *International Journal of Research in Marketing* 27, 4 (2010), 329–341.

[122] Divy Thakkar, Neha Kumar, and Nithya Sambasivan. 2020. Towards an AI-Powered Future That Works for Vocational Workers. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. Association for Computing Machinery, New York, NY, USA, 1–13. https://doi.org/10.1145/3313831.3376674

[123] David R Thomas. 2006. A general inductive approach for analyzing qualitative evaluation data. *American journal of evaluation* 27, 2 (2006), 237–246.

[124] Ehsan Toreini, Mhairi Aitken, Kovila Coopamootoo, Karen Elliott, Carlos Gonzalez Zelaya, and Aad van Moorsel. 2020. The Relationship between Trust in AI and Trustworthy Machine Learning Technologies. In *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency* (Barcelona, Spain) *(FAT* '20)*. Association for Computing Machinery, New York, NY, USA, 272–283. https://doi.org/10.1145/3351095.3372834

[125] Joe Toscano. 2018. Optimizing for Trust: The New Metric of Success for Today's Economy | by Joe Toscano RE: Write | Medium. https://medium.com/re-write/optimizing-for-trust-the-new-metric-of-success-for-todays-economy-92408e832a53. (Accessed on 09/09/2021).

[126] European Union. 2019. General Data Protection Regulation (GDPR) Compliance Guidelines. https://gdpr.eu/. (Accessed on 09/10/2021).

[127] Norman Uphoff. 1989. Distinguishing power, authority & legitimacy: Taking Max Weber at his word by using resources-exchange analysis. *Polity* 22, 2 (1989), 295–322.

[128] C Vijai and Worakamol Wisetsri. 2021. Rise of Artificial Intelligence in Healthcare Startups in India. *Advances In Management* 14, 1 (2021), 48–52.

[129] Voice+Code. 2019. Trust is a Critical Customer Experience Metric: Determine How Perceptions of Data Privacy and Security Affect Your Business. https://www.voiceandcode.com/our-insights/2019/3/26/trust-is-a-critical-customer-experience-metric-determine-how-perceptions-of-data-privacy-and-security-affect-your-business. (Accessed on 09/09/2021).

[130] Ruotong Wang, F. Maxwell Harper, and Haiyi Zhu. 2020. Factors Influencing Perceived Fairness in Algorithmic Decision-Making: Algorithm Outcomes, Development Procedures, and Individual Differences. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '20)*. Association for Computing Machinery, New York, NY, USA, 1–14. https://doi.org/10.1145/3313831.3376813

[131] Similar Web. 2021. Top Apps Ranking - Most Popular Apps in India | Similarweb. https://www.similarweb.com/apps/top/google/app-index/in/all/top-free/. (Accessed on 09/10/2021).

[132] Max Weber. 1978. *Economy and society: An outline of interpretive sociology.* Vol. 1. Univ of California Press, Berkeley, CA.

[133] Conor Wilcock. 2020. Cultural Biases in Market Research - B2B International. https://www.b2binternational.com/publications/understanding-accounting-cultural-bias-global-b2b-research/. (Accessed on 09/07/2021).

[134] Marisol Wong-Villacres, Carl DiSalvo, Neha Kumar, and Betsy DiSalvo. 2020. Culture in Action: Unpacking Capacities to Inform Assets-Based Design. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. Association for Computing Machinery, New York, NY, USA, 1–14. https://doi.org/10.1145/3313831.3376329

[135] Allison Woodruff, Yasmin Asare Anderson, Katherine Jameson Armstrong, Marina Gkiza, Jay Jennings, Christopher Moessner, Fernanda B. Viégas, Martin Wattenberg, Lynette Webb, Fabian Wrede, and Patrick Gage Kelley. 2020. "A cold, technical decision-maker": Can AI provide explainability, negotiability, and humanity? *CoRR* abs/2012.00874 (2020), 1–23. arXiv:2012.00874 (https://arxiv.org/abs/2012.00874)

[136] Allison Woodruff, Sarah E. Fox, Steven Rousso-Schindler, and Jeffrey Warshaw. 2018. A Qualitative Exploration of Perceptions of Algorithmic Fairness. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. Association for Computing Machinery, New York, NY, USA, 1–14. https://doi.org/10.1145/3173574.3174230

[137] S. Woods, M. Walters, Kheng Lee Koay, and K. Dautenhahn. 2006. Comparing human robot interaction scenarios using live and video based methods: towards a novel methodological approach. In *9th IEEE International Workshop on Advanced Motion Control, 2006*. IEEE, Istanbul, Turkey, 750–755. https://doi.org/10.1109/AMC.2006.1631754

[138] Fumeng Yang, Zhuanyi Huang, Jean Scholtz, and Dustin L. Arendt. 2020. How Do Visual Explanations Foster End Users' Appropriate Trust in Machine Learning?. In *Proceedings of the 25th International Conference on Intelligent User Interfaces* (Cagliari, Italy) *(IUI '20)*. Association for Computing Machinery, New York, NY, USA, 189–201. https://doi.org/10.1145/3377325.3377480

A INTERVIEW SCENARIOS

Loan assessment (Scenario 1). Suppose you (or your family members) are applying for a loan for a sum of Rs. 10 lakhs. The bank is using an AI system called MyLoans for evaluating loan applications. You will be required to fill out the loan application form, specifying the loan amount, applicant financial history, and submit them along with relevant documents like address or ID proof to MyLoans. Once assessed, the status of your loan application will be updated on the MyLoans application. The difference from a regular loan assessment process is that MyLoans (an AI system) will determine whether your loan application is approved/successful or not, instead of a loan officer.

Financial Advisor (Scenario 2). Consider the situation where you are planning to start saving for a financial goal such as education, vehicle, marriage, for yourself or a family member. There is an application, MoneyAdvisor, which using AI, helps its users to understand their financial behavior. It takes into account the user's age, marital status, employment status, annual income, their assets and expenses. The application creates customized advice for their users, recommending ways to budget income and expenditure, and offers suggestions for investments. The difference from a regular financial advising process is that MoneyAdvisor (an AI system) will determine generate recommendations on how to budget and where to invest, instead of a human financial advisor.

Medical Diagnosis (Scenario 3). Consider the situation where you're experiencing symptoms like fatigue, shortness of breath, and chest pain. The hospital uses a system– GetWell– which using Artificial Intelligence is capable of reading your ECG reports, detect cardiac abnormalities, provides a diagnosis, and responds with the next steps which best match your diagnosis. You would have to enter your age, gender and upload the ECG report on GetWell at the remote centre. The difference from a regular medical diagnosis process is that GetWell (an AI system) will determine your medical diagnosis, instead of a doctor.

Hospital Finder (Scenario 4). Consider the situation where you are experiencing symptoms of nausea, fever, and stomach ache. But you are not sure which hospital to visit. There is an application called FindADoctor which using Artificial Intelligence, helps its users in finding nearby hospitals, nursing homes and clinics. The user needs to answer a few questions about their location and their symptoms.

B SURVEY RESULTS

AI Acceptability	Not at all	Somewhat	Moderately	Very	Extremely
General	3.5%	17.2%	22.0%	30.1%	27.2%
Medical Diagnosis	7.8%	19.4%	22.7%	28.3%	21.8%
Loan Approval	7.2%	16.6%	22.4%	30.5%	23.3%
Hiring Decision	9.6%	18.1%	22.4%	26.6%	23.3%
Song Recommendation	2.2%	12.9%	18.3%	34.9%	31.8%
Route Recommendation	1.3%	11.3%	18.3%	36.8%	32.2%
Auto Pricing Decision	3.9%	15.9%	20.9%	31.4%	27.9%

Table 4: Detailed results from the survey on the acceptance of AI-based outcomes in general, and across the six scenarios listed above.

	General	High-stakes	Low-stakes
	Independent samples t-test		
Gender	0.37 (0.10), $p < 0.001$	0.27 (0.10), $p < 0.01$	0.22 (0.08), $p < 0.01$
	One-way ANOVA		
Age	$F(5, 453) = 7.19$, $p < 0.001$	$F(5, 453) = 11.55$, $p < 0.001$	$F(5, 453) = 6.71$, $p < 0.001$
Internet exposure	$F(4, 454) = 0.77$, $p > 0.05$	$F(4, 454) = 1.21$, $p > 0.05$	$F(4, 454) = 1.43$, $p > 0.05$

Table 5: Results from our survey: Influence of personal characteristics on acceptability