STUDENT **PAPER** 1

Building Trust in the AI Ecosystem by Re-Evaluating Public Perception

Researcher: Christian Flores | Mentor: Dr. Sean Kross

ABSTRACT

Artificial intelligence systems leverage large datasets with iterative processing algorithms that identify patterns to create an additional layer of expertise. This transformational power operates in tandem with ethical risks. The dominant narrative behind AI is simultaneously stigmatized and misunderstood: with exponential growth of the ubiquitous technology leaving public awareness in the dust, it's becoming increasingly important to balance enthusiasm for AI's enormous promise with a sober understanding of its moral risks. This study seeks to characterize the public opinion of AI in high-risk, domain-specific applications. To that end, a poll was administered to American adults. The results of the study reveal that the great majority of survey respondents have a neutral or optimistic perspective on AI in particular high-risk domains. The study concludes by presenting a standard heuristic for understanding public perception where ethics may fail to preserve a human factors' approach. In this way, researchers and developers can undertake coordinated efforts to mitigate the harm caused by AI while promoting rational optimism in vulnerable populations.

INTRODUCTION

AI systems leverage machine learning algorithms to maximize the potential of big data. However, deploying machine-learning algorithms at scale comes with risk. This, in part, can be explained by the inability for artificial intelligence to replicate human capabilities including individual flexibility, context-relevant judgements, empathy, as well as complex moral judgements (Webb et al., 2021). Now more than ever, the dilemma of socially unaware AI is synonymous with scandal (Dressel & Farid, 2018) (Bartlett et al., 2022). With eager news sources and online publications acting as doomsayers, the narrative surrounding AI has shifted from harnessing life-changing potential to impending doom. A story of promise and peril. The pressure is mounting to design ethical AI: a report made by researchers at Cambridge and Oxford enumerates a number of priority research areas for AI development, one of them highlighting the importance behind balancing optimism about the vast potential of AI technology with a level-headed recognition of the risks involved (Brundage et al., 2018). Prior research has examined how the general public perceives the impact of the technology (Müller & Bostrom, 2016) (Grace et al., 2018). However, the research lacks a standard method for understanding public perception and sentiment behind AI. Furthermore, prior research has not concentrated on AI application in specific domains. Instead, the vast majority have opted to monitor perception and sentiment on the topic of AI in a broad sense, using select media news source coverage and broadly encompassing survey topics.

In the study presented in this paper, 280 adults aged 18 or older in the United States completed an online survey. To promote generalizability, the survey sample was demographically balanced to include a wide range of respondents with different gender, age, and level of education. The research inherently promotes a risk-benefit evaluation strategy to evaluate the data critically. We investigate one essential research question:

RQ) What are the attitudes and sentiments towards AI-application in high-risk domains?

The results from the poll indicate that the great majority of respondents had a neutral or optimistic perspective on AI in particular high-risk domains. In addition, by analyzing the attitude of survey respondents, we may conduct a more thorough analysis of perception. While there is a greater range of keywords used to describe "negative" attitudes, "positive" words were more prominent in the data, as determined by sentiment analysis. The findings of this study shed light on how the general public perceives high-risk domains in the AI ecosystem. The results also give developers and lawmakers who regulate AI with a compass for future development and governance.

In general, survey respondents accept both the positive and negative aspects of domain-specific AI applications, as opposed to accepting only one side. However, further investigation reveals that emotional evaluation questions across all domains tend to have a positive sentiment. This outlook reveals wide-ranging insights with an overall optimistic tone on the future of AI.

A risk-benefit analysis is a robust tool for determining public acceptance and support for a particular technology. The experimental heuristic described in the Discussion section could assist researchers in detecting whether AI could aid researchers in determining if AI applications are losing public favor.

Related Literature

Recent research tries to elucidate the underlying effects of algorithm-driven AI across many aspects of modern life. A growth in large survey studies attempts to address the disconnect between the disruptive technology and the general population. Despite a gradual shift from "robot apocalypse" and "automation boon" tropes to a more open-minded approach in discourse surrounding the subject, the societal effects of AI continue to draw intense public attention (Littman et al., 2021).

Analysis on Mass-Media Discourse

In the past, text mining techniques have been used to determine how the media feels about AI. Fast and Horvitz conducted a long-term study of how the public felt about articles published by the New York Times between January 1986 and May 2016, which added up to a staggering 3 million pieces. The purpose of the study was to determine how people's hopes and fears regarding AI have evolved over time. Indicators for a set of hopes and fears about AI are collected in the study. To accomplish this, they examined how ideas have changed over time by searching article text for positive or negative keywords. This enabled them to sort article topics into two primary groups: those that express a hope or a fear. The research article concludes that existential fear and worry about a number of AI applications are on the rise. In addition, there has been an increase in ethical concerns regarding AI over the past three decades, which have been driven in part by existential concerns. The essay concludes with a summary report on the optimistic or pessimistic perceptions of AI-related publications. Surprisingly, the study reveals a cheerful outlook for the future of AI. Since 2009, there has been a sharp increase in the number of public discussions about AI, with the majority of media sources expressing optimism (Fast & Horvitz, 2017).

Other studies of mass media discourse have tracked the evolution of AI as covered by major news media outlets. Zhai et al. analyze five major news media outlets over the past three decades using seven dimensions: scientific subject, keyword, country, institution, people, topic and opinion polarity. While the study acknowledges that AI has become an important force in the new era, the public's perception of AI has been contested. According to a widely held belief in the media, AI is the driving force behind the modernization of conventional industries. However, the notion of AI as a "humanized" technology is not yet widely accepted. This can be summed up by one single insight: when we transfer the right of judgment to computer systems, there will be a number of moral and ethical dilemmas involved (Zhai et al., 2020).

Chuan et al. explores framing theory in journalism and science communication on the topic of artificial intelligence. The objectives of the study were to evaluate how the topic of AI was presented in major American newspapers during a ten-year period as well as the themes that were covered more frequently. After a thorough examination of the articles reviewed for the study, it became evident that ethical or moral issues were the most prevalent topics. The study concluded that a more in-depth discussion of the risks and benefits of AI is required for a critical evaluation of the technology's use and regulation (Chuan et al., 2019) .

Survey outreach

Other literature has used a methodological approach that is more individualized to gather information on how the general population perceives AI. Yeh et al. examine the perceived understanding and involvement with AI among Taiwanese survey respondents. Forty-three percent of the 1108 respondents identified themselves as "slightly" understanding AI, according to the study. When asked about their specific involvements with various AI-enabled devices and applications, however, over fifty-seven percent of respondents reported a moderate to high level of familiarity with the technology (Yeh et al., 2021). This accentuates public oblivion in nations abroad. The study presented in this paper adds to the discussion by characterizing public perception and sentiment on AI with a primary goal of providing a foundation for conducting similar empirical research in the future.

Methods

A survey containing likert assessments and affective questions was used to collect data regarding the overall impact of AI as well as its impact in three specific high-risk domains. A high-risk domain is an application area where the ethics of the technology's use are questioned due to unintended consequences. The job equality domain refers to the fairness in obtaining and keeping an easily automatable job. The user behavior domain describes the use of technology to influence user behavior, typically for monetary gain. The decision-making domain refers to the method of synthesizing large amounts of data to automate the process of making significant decisions. When investigating public opinion and attitude regarding AI, these three domains are of primary interest.

Data collection

The survey firm Survey Monkey was contracted to administer the questionnaire to American adults in the United States. Respondents were able to fill out the survey hosted on the platform using a computer or mobile device. There were over 280 valid survey responses to our survey.

Survey Questionnaire

Respondents were asked about the influence of AI on humanity in general as well as on three distinct domains: job equality, user behavior, and decision-making. The global impact of AI was assessed by a scale that asked: "How would you assess the overall impact of AI on humankind?". Similar wording was used to assess the perceived impact of the three domain-specific AI applications. Respondents answered using a five-point likert scale, ranging from 1 (extremely negative) to 5 (extremely positive). For each

domain, respondents were also asked to rate their level of agreement on a five-point likert scale, ranging from 1 (strongly disagree) to 5 (strongly agree), on four different statements that were equally distributed as either a benefit or concern of the application of AI in that domain. Statements were formulated to emphasize either a positive or negative outcome resulting from the implementation of AI in the specified domain.

Following each domain-specific general likert assessment and matrix-style question, survey respondents were also asked to answer an affective question about the impact that AI has on each domain-specific application, expressing how they feel about it. In particular, respondents were asked, "How do you feel about the impact of AI on employment opportunities for the general population?". Modifications were made to the query's language to accommodate the relevant domain context.

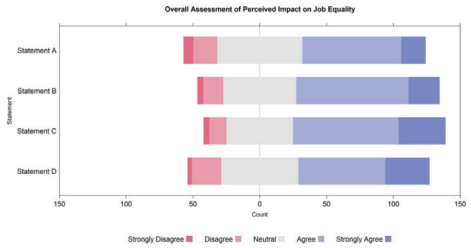
Results

Perceived impact of AI on humankind and within three domains.

The data collected reveals a hopeful tone for the receptivity of AI. The majority of respondents had a neutral or favorable opinion of the technology based on their responses to the likert scales that assessed the overall impact that AI was perceived to have on humanity and the three specific domains. Many respondents rated the overall impact of AI on humanity and across the three domains as positive or extremely positive (38.9%). Forty and three hundredths of a percent deemed the overall impact of AI on humanity and the three domains to be neutral (40.3%). Twenty and eight hundredths percent of respondents rated the overall impact of AI on humanity and across the three domains as negative or extremely negative (20.8%).

FIGURE 1:

Proportions of perceived impact that AI has on humankind generally and across three specific domains.

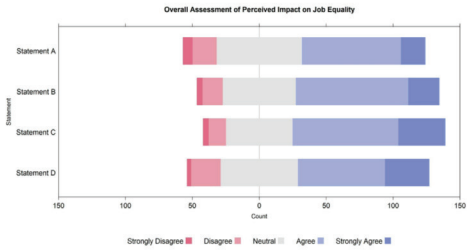


Risk-benefit assessment in the job equality domain.

The level of respondents' agreement with four statements formulated and distributed equally as either a benefit or a concern in each domain was analyzed using matrix-style questions. Over half of respondents agreed or strongly agreed with the benefit statements for the job equality domain (55%). Many responses were neutral regarding the benefit statements (32.8%), while a small percentage disagreed or strongly disagreed with them (12.2%). Similar tendencies can be observed in the level of agreement with the concern statements. More than half of respondents (58.4%) expressed agreement or strong agreement with the concern statements. Plenty indicated a neutral stance on the concern statements (29.8%), while a minority disagreed or strongly disagreed (11.7%).

FIGURE 2:

Proportions of perceived impact of AI on job equality through assessment of benefits and concerns (Statements A & B are related to benefits. Statements C & D are related to concerns).

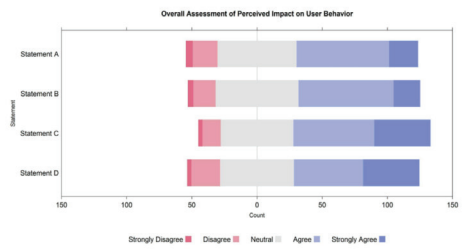


Risk-benefit assessment in the user behavior domain.

In the user behavior domain, slightly over half of the respondents agreed or strongly agreed with the benefits statements (52.2%). Many responses indicated a neutral stance on the benefits statements (35.1%), while a minority indicated disagreement or strong disagreement (12.7%). More than half of the respondents either agreed or strongly agreed with the concern statements. A sizable proportion of respondents held a neutral stance on the concern statements (31.7%), while a minority disagreed or strongly disagreed with the concern statements (11.8%).

FIGURE 3:

Proportions of perceived impact of AI on user behavior through assessment of benefits and concerns (Statements A & B are related to benefits. Statements C & D are related to concerns).

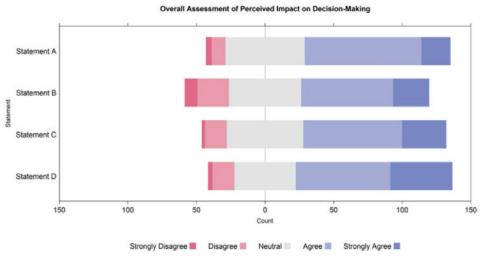


Risk-benefit assessment in decision-making domain.

For the domain of decision-making, the same trends observed in the ranking of benefit and concern statements in previous domains are observed. Over half of the responses indicated they agreed or strongly agreed with benefit statements (55.9%). A sizable proportion of respondents expressed a neutral stance on the benefit statements (31.2%), while a minority of responses indicated they disagreed or strongly disagreed with the benefit statements (12.9%). The overwhelming majority of respondents agreed or strongly agreed with the concern statements (61.2%). Many responses held a neutral stance on the concern statements (28.4%), while a minority disagreed or strongly disagreed (10.4%).

FIGURE 4:

Proportions of perceived impact of AI on decision-making through assessment of benefits and concerns (Statements A & B are related to benefits. Statements C & D are related to concerns).



Text mining for sentiment analysis on perceived impact of AI for all domains.

For a more rigorous analysis of the two types of likert assessments across all domains, affective question response data was collected for each domain. After reviewing each matrix-style evaluation of benefits and concerns, survey respondents were able to express their feelings regarding AI's impact on a specific domain. Using techniques for sentiment analysis in the programming language R, the results indicate that the majority of respondents viewed the impact of AI positively across all domains. Sixty percent of text responses contributed to a positive sentiment.

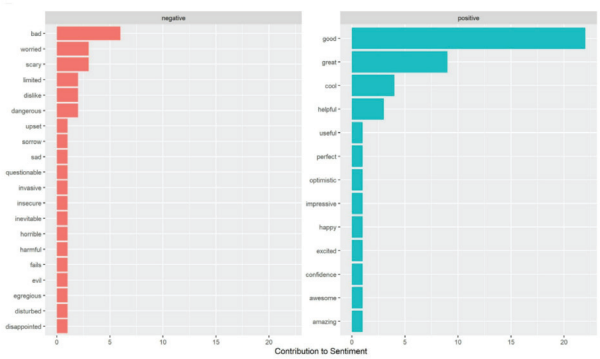
FIGURE 4:

The relative importance of specific sentiment keywords as represented by font size.

While there was a wider range of keywords used to describe "negative" sentiments, "positive" words were more abundant from the data.

FIGURE 5:

The distribution of negative (red) and positive (blue) keywords used to discover overall sentiment trends.



Discussion

While the majority of respondents expressed a neutral, positive, or extremely positive assessment of perceived impact that AI has on humanity and within the three domains, the matrix-assessments for each domain reveal a conflicting representation on the perception of AI. The majority of respondents concur with both the benefit and concern statements for each domain. Respondents accept both the positive and negative aspects of domain-specific AI applications, rather than agreeing with one side or the other. However, further investigation reveals that emotional evaluation questions across all domains tend to have a positive sentiment. Although it is impossible to draw a definitive conclusion from this study, the data is consistent with previous research in that it indicates a general decline in existentialist beliefs.

In addition, we develop a heuristic for evaluating public opinion on related topics and suggest future research areas based on the findings of this study.

A multidimensional and relational perspective with a keen focus on context.

According to the findings, the general public perceives that AI has a good influence on humanity and within three distinct categories, however most respondents equally accept the advantages and risks of each application's unique domain. This inference shows a favorable attitude toward the technology but also points to skepticism about certain applications of AI.

Parsing the matrix assessment data reveals potential justifications for the polarity. By choosing to focus on respondent answer trends for benefits and concerns across all domain-specific applications, new dimensions behind the problem are explored. Contextualizing benefit and concern assessments is critical to understanding why survey results indicate an equal level of agreement with risks and benefits associated with domain-specific AI applications. We hypothesize that certain domain-specific applications will appeal more to specific individuals. This can be driven by the rationale that respondent exposure to AI technology is context-dependent. Given that persons of any age 18 and over were eligible to participate in the poll, it is important to take into account the potential disparities in technology use among young and older respondents. While no attempt was made to decode the technological generation gap in this

study, a number of studies have linked the current generation to increased technology use (Vogels, 2019). A younger respondent may have a greater likelihood of being exposed to AI-enabled technology than an older one who has no need for it. Furthermore, a subset of individuals may opt to consume sensationalized media coverage of AI, which may exacerbate the opinion polarity.

Conducting correlational research to determine significant attitude predictors.

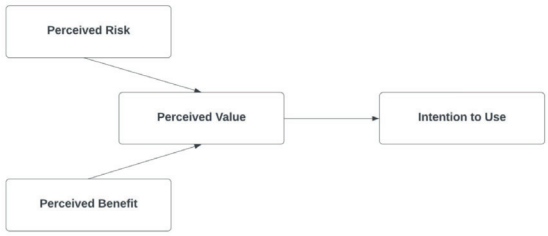
The findings of this study could be improved by evaluating background factors. Taking into account variables such as domain-specific experience, technological exposure, and subscription to major news channels, it may be possible to gain a broader perspective on some domain-specific applications of AI. Reframing the methodological approach by questioning about the respondent's history can enhance our understanding of public receptivity to AI. In essence, it would provide a more holistic perspective that would assist guide developers and legislators more accurately.

Establishing a heuristic for future discussion.

The findings contribute to the discussion and provide useful information that might influence future research. More importantly, the methodology will give a solid baseline for undertaking comparable empirical research. A risk-benefit analysis is a valuable instrument for gauging public acceptance and support for a certain technology. The experimental heuristic outlined below could aid researchers in determining whether AI applications are losing public acceptance.

FIGURE 3:

A flowchart representing a risk-benefit framework for conducting perception inquiry.



When doing comparable perception research, researchers may use this graphic to inform their methodological approach. In an experiment, participants may be asked about their opinions on a certain technology, for instance. In an effort to promote or regulate a technology, it is possible to assess if the benefits outweigh the risks by comparing the risks and benefits associated with the technology. This approach to empirical research proves to be concrete and quantifiable, serving as a consistent indicator for perceived product value.

Conclusion

In this study, we present the results of a survey of 280 individuals regarding how they see the influence of artificial intelligence in both general and domain-specific applications. While the majority of respondents in the survey have a favorable view toward AI, there is a similar degree of consensus about the advantages and disadvantages of domain-specific applications. Consistent with past findings, the results of the sentiment analysis help to underscore the positive acceptance of artificial intelligence. To comprehend the perceptual landscape of AI technology, further contextual study is required.

Understanding how the public perceives artificial intelligence is crucial for product creation, research, and public policy. Therefore, a feedback loop is essential in field research and development. To control the development of artificial intelligence effectively, the general public must be a stakeholder in the technology, and academia must be prepared to bridge the gap between public opinion and policy. The purpose of the present study is to characterize how the American public now views AI. Importantly, the study closes with a paradigm that, when applied to available research on the topic, can potentially estimate relative product value. The purpose of the study is to improve the design of future research and provide information on artificial intelligence deployment areas where public support is waning. When these results are considered collectively, relevant stakeholders may operationalize governance in AI-enabled applications lacking a robust human factors approach.

CITATIONS

[1] Webb, M.E., Fluck, A., Magenheim, J., et al. (2021). Machine Learning for Human Learners: Opportunities, Issues, Tensions and Threats. Education Tech Research Dev, 69, 2109–2130. https://doi.org/10.1007/s11423-020-09858-2

[2] Dressel, J., & Farid, H. (2018). The Accuracy, Fairness, and Limits of Predicting Recidivism. Science Advances, 4(1). https://doi.org/10.1126/sciadv.aao5580

[3] Bartlett, R., Morse, A., Stanton, R., Wallace, N. (2022). Consumer-Lending Discrimination in the FinTech Era. Journal of Financial Economics, 143 (1), 30 - 56. https://doi.org/10.1016/j.jfineco.2021.05.047

[4] Brundage, M., Avin, S., Clark, J., Toner, H., Eckersley, P., Garfinkel, B., Dafoe, A., et al. (2018). The Malicious Use of Artificial Intelligence: Forecasting, Prevention, and Mitigation. https://doi.org/10.17863/CAM.22520.

[5] Müller, V., & Bostrom, N. (2016). Future Progress in Artificial Intelligence: A Survey of Expert Opinion. Fundamental Issues of Artificial Intelligence, 553-571. https://doi.org/10.1007/978-3-319-26485-1_33

[6] Grace, K., Salvatier, J., Dafoe, A., Zhang, B., & Evans, O. (2018). When will AI Exceed Human Performance? Evidence from AI Experts. Journal of Artificial Intelligence Research, 62, 729-754. https://doi.org/10.1613/jair.1.11222.

[7] Littman, M., Ajunwa, I., Berger, Guy., Boutilier, C., Currie, M., Doshi-Velez, F., Hadfield, Gillian., et al. (2021) Gathering Strengths, Gathering Storms: One Hundred Year Study on Artificial Intelligence Panel Report. AI100. http://ai100.stanford.edu/2021-report.

[8] Fast, E., & Horvitz, E. (2017). Long-Term Trends in the Public Perception of Artificial Intelligence. AAAI. https://doi.org/10.48550/arXiv.1609.04904.

[9] Zhai, Y., Yan, J., Zhang, H. and Lu, W. (2020). "Tracing the Evolution of AI: Conceptualization of Artificial Intelligence in Mass Media Discourse". Information Discovery and Delivery, 48 (3), 137-149. https://doi.org/10.1108/IDD-01-2020-0007.

[10] Chuan, Ching-Hua., Sunny Tsai, Wan Hsiu. and Cho, Su Yeon. (2019). Framing Artificial Intelligence in American Newspapers. AIES. https://doi.org/10.1145/3306618.3314285.

[11] Yeh S-C, Wu A-W, Yu H-C, Wu HC, Kuo Y-P, Chen P-X. (2021). Public Perception of Artificial Intelligence and Its Connections to the Sustainable Development Goals. Sustainability, 13(16), 9165. https://doi.org/10.3390/su13169165.

[12] Vogels, E. (2019). Millennials Stand Out for their Technology Use, but Older Generations Also Embrace Digital Life. Pew Research Center.Washington, D.C. https://www.pewresearch.org/fact-tank/2019/09/09/us-generations-technology-use/ Flores 19

Christian Flores

McNair Cohort: 2022

Biography:

My name is Christian Flores. I am a recent 2022 graduate with a bachelor's degree in Cognitive Science. During my time at UCSD, I entertained the practical and theoretical aspects of my major. I recently completed a self-guided research project that explored the social impact of artificial intelligence. My research interests generally lie at the intersection of computing and social impact. I will pursue a masters degree in Computer Science beginning in the Fall of 2023. My plan is to obtain a PhD in Computer Science to pursue a career in academia.

> *"I was motivated to engage in research by a curiosity to explore beyond the proficiency and application I gained from regular coursework, seeking deeper insights into a relevant domain."*

Acknowledgements:

I would like to thank my mentor, Dr. Sean Kross for his invaluable advice in completing my research paper. I would also like to thank the wonderful staff in the Undergraduate Research Hub and within the McNair Scholars Program for providing me with relentless support and guidance throughout the research process.