Selective aversion for Artificial Intelligence: Domain specificity and perceived understanding influences algorithm aversion

Authors: Sumitava Mukherjee [1] , Deeptimayee Senapati [1] , and Isha Mahajan [2]

Introduction

•With AI slowly paving its way into the public sector, ***Niti Aayog***, the “Think Tank” of the Government of India, has launched the #AIforAll[1] initiative to introduce AI in various sectors, including healthcare, education, agriculture, mobility, and smart cities & infrastructure.

•These initiatives' success largely depends on the public's acceptance and trust because people are aversive to use AI algorithms in some domains, even when Algorithms perform better than humans (Algorithm Aversion)[5].

•Lack of transparency[5], domain specificity[11], task sensitivity[2], stakes involved[10], and uniqueness-neglect[7] are some of the factors that contribute to algorithm aversion.

•Providing social proof [1], making it look like it is customized[7], giving humans some control over the algorithm[4], understanding how the algorithms work[11], and increasing algorithmic processing transparency can favor reducing aversion.

•The current study is the first attempt to explore public acceptability in different sectors of governance where we tried to gauge people's acceptability, understanding, and preference for the use of AI.

Google Meet Link : https://meet.google.com/wrm-jgbz-eac

sm1@hss.iitd.ac.in / deeptimayeesenapati@gmail.com

Method

Participants: One Hundred ninety-three (Females=96 ; mean age = 23.61 years, *SD = 8.07)* individuals participated in the online survey voluntarily.

Procedure: We explored people’s judgement towards AI in five key domains proposed by #AIforAll initiative.

• For each domain either two or three task scenarios were mentioned. All tasks had four questions which were each rated on a 10-point scale, these items measured acceptability, perceived sense of understanding, perceived nature of the task and algorithm aversion respectively.

Questions:

(a) How acceptable it would be for them if an AI program performed the particular task (1 = "completely not acceptable" to 10 = "completely acceptable");

(b) The extent of their understanding about how an AI might perform the particular task (1 = "I do not understand at all how an AI program will do this task" to10 ="I completely understand how an AI program will do this task")

(c) Perceived nature of the task (1 = "completely objective" to 10 = "completely subjective")

(d) whether they are aversive towards algorithms to do that task (1= "I prefer this task to be done by a qualified human" to 10 = "I prefer this task to be done by an AI program").

Results

• We found lower acceptability of AI in education (Mean= 5.51, SD= 2.21) and healthcare (Mean=5.87, SD=2.36) but higher acceptability in agriculture (Mean=8.13, SD=1.74), smart mobility (Mean=7.78, SD=1.49) and smart cities & infrastructure (Mean=7.25, SD=1.81).

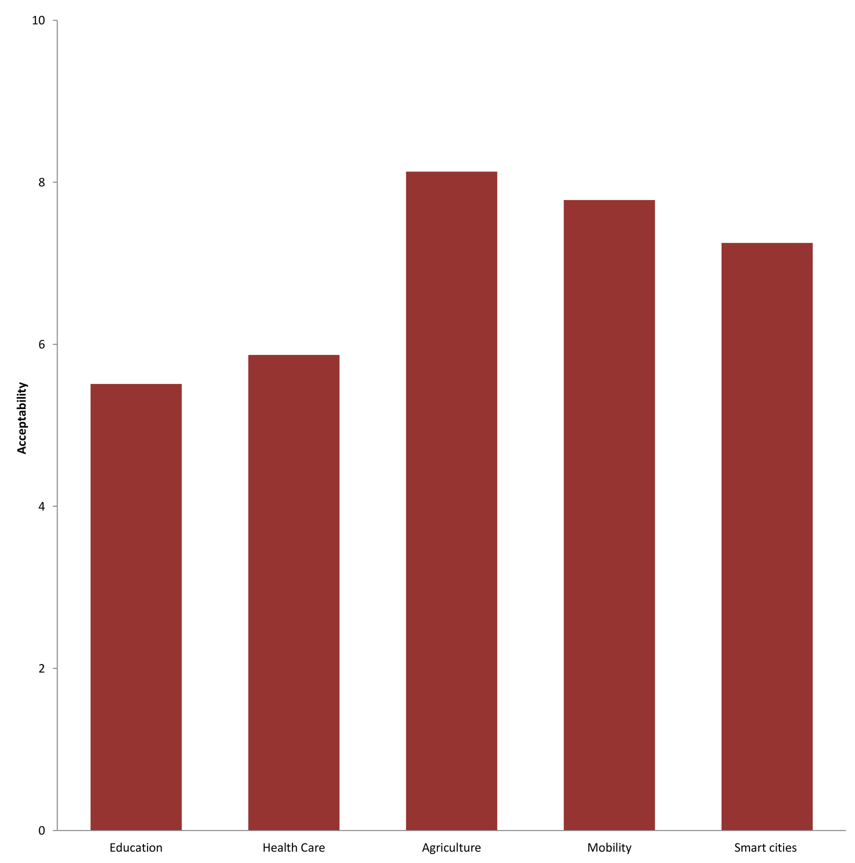

Figure 1: Acceptability across the domains. Acceptability in education and healthcare is lower compared to other domains.

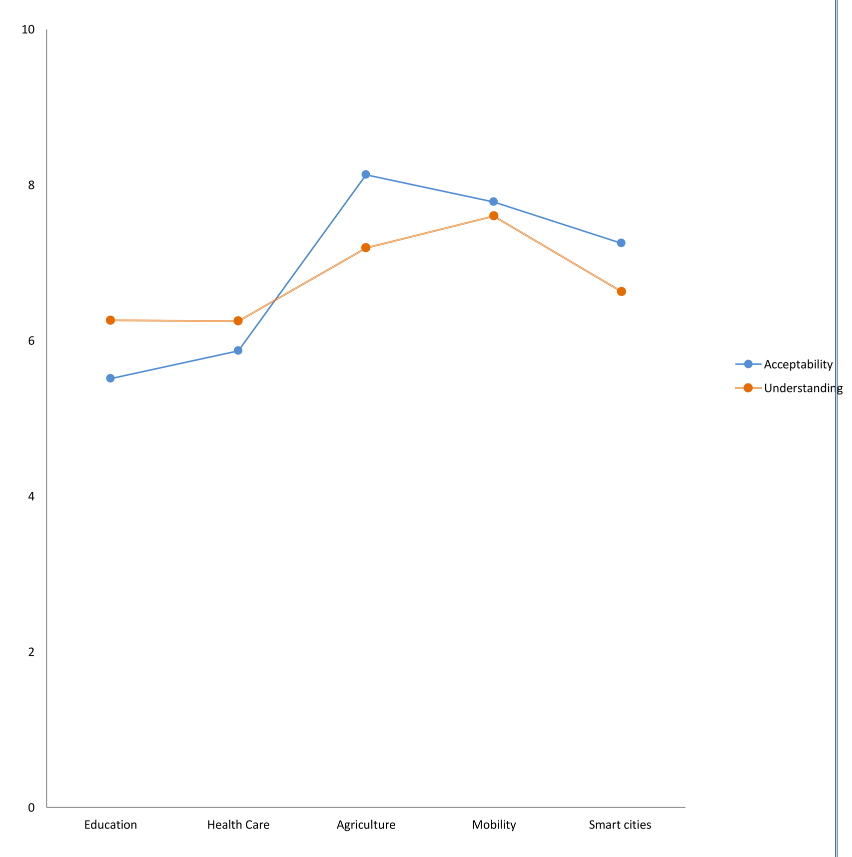

Figure 2: Acceptability and understanding across the domains.

• A simple linear regression revealed that perceived understanding drives acceptability towards AI; (F(1,191) = 75.23, *p < .001); R2 = .283*

• We also observed higher acceptability leads to lower algorithm aversion ; F(1,191) = 272.89, *p < .001 with R2 of .588*

• A stepwise regression model was obtained which is “Acceptability= 2.515+ (.184* Perceived Understanding)+ (.575* Algorithm Aversion)”.

Discussion

•Acceptability is the key to decrease algorithm aversion.

• Domain specificity exists in acceptance of algorithms, such as in some domains, people are willing to use algorithms, and in others, they do not.

•People's perceived understanding of how AI works drives both algorithm acceptability and aversion in various domains.

•The study provides a clear picture of the factors affecting acceptability of AI, thus representing a more precise picture than a range of previous diversified findings, which has important implications for national tech strategies

References

1. Alexander, V., Blinder, C., Zak, P. (2018). Why trust an algorithm?. *Computers in Human Behavior, 89. pp. 279-288. http://doi.org/10.1016/j.chb.2018.07.026*
2. Castelo, N., Maarten W. B., & Lehmann, D. R. (2019). Task-dependent algorithm aversion, *Journal of Marketing Research .*
3. Dawes, R. M., Faust, D., & Meehl, P. E. (1989). Clinical versus actuarial judgment. *Science, 243 , 1668–1674. http://dx.doi.org/10.1126/science.2648573*
4. Dietvorst, B. J., Simmons, J. P., & Massey, C. (2016). Overcoming algorithm aversion: People will use imperfect algorithms if they can (even slightly) modify them. *Management Science.*
5. Dietvorst, Berkeley & Simmons, Joseph & Massey, Cade. (2014). Algorithm aversion: People erroneously avoid algorithms after seeing them err. *Journal of experimental psychology.*
6. Dijkstra, J. J. (1999). User agreement with incorrect expert system advice. *Behaviour and Information Technology, 18(6), 399-411.*
7. Longoni, Chiara &Bonezzi, Andrea &Morewedge, Carey. (2019). Resistance to medical artificial intelligence. *Journal of Consumer Research. 46.* http://doi.org/10.1093/jcr/ucz013 .
8. Meehl, P. E. (1954). Clinical versus statistical prediction: A theoretical analysis and review of the literature. *Minneapolis, MN: University ofMinnesota Press.*
9. Niti Aayog. (2018). *National Strategy for Artificial Intelligence.*
10. Promberger, M., & Baron, J. (2006). Do patients trust computers? *Journal of Behavioral Decision Making, 19, 455–468.*
11. Yeomans, Michael & Shah, Anuj & Mullainathan, Sendhil & Kleinberg, Jon. (2019).Making sense of recommendations. *Journal of Behavioral Decision Making.*

1 Department of Humanities and Social Sciences, Indian Institute of Technology Delhi, New Delhi
2 Symbiosis School for Liberal Arts, Symbiosis International University, Pune