Influence on human decision-making of a biased AI health recommendation system (#96488)

Author(s)
Lucía Vicente (University of Deusto) - lucia.vicente@deusto.es
Helena Matute (University of Deusto) - matute@deusto.es

1) Have any data been collected for this study already?
No, no data have been collected for this study yet.

2) What's the main question being asked or hypothesis being tested in this study?
Our first prediction concerns the first phase of the experiment: We predict that participants assisted by a biased recommendation system during this phase will make more mistakes than an unassisted group of participants. This should replicate the results of a previous experiment. Our second prediction is that the group of volunteers assisted by the AI during the first phase will make more errors than the unassisted group also in the second phase of the experiment when both groups perform the task without assistance. That is, the AI-assisted group will inherit the bias of the recommendation system. Third, we will also test whether the bias in the AI-assisted group generalizes to ambiguous (50/50) stimuli during Phase 2, so that these ambiguous stimuli become classified more often in the direction of the errors made by the AI during Phase 1. The control group should show no bias and should classify these ambiguous stimuli randomly (i.e., 50% positive, 50% negative).

3) Describe the key dependent variable(s) specifying how they will be measured.
During the first phase of the experiment, the main dependent variable is the number of mistakes that participants make in all the trials of the task and particularly in the trials where the AI recommendation is erroneous, that is, where the recommendation contradicts the objective information (i.e. proportion of dark/light cells) presented in a fictitious tissue sample. (The bias of the AI consists on a systematic wrong recommendation in trials where the proportion of dark and light cells in the tissue sample is 40/60. If participants in the assisted group follow the AI recommendation during the task they will made the same mistakes than the system, on the contrary, if participants observe the tissue sample to make the decision, ignoring the recommendation, there will be no differences between the assisted and unassisted group). During the second phase of the experiment, where all participants are unassisted, the main dependent variables are also the number of errors in all trials of the classification task and particularly the number of errors in the classification of the tissue samples with the proportion 40/60. We expect participants from the AI-assisted group to classify these samples in the same way that the recommendation system did in the previous phase. Third, during phase 2 we will also assess the number of errors in ambiguous (50/50) stimuli, in order to test the generalization of the inherited bias (see Prediction 3).

4) How many and which conditions will participants be assigned to?
The experiment comprises two groups. Half of the participants will be randomly assigned to the AI-assisted group, while the other half will be randomly assigned to the unassisted group. In phase 2, both groups will have no assistance during the task.

5) Specify exactly which analyses you will conduct to examine the main question/hypothesis.
Analyses to replicate previous findings in Phase 1: (1) between groups t-test on number of errors in the 40/60 samples in Phase 1 and (2) between groups t-test on total number of errors during all trials of Phase 1.
Analyses in Phase 2: (1) between groups t-test on number of errors in the 40/60 samples in Phase 2, (2) between groups t-test on total number of errors during all trials of Phase 2 and (3) between groups t-test on number of errors in the 50/50 samples of Phase 2.

6) Describe exactly how outliers will be defined and handled, and your precise rule(s) for excluding observations.
If participants do not get at least 5 correct responses (out of 6) on the second repetition of the second block of trials in the practice phase, their data will be excluded from the analysis. Also, if participants do not perform better than chance (more than 30 hits out of 60) during phase 1, we consider that they either did not pay attention to the task, or did not understand the instructions, and their data will be excluded from the analysis.

7) How many observations will be collected or what will determine sample size? No need to justify decision, but be precise about exactly how the number will be determined.
We have planned a sample size of 200 participants (100 per group). This sample size should allow us to detect a size effect of $d = 0.35$ on the Student's t test for the difference between two independent means with 80% power.

8) Anything else you would like to pre-register? (e.g., secondary analyses, variables collected for exploratory purposes, unusual analyses planned?)
We will also test whether errors in the AI-assisted group occur more frequently during the first trials in which the 40/60 stimuli are presented, with participants progressively learning to reduce their errors. Thus, a reduction in the number of errors should be observed as training proceeds in the AI-Assisted group; the control group should show fewer errors than the experimental group, particularly during the early trials. In addition, we included some post-experimental questions to know if participants have seen and followed the recommendation of the AI and to explore the confidence participants have felt in their own ability to perform the task, and the confidence they place in general in artificial intelligence algorithms in health. This information is collected with exploratory purposes and thus we do not pre-resister any prediction.