The impact of AI errors in judgments in a human-in-the-loop process (#139164)

Created: 07/24/2023 04:23 AM (PT)
Public: 09/27/2023 12:17 PM (PT)

Author(s)
Ujué Agudo (Bikolabs; University of Deusto) - ujue.agudo@biko2.com
Karlos G. Liberal (Bikolabs) - karlosgliberal@gmail.com
Miren Arrese (Bikolabs) - miren.arrese@biko2.com
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?
Participants who receive AI support at the beginning of the process of assessing the guilt of an offender will show more compliance with the erroneous AI support than those who receive the erroneous AI support at the end of the process.
In addition, we expect that participants who judge the defendant without having seen the AI's erroneous support will consider the defendant less guilty when the testimonies indicate innocence than participants who judge the defendant after having previously seen the AI's erroneous support; and also that participants who judge without having seen the AI's erroneous support will consider the defendant more guilty when the testimonies indicate guilt than participants who previously saw the AI's erroneous support.

3) Describe the key dependent variable(s) specifying how they will be measured.
One dependent variable will be compliance with the erroneous AI support. When participants receive the erroneous support from the AI, they will have to choose between two options: confirm the AI's verdict or modify it. We will measure how many participants choose to confirm, thus showing compliance with AI support.
The other dependent variable will be the participants' judgment of the defendants's guilt when the testimonies suggest innocence and when they suggest guilt. This judgment will be assessed using a 0-100 scale, with 0 being absolutely innocent, and 100 being absolutely guilty.

4) How many and which conditions will participants be assigned to?
We will use a mixed 2x2 design, with one variable being between groups and the other one within participants. Thus, participants will be randomly assigned to one of two groups: the group in which they make their judgment and then they receive the AI support and are required to indicate whether or not they confirm this support (Judgment-->ComplianceAI); or the group in which they first receive the AI support and indicate whether or not they confirm it and then make their judgment (ComplianceAI-->Judgment).
The other variable will be within participants. In half of the cases the testimonies suggest innocence (and when the AI errs, it suggests guilt); in the other cases, the testimonies suggest guilt (and when the AI errs, it suggests innocence).

5) Specify exactly which analyses you will conduct to examine the main question/hypothesis.
We will run a binomial test to analyze the difference between the groups in terms of the participants' compliance with erroneous AI support. We expect lower compliance in the Judgment-->ComplianceAI group.
We will also run a 2 (group: Judgment-->ComplianceAI vs. ComplianceAI-->Judgment) x 2 (type of testimonies: innocence vs. guilt) mixed ANOVA on the participants judgments of guilt. We expect that the participants in the Judgment-->ComplianceAI group will judge the defendant less guilty than in the ComplianceAI-->Judgment group when the testimonies indicate innocence; and that the participants in the Judgment-->ComplianceAI group will judge the defendant more guilty than in the ComplianceAI-->Judgment group when the testimonies indicate guilt.

6) Describe exactly how outliers will be defined and handled, and your precise rule(s) for excluding observations.
We will exclude the data of participants who complete the survey in more than 45 minutes. The duration of the study is estimated at 12 minutes.

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 will recruit 260 participants in the Prolific Academic online platform. We will use Prolific's internal selection service to recruit this specific sample: participants over 18 years of age and with Spanish as their first language, that have not previously participated in other experiments conducted by our research team on the Prolific platform.

8) Anything else you would like to pre-register? (e.g., secondary analyses, variables collected for exploratory purposes, unusual analyses planned?)
We collect some other secondary variables for exploratory purposes, such as the participants' compliance with accurate AI support and judgment in all criminal cases, not just those in which the AI support is erroneous.