

Credit where credit is due: how can AI's role in credit decisions be explained?

Data Dictionary

'Age' – Participant age in years

'agegroup' – Categorical variable for age. Options include: 18-34, 35-44, 45-54, 55-64, 65-74, 75+.

'AQ1' – Categorical variable indicating the answer option the participant selected when asked how important they found the information (options included 'Very important', 'Slightly important', 'Neither unimportant nor important', 'Slightly unimportant', 'Very unimportant').

'AQ2' – Categorical variable indicating the answer option the participant selected when asked how helpful they found the information (options included 'Very helpful', 'Slightly helpful', 'Neither unhelpful nor helpful', 'Slightly unhelpful', 'Very unhelpful').

'AQ3' – Categorical variable indicating the answer option the participant selected when asked to what extent they agree that the information provided was sufficient (options included 'Strongly agree', 'Somewhat agree', 'Neither agree nor disagree', 'Somewhat disagree', 'Strongly disagree').

'AQ4' – Categorical variable indicating the answer option the participant selected when asked whether they felt confident in their ability to disagree with the decision (options included 'Very confident', 'Somewhat confident', 'Neither confident nor unconfident', 'Somewhat unconfident', 'Very unconfident').

'Comp1' – Binary variable indicating whether the participant correctly answered the first comprehension question (CQ1/T1Q1/T2Q1/T3Q1), relating to the comprehension of basic information about how the algorithm works (1 = correct response, 0 = incorrect response).

'Comp2' – Binary variable indicating whether the participant correctly answered the second comprehension question (CQ2/T1Q2/T2Q2/T3Q3), relating to the comprehension of information about the directionality of the 'debt collection accounts opened against you' feature information (1 = correct response, 0 = incorrect response).

'Comp3' – Binary variable indicating whether the participant correctly answered the third comprehension question (CQ3/T1Q3/T2Q3/T4Q3), relating to the comprehension of features importance information (1 = correct response, 0 = incorrect response).

'Completed' – Binary variable. NA indicates that participant did not respond to all questions

'CompTotal' – Sum of Comp1, Comp2, Comp3.

'confident' - Binary variable coded 1 if the participant reports that information provided was important by selecting 'Somewhat confident' or 'Very confident' (AQ4). Coded 0 if participant selects any other option or if no response is given. Participants who drop out of the experiment after exposure to treatment are also coded as 0.

'CQ1' - Categorical variable indicating the answer option participants in the control group selected for Comprehension Question 1.

'CQ2' - Categorical variable indicating the answer option participants in the control group selected for Comprehension Question 2.

'CQ3' - Categorical variable indicating the answer option participants in the control group selected for Comprehension Question 3.

'E5' - Binary variable indicating whether participants who correctly identified the incorrect AI-assisted decision on profile 3 correctly identified the corresponding reasoning (1 = correctly identified reasoning, 0 = did not correctly identify reasoning).

'E6' - Binary variable indicating whether participants who correctly identified the incorrect AI-assisted decision on profile 4 correctly identified the corresponding reasoning (1 = correctly identified reasoning, 0 = did not correctly identify reasoning).

'E7' - Binary variable indicating whether participants who correctly identified the incorrect AI-assisted decision on profile 5 correctly identified the corresponding reasoning (1 = correctly identified reasoning, 0 = did not correctly identify reasoning).

'Finished' - Binary variable. TRUE indicates that the participant completed the experiment. FALSE indicated that the participant did not complete the experiment.

'helpful' - Binary variable coded 1 if the participant reports that information provided was helpful by selecting 'Slightly helpful' or 'Very helpful' (Preference Q2). Coded 0 if participant selects any other option or if no response is given. Participants who drop out of the experiment after exposure to treatment are also coded as 0.

'importance' - Binary variable coded 1 if the participant reports that information provided was important by selecting 'Slightly important' or 'Very important' (Preference Q1). Coded 0 if participant selects any other option or if no response is given. Participants who drop out of the experiment after exposure to treatment are also coded as 0.

'JTotal' - Absolute number of AI-assisted decisions judged correctly (i.e., correct identification of correct/incorrect decisions made by algorithm; min = 0, max = 5)

'P1J' - Binary variable indicating whether the participant accepted or challenged Scenario 1 (Accept; Challenge).

'P2J' – Binary variable indicating whether the participant accepted or challenged Scenario 2 (Accept; Challenge).

'P3J' – Binary variable indicating whether the participant accepted or challenged Scenario 3 (Accept; Challenge).

'P4J' – Binary variable indicating whether the participant accepted or challenged Scenario 4 (Accept; Challenge).

'P5J' – Binary variable indicating whether the participant accepted or challenged Scenario 5 (Accept; Challenge).

'PO1.P1J' – Binary variable indicating whether the participant correctly judged the AI-assisted decision when the profile was a correct acceptance (Profile 1; 1 = correct judgement, 0 = incorrect judgement).

'PO1.P3J' – Binary variable indicating whether the participant correctly judged the AI-assisted decision when the profile had an 'incorrect prediction due to data input' error (Profile 3; 1 = correct judgement, 0 = incorrect judgement).

'PO1.P4J' – Binary variable indicating whether the participant correctly judged the AI-assisted decision when the profile had an 'over-reliance on one feature' error (Profile 3; 1 = correct judgement, 0 = incorrect judgement).

'PO1.P5J' – Binary variable indicating whether the participant correctly judged the AI-assisted decision when the profile had a 'failure to consider relevant feature' error (Profile 3; 1 = correct judgement, 0 = incorrect judgement).

'Progress' – Percentage of the experiment the participant progressed through

'Sex' – Sex assignment at birth. Categorical variable: Male, Female, PNTS (Prefer Not To Say)

'sufficient' - Binary variable coded 1 if the participant reports that information provided was sufficient by selecting 'Somewhat agree' or 'Strongly agree' (AQ3). Coded 0 if participant selects any other option or if no response is given. Participants who drop out of the experiment after exposure to treatment are also coded as 0.

'T1Q1' – Categorical variable indicating the answer option participants in Treatment 1 selected for Comprehension Question 1.

'T1Q2' – Categorical variable indicating the answer option participants in Treatment 1 selected for Comprehension Question 2.

'T1Q3' – Categorical variable indicating the answer option participants in Treatment 1 selected for Comprehension Question 3.

'T2Q1' – Categorical variable indicating the answer option participants in Treatment 2 selected for Comprehension Question 1.

'T2Q2' – Categorical variable indicating the answer option participants in Treatment 2 selected for Comprehension Question 2.

'T2Q3' – Categorical variable indicating the answer option participants in Treatment 2 selected for Comprehension Question 3.

'T3Q1' – Categorical variable indicating the answer option participants in Treatment 3 selected for Comprehension Question 1.

'T3Q2' – Categorical variable indicating the answer option participants in Treatment 3 selected for Comprehension Question 2.

'T3Q3' – Categorical variable indicating the answer option participants in Treatment 3 selected for Comprehension Question 3.

'Time taken' – Number of seconds the participant spent completing the experiment.

'treatment' – Categorical variable indicating to which treatment group the participant was assigned. 0: control; 1: features-based; 2: combination data-centric and features-based; 3: combination data-centric, features-based, and decision rule.