Financial Conduct Authority

Research Note: Annex

24 February 2025

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

Contents

Annex

Sample and power calculations

To ensure robust statistical conclusions, we conducted power calculations under the following assumptions:

- Significance level (α): 0.05, adjusted for multiple comparisons using Bonferroni correction (3 primary analyses)
- Statistical power: 0.8 (80%)
- Effect size determination: Baseline rates for the "correct judgement" metric were derived from a pre-test, which indicated a baseline proportion of 87.5%. Given the increased complexity of features in the main experiment, we conservatively assumed a lower baseline of 50% for power calculations

The parameters for the power analysis were therefore:

- Baseline proportion (P1): 0.50
- Minimum Detectable Effect (MDE): 5 percentage points (pp)
- Test type: Two-sided
- Sample size per trial arm (N): 2,215

This sample size was calculated to achieve the stated power and significance thresholds, yielding a total required sample size of 8,860 participants across 4 trial arms. This allocation maximised power to detect an effect size of 5 pp within the constraints of our budget and logistical considerations. The MDE of 5 pp was established as a meaningful threshold in consultation with both policy and academic stakeholders.

Missing data strategy

Our approach to missing outcome data was to code missing/incomplete responses as 0, conditional on exposure to treatment. This means we analysed all participants' data, as long as they were exposed to treatment. This approach is effectively the 'lower bound' estimate of Horowitz-Manski (i.e., assumes all missing data is 'incorrect'). This means that if attrition is random across treatments, the relative percentage difference in outcomes across treatments should be the same, but the absolute percentage point difference will be slightly lower than reality (because it is unlikely that *all* participants who drop out or do not respond to a question would necessarily have been incorrect).

Our approach to missing covariates was to code them as 'prefer not to say' (PNTS).

We did not find differential attrition across treatments. This means that participants did not drop out of the experiment at statistically significantly different rates across treatment groups. This further reduces the risk of introducing bias to the results with our missing data strategy.

Sensitivity analysis

The main analysis reported was based on an Ordinary Least Squares (OLS) regression model with covariates. However, as part of the sensitivity analyses, we also ran a quasi-binomial regression (a binomial model corrected for overdispersion) to estimate the impact of treatment assignment on the average number of correct judgements about whether AI-assisted decisions were correct across five profiles. We did not find significant differences between the quasi-binomial and OLS models.

Where we ran OLS regressions, we checked that the baseline proportion was > 5% or < 95% for the given outcome measure, as the linear approximation assumption would otherwise be violated. Where this was the case, we then ran logistic regression and compared results to ensure there was no significant difference. We only needed to run this sensitivity analysis for S4 and found no significant differences across regression models, so reported the OLS regression for ease of interpretation.

Multiple comparisons

We corrected for multiple hypotheses testing using the Bonferroni correction approach (Abdi, 2007), which involved dividing the traditional significance threshold (α = 0.05) by the number of comparisons made. With 3 treatments compared to the control, our primary and secondary analyses, as well as our analyses of attitudinal outcomes, we adopted a significance threshold of α = 0.0167. We did this because the more comparisons across groups we make, the greater is the risk that a result is a 'false positive' (where a model indicates a finding as statistically significant by sheer chance rather than as the result of an actual effect). This adjustment helps mitigate this risk by making the significance threshold more conservative.

Regression models and covariates

Primary analysis: effect of treatment on proportion of correct judgements

Outcome: Proportion of correctly judged profiles (0 to 1) across five profiles.

Model Specification:

$$\mathrm{Y}_i = \beta_0 + \beta_{1-3}X_i + \beta_X X_i + \omega_i$$

Where:

- Y_i is a proportion of correctly judged profiles out of 5 (from 0 to 1); and
- β_{1-3} are the three treatment allocation dummies (one for each treatment group apart from the control); and
- β_X is the matrix of covariates, as specified below; and
- ω_i are Huber White robust standard errors.

Secondary Analysis:

Correct Judgements by Error Type/Scenario

Outcomes:

- S1: Correct judgement for 'incorrect prediction due to data input' error.
- S2: Correct judgement for 'overreliance on one feature' error.
- S3: Correct judgement for 'failure to consider relevant feature' error.
- S4: Correct acceptance (Scenario 1).
- S5: Correct rejection (Scenario 2).

Model Specification (for each outcome):

$$\mathrm{Y}_i = \beta_0 + \beta_{1-3} X_i + \beta_X X_i + \omega_i$$

Where:

- Yi is the proportion of those coded as 1 for each of the Scenarios;
- β_{1-3} are the three treatment allocation dummies (one for each treatment group apart from the control);
- β_X is the matrix of covariates, as specified below; and
- ω_i are Huber White robust standard errors.

Comprehension Outcomes

Outcomes:

- S6: Comprehension of basic information about algorithm use.
- S7: Comprehension of directionality of feature information.
- S8: Comprehension of feature importance information.

Model Specification:

$$\mathrm{Y}_i = \beta_0 + \beta_{1-3} X_i + \beta_X X_i + \omega_i$$

Where:

- Yi is the proportion of those coded as 1 for each of the comprehension outcomes;
- β_{1-3} are the three treatment allocation dummies (one for each treatment group apart from the control);
- β_X is the matrix of covariates, as specified below; and
- ω_i are Huber White robust standard errors.

Exploratory Analyses: Effect of Treatment on Attitudinal Outcomes

Outcomes:

E1: Importance of information.

E2: Helpfulness of information.

E3: Sufficiency of information.

E4: Confidence in disagreement.

Model Specification (for each outcome):

$$\mathrm{Y}_i = \beta_0 + \beta_{1-3} X_i + \beta_X X_i + \omega_i$$

Where:

- Yi is the proportion of those coded as 1 for each of the attitudinal outcomes;
- β_{1-3} are the three treatment allocation dummies (one for each treatment group apart from the control);
- β_X is the matrix of covariates, as specified below; and
- ω_i are Huber White robust standard errors.

Covariates

All models included the following covariates to increase statistical power:

- Sex assignment at birth: Female (base group), Male, PNTS
- Age group: 18-24 (base group), 25-34, 35-44, 45-54, 55-64, 65-74, 75+, prefer not to say (PNTS)

Covariates were approximately balanced across treatment groups. They were included to increase the precision of the estimated treatment effects.

Explanation genres and scenarios

Please note the below section includes all scenarios, in order from Scenario 1 – Scenario 5. For each scenario, we include each explanation genre.

Scenario 1 (Correct acceptance), Data-centric explanation (control)

Message received from the credit provider:

Thank you for your application for our Regular Credit Card. We would like to inform you that the result of your application is: **Accepted**

We have a tool that uses algorithms to assess credit card applications. The tool doesn't make decisions on its own and the application would not be rejected by the tool alone, but it helps our credit card officers find applications that might have a higher risk of not being repaid.

How the algorithm helped us make this decision:

We use information about you to see how your application compares to other people's data in our system. This helps us understand your situation better. The table below shows information about you, how it compares to the average of past applicants, and where the information came from. However, not all the information shown below is always considered by the algorithm.

Information type	Your information	Average for our past applicants	Source of information
Debt collection accounts opened against you in last 24 months	0	0.10	Credit reporting agencies
Late payments or overdue accounts in last 24 months	0	0.14	Credit reporting agencies
Percentage of credit that you've already paid off now	100%	47%	Credit reporting agencies
Percentage of credit limit that you're using	11%	38%	Credit reporting agencies
Annual income	£40,000	£43,166	You provided
Current total credit card limit	£2,300	£10,257	Credit reporting agencies

If you believe there is something wrong with how the algorithm assisted us in making the decision, **you can challenge the decision**. Please note that we can only reconsider the decision if you provide a **valid, appropriate reason** for why you'd like to challenge the decision.

Scenario 1 (Correct acceptance), Features-based explanation

Message received from the credit provider:

Thank you for your application for our Regular Credit Card. We would like to inform you that the result of your application is: **Accepted**

We have a tool that uses algorithms to assess credit card applications. The tool doesn't make decisions on its own and the application would not be rejected by the tool alone, but it helps our credit card officers find applications that might have a higher risk of not being repaid.

How the algorithm helped us make this decision:

We use information about you to see how your application compares to other people's data in our system. This helps us understand your situation better. The table below shows information about you that the algorithm considered for your application as well as how important that information is to the decision.

Information type	Your application	Importance of information	Effect of information
Debt collection accounts opened against you in last 24 months	0	Most important	Increased your likelihood of approval
Late payments or overdue accounts in last 24 months	0	Very important	Increased your likelihood of approval

If you believe there is something wrong with how the algorithm assisted us in making the decision, **you can challenge the decision**. Please note that we can only reconsider the decision if you provide a **valid, appropriate reason** for why you'd like to challenge the decision.

Scenario 1 (Correct acceptance), Combination of data-centric and features-based explanation

Message received from the credit provider:

Thank you for your application for our Regular Credit Card. We would like to inform you that the result of your application is: **Accepted**

We have a tool that uses algorithms to assess credit card applications. The tool doesn't make decisions on its own and the application would not be rejected by the tool alone, but it helps our credit card officers find applications that might have a higher risk of not being repaid.

How the algorithm helped us make this decision:

We use information about you to see how your application compares to other people's data in our system. This helps us understand your situation better. The table below shows information about you that the algorithm considered for your application, where the information came from, how important that information is to the decision, and how it compares to the average of past applicants.

Information type	Your application	Average for past applicants	Importance of information	Effect of information	Source of information
Debt collection accounts opened against you in last 24 months	0	0.10	Most important	Increased your likelihood of approval	Credit reporting agencies
Late payments or overdue accounts in last 24 months	0	0.14	Very important	Increased your likelihood of approval	Credit reporting agencies

If you believe there is something wrong with how the algorithm assisted us in making the decision, **you can challenge the decision**. Please note that we can only reconsider the decision if you provide a **valid, appropriate reason** for why you'd like to challenge the decision.

Scenario 1 (Correct acceptance), Combination + decision rule

Message received from the credit provider:

Thank you for your application for our Regular Credit Card. We would like to inform you that the result of your application is: **Accepted**

We have a tool that uses algorithms to assess credit card applications. The tool doesn't make decisions on its own and the application would not be rejected by the tool alone, but it helps our credit card officers find applications that might have a higher risk of not being repaid.

How the algorithm helped us make this decision:

We use information about you to see how your application compares to other people's data in our system. This helps us understand your situation better. The table below shows information about you that the algorithm considered for your application, where the information came from, how important that information is to the decision, and how it compares to the average of past applicants. The decision rule explains the basis for this decision.

Information type	Your application	Average for past applicants	Importance of information	Effect of information	Source of information
Debt collection accounts opened against you in last 24 months	0	0.10	Most important	Increased your likelihood of approval	Credit reporting agencies
Late payments or overdue accounts in last 24 months	0	0.14	Very important	Increased your likelihood of approval	Credit reporting agencies

The following decision rule was applied:
If debt collection accounts opened against you in last 24 months is less than 0.5 and late payments or overdue accounts in last 24 months is less than 0.5 then Accept the application

If you believe there is something wrong with how the algorithm assisted us in making the decision, **you can challenge the decision**. Please note that we can only reconsider the decision if you provide a **valid, appropriate reason** for why you'd like to challenge the decision.

Scenario 2 (Correct rejection), Data-centric explanation (control)

Message received from the credit provider:

Thank you for your application for our Regular Credit Card. We would like to inform you that the result of your application is: **Rejected**

We have a tool that uses algorithms to assess credit card applications. The tool doesn't make decisions on its own and the application would not be rejected by the tool alone, but it helps our credit card officers find applications that might have a higher risk of not being repaid.

How the algorithm helped us make this decision:

We use information about you to see how your application compares to other people's data in our system. This helps us understand your situation better. The table below shows information about you, how it compares to the average of past applicants, and where the information came from. However, not all the information shown below is always considered by the algorithm.

Information type	Your information	Average for our past applicants	Source of information
Debt collection accounts opened against you in last 24 months	4	0.10	Credit reporting agencies
Late payments or overdue accounts in last 24 months	1	0.14	Credit reporting agencies
Percentage of credit that you've already paid off now	20%	47%	Credit reporting agencies
Percentage of credit limit that you're using	96%	38%	Credit reporting agencies
Annual income	£21,117	£43,166	You provided
Current total credit card limit	£400	£10,257	Credit reporting agencies

If you believe there is something wrong with how the algorithm assisted us in making the decision, **you can challenge the decision**. Please note that we can only reconsider the decision if you provide a **valid, appropriate reason** for why you'd like to challenge the decision.

Scenario 2 (Correct rejection), Features-based explanation

Message received from the credit provider:

Thank you for your application for our Regular Credit Card. We would like to inform you that the result of your application is: **Rejected**

We have a tool that uses algorithms to assess credit card applications. The tool doesn't make decisions on its own and the application would not be rejected by the tool alone, but it helps our credit card officers find applications that might have a higher risk of not being repaid.

How the algorithm helped us make this decision:

We use information about you to see how your application compares to other people's data in our system. This helps us understand your situation better. The table below shows information about you that the algorithm considered for your application as well as how important that information is to the decision.

Information type	Your application	Importance of information	Effect of information
Debt collection accounts opened against you in last 24 months	4	Most important	Decreased your likelihood of approval

If you believe there is something wrong with how the algorithm assisted us in making the decision, **you can challenge the decision**. Please note that we can only reconsider the decision if you provide a **valid, appropriate reason** for why you'd like to challenge the decision.

Scenario 2 (Correct rejection), Data-centric and features-based explanation

Message received from the credit provider:

Thank you for your application for our Regular Credit Card. We would like to inform you that the result of your application is: **Rejected**

We have a tool that uses algorithms to assess credit card applications. The tool doesn't make decisions on its own and the application would not be rejected by the tool alone, but it helps our credit card officers find applications that might have a higher risk of not being repaid.

How the algorithm helped us make this decision:

We use information about you to see how your application compares to other people's data in our system. This helps us understand your situation better. The table below shows information about you that the algorithm considered for your application, where the information came from, how important that information is to the decision, and how it compares to the average of past applicants.

Information type	Your application	Average for past applicants	Importance of information	Effect of information	Source of information
Debt collection accounts opened against you in last 24 months	4	0.10	Most important	Decreased your likelihood of approval	Credit reporting agencies

If you believe there is something wrong with how the algorithm assisted us in making the decision, **you can challenge the decision**. Please note that we can only reconsider the decision if you provide a **valid, appropriate reason** for why you'd like to challenge the decision.

Scenario 2 (Correct rejection), Combination + decision rule explanation

Message received from the credit provider:

Thank you for your application for our Regular Credit Card. We would like to inform you that the result of your application is: **Rejected**

We have a tool that uses algorithms to assess credit card applications. The tool doesn't make decisions on its own and the application would not be rejected by the tool alone, but it helps our credit card officers find applications that might have a higher risk of not being repaid.

How the algorithm helped us make this decision:

We use information about you to see how your application compares to other people's data in our system. This helps us understand your situation better. The table below shows information about you that the algorithm considered for your application, where the information came from, how important that information is to the decision, and how it compares to the average of past applicants. The decision rule explains the basis for this decision.

Information type	Your application	Average for past applicants	Importance of information	Effect of information	Source of information
Debt collection accounts opened against you in last 24 months	4	0.10	Most important	Decreased your likelihood of approval	Credit reporting agencies

The following decision rule was applied:
If debt collection accounts opened against you in last 24 months is greater than 0.5 then Reject the application

If you believe there is something wrong with how the algorithm assisted us in making the decision, **you can challenge the decision**. Please note that we can only reconsider the decision if you provide a **valid, appropriate reason** for why you'd like to challenge the decision.

Scenario 3 (Incorrect rejection – data input error), Data-centric explanation (control)

Message received from the credit provider:

Thank you for your application for our Regular Credit Card. We would like to inform you that the result of your application is: **Rejected**

We have a tool that uses algorithms to assess credit card applications. The tool doesn't make decisions on its own and the application would not be rejected by the tool alone, but it helps our credit card officers find applications that might have a higher risk of not being repaid.

How the algorithm helped us make this decision:

We use information about you to see how your application compares to other people's data in our system. This helps us understand your situation better. The table below shows information about you, how it compares to the average of past applicants, and where the information came from. However, not all the information shown below is always considered by the algorithm.

Information type	Your information	Average for our past applicants	Source of information
Debt collection accounts opened against you in last 24 months	1	0.10	Credit reporting agencies
Late payments or overdue accounts in last 24 months	0	0.14	Credit reporting agencies
Percentage of credit that you've already paid off now	100%	47%	Credit reporting agencies
Percentage of credit limit that you're using	11%	38%	Credit reporting agencies
Annual income	£40,000	£43,166	You provided
Current total credit card limit	£2,300	£10,257	Credit reporting agencies

If you believe there is something wrong with how the algorithm assisted us in making the decision, **you can challenge the decision**. Please note that we can only reconsider the decision if you provide a **valid, appropriate reason** for why you'd like to challenge the decision.

Scenario 3 (Incorrect rejection – data input error), Features-based explanation

Message received from the credit provider:

Thank you for your application for our Regular Credit Card. We would like to inform you that the result of your application is: **Rejected**

We have a tool that uses algorithms to assess credit card applications. The tool doesn't make decisions on its own and the application would not be rejected by the tool alone, but it helps our credit card officers find applications that might have a higher risk of not being repaid.

How the algorithm helped us make this decision:

We use information about you to see how your application compares to other people's data in our system. This helps us understand your situation better. The table below shows information about you that the algorithm considered for your application as well as how important that information is to the decision.

Information type	Your application	Importance of information	Effect of information
Debt collection accounts opened against you in last 24 months	1	Most important	Decreased your likelihood of approval

If you believe there is something wrong with how the algorithm assisted us in making the decision, **you can challenge the decision**. Please note that we can only reconsider the decision if you provide a **valid, appropriate reason** for why you'd like to challenge the decision.

Scenario 3 (Incorrect rejection – data input error), Combination data-centric and features-based explanation

Message received from the credit provider:

Thank you for your application for our Regular Credit Card. We would like to inform you that the result of your application is: **Rejected**

We have a tool that uses algorithms to assess credit card applications. The tool doesn't make decisions on its own and the application would not be rejected by the tool alone, but it helps our credit card officers find applications that might have a higher risk of not being repaid.

How the algorithm helped us make this decision:

We use information about you to see how your application compares to other people's data in our system. This helps us understand your situation better. The table below shows information about you that the algorithm considered for your application, where the information came from, how important that information is to the decision, and how it compares to the average of past applicants.

Information type	Your application	Average for past applicants	Importance of information	Effect of information	Source of information
Debt collection accounts opened against you in last 24 months	1	0.10	Most important	Decreased your likelihood of approval	Credit reporting agencies

If you believe there is something wrong with how the algorithm assisted us in making the decision, **you can challenge the decision**. Please note that we can only reconsider the decision if you provide a **valid, appropriate reason** for why you'd like to challenge the decision.

Scenario 3 (Incorrect rejection – data input error), Combination + decision rule explanation

Message received from the credit provider:

Thank you for your application for our Regular Credit Card. We would like to inform you that the result of your application is: **Rejected**

We have a tool that uses algorithms to assess credit card applications. The tool doesn't make decisions on its own and the application would not be rejected by the tool alone, but it helps our credit card officers find applications that might have a higher risk of not being repaid.

How the algorithm helped us make this decision:

We use information about you to see how your application compares to other people's data in our system. This helps us understand your situation better. The table below shows information about you that the algorithm considered for your application, where the information came from, how important that information is to the decision, and how it compares to the average of past applicants. The decision rule explains the basis for this decision.

Information type	Your application	Average for past applicants	Importance of information	Effect of information	Source of information
Debt collection accounts opened against you in last 24 months	1	0.10	Most important	Decreased your likelihood of approval	Credit reporting agencies

The following decision rule was applied:
If debt collection accounts opened against you in last 24 months is greater than 0.5 then Reject the application

If you believe there is something wrong with how the algorithm assisted us in making the decision, **you can challenge the decision**. Please note that we can only reconsider the decision if you provide a **valid, appropriate reason** for why you'd like to challenge the decision.

Scenario 4 (Incorrect rejection – over-reliance on one feature), Data-centric explanation (control)

Message received from the credit provider:

Thank you for your application for our Regular Credit Card. We would like to inform you that the result of your application is: **Rejected**

We have a tool that uses algorithms to assess credit card applications. The tool doesn't make decisions on its own and the application would not be rejected by the tool alone, but it helps our credit card officers find applications that might have a higher risk of not being repaid.

How the algorithm helped us make this decision:

We use information about you to see how your application compares to other people's data in our system. This helps us understand your situation better. The table below shows information about you, how it compares to the average of past applicants, and where the information came from. However, not all the information shown below is always considered by the algorithm.

Information type	Your information	Average for our past applicants	Source of information
Debt collection accounts opened against you in last 24 months	1	0.10	Credit reporting agencies
Late payments or overdue accounts in last 24 months	0	0.14	Credit reporting agencies
Percentage of credit that you've already paid off now	100%	47%	Credit reporting agencies
Percentage of credit limit that you're using	11%	38%	Credit reporting agencies
Annual income	£180,000	£43,166	You provided
Current total credit card limit	£6,000	£10,257	Credit reporting agencies

If you believe there is something wrong with how the algorithm assisted us in making the decision, **you can challenge the decision**. Please note that we can only reconsider the decision if you provide a **valid, appropriate reason** for why you'd like to challenge the decision.

Scenario 4 (Incorrect rejection – over-reliance on one feature), Features-based explanation

Message received from the credit provider:

Thank you for your application for our Regular Credit Card. We would like to inform you that the result of your application is: **Rejected**

We have a tool that uses algorithms to assess credit card applications. The tool doesn't make decisions on its own and the application would not be rejected by the tool alone, but it helps our credit card officers find applications that might have a higher risk of not being repaid.

How the algorithm helped us make this decision:

We use information about you to see how your application compares to other people's data in our system. This helps us understand your situation better. The table below shows information about you that the algorithm considered for your application as well as how important that information is to the decision.

Information type	Your application	Importance of information	Effect of information
Debt collection accounts opened against you in last 24 months	1	Most important	Decreased your likelihood of approval

If you believe there is something wrong with how the algorithm assisted us in making the decision, **you can challenge the decision**. Please note that we can only reconsider the decision if you provide a **valid, appropriate reason** for why you'd like to challenge the decision.

Scenario 4 (Incorrect rejection – over-reliance on one feature), Combination of data-centric and features-based explanation

Message received from the credit provider:

Thank you for your application for our Regular Credit Card. We would like to inform you that the result of your application is: **Rejected**

We have a tool that uses algorithms to assess credit card applications. The tool doesn't make decisions on its own and the application would not be rejected by the tool alone, but it helps our credit card officers find applications that might have a higher risk of not being repaid.

How the algorithm helped us make this decision:

We use information about you to see how your application compares to other people's data in our system. This helps us understand your situation better. The table below shows information about you that the algorithm considered for your application, where the information came from, how important that information is to the decision, and how it compares to the average of past applicants.

Information type	Your application	Average for past applicants	Importance of information	Effect of information	Source of information
Debt collection accounts opened against you in last 24 months	1	0.10	Most important	Decreased your likelihood of approval	Credit reporting agencies

If you believe there is something wrong with how the algorithm assisted us in making the decision, **you can challenge the decision**. Please note that we can only reconsider the decision if you provide a **valid, appropriate reason** for why you'd like to challenge the decision.

Scenario 4 (Incorrect rejection – over-reliance on one feature), Combination + decision rule explanation

Message received from the credit provider:

Thank you for your application for our Regular Credit Card. We would like to inform you that the result of your application is: **Rejected**

We have a tool that uses algorithms to assess credit card applications. The tool doesn't make decisions on its own and the application would not be rejected by the tool alone, but it helps our credit card officers find applications that might have a higher risk of not being repaid.

How the algorithm helped us make this decision:

We use information about you to see how your application compares to other people's data in our system. This helps us understand your situation better. The table below shows information about you that the algorithm considered for your application, where the information came from, how important that information is to the decision, and how it compares to the average of past applicants. The decision rule explains the basis for this decision.

Information type	Your application	Average for past applicants	Importance of information	Effect of information	Source of information
Debt collection accounts opened against you in last 24 months	1	0.10	Most important	Decreased your likelihood of approval	Credit reporting agencies

The following decision rule was applied:
If debt collection accounts opened against you in last 24 months is greater than 0.5 then Reject the application

If you believe there is something wrong with how the algorithm assisted us in making the decision, **you can challenge the decision**. Please note that we can only reconsider the decision if you provide a **valid, appropriate reason** for why you'd like to challenge the decision.

Scenario 5 (Incorrect rejection – failure to consider relevant features), Data-centric explanation (control)

Message received from the credit provider:

Thank you for your application for our Regular Credit Card. We would like to inform you that the result of your application is: **Rejected**

We have a tool that uses algorithms to assess credit card applications. The tool doesn't make decisions on its own and the application would not be rejected by the tool alone, but it helps our credit card officers find applications that might have a higher risk of not being repaid.

How the algorithm helped us make this decision:

We use information about you to see how your application compares to other people's data in our system. This helps us understand your situation better. The table below shows information about you, how it compares to the average of past applicants, and where the information came from. However, not all the information shown below is always considered by the algorithm.

Information type	Your information	Average for our past applicants	Source of information
Debt collection accounts opened against you in last 24 months	0	0.10	Credit reporting agencies
Late payments or overdue accounts in last 24 months	1	0.14	Credit reporting agencies
Percentage of credit that you've already paid off now	2%	47%	Credit reporting agencies
Percentage of credit limit that you're using	80%	38%	Credit reporting agencies
Annual income	£280,000	£43,166	You provided
Current total credit card limit	£2,000	£10,257	Credit reporting agencies

If you believe there is something wrong with how the algorithm assisted us in making the decision, **you can challenge the decision**. Please note that we can only reconsider the decision if you provide a **valid, appropriate reason** for why you'd like to challenge the decision.

Scenario 5 (Incorrect rejection – failure to consider relevant features), Features-based explanation

Message received from the credit provider:

Thank you for your application for our Regular Credit Card. We would like to inform you that the result of your application is: **Rejected**

We have a tool that uses algorithms to assess credit card applications. The tool doesn't make decisions on its own and the application would not be rejected by the tool alone, but it helps our credit card officers find applications that might have a higher risk of not being repaid.

How the algorithm helped us make this decision:

We use information about you to see how your application compares to other people's data in our system. This helps us understand your situation better. The table below shows information about you that the algorithm considered for your application as well as how important that information is to the decision.

Information type	Your application	Importance of information	Effect of information
Debt collection accounts opened against you in last 24 months	0	Most important	Increased your likelihood of approval
Late payments or overdue accounts in last 24 months	1	Very important	Decreased your likelihood of approval
Percentage of credit that you've already paid off	2%	Very important	Decreased your likelihood of approval
Percentage of credit limit that you're using now	80%	Important	Decreased your likelihood of approval

If you believe there is something wrong with how the algorithm assisted us in making the decision, **you can challenge the decision**. Please note that we can only reconsider the decision if you provide a **valid, appropriate reason** for why you'd like to challenge the decision.

Scenario 5 (Incorrect rejection – failure to consider relevant features), Combination of data-centric and features-based explanation

Message received from the credit provider:

Thank you for your application for our Regular Credit Card. We would like to inform you that the result of your application is: **Rejected**

We have a tool that uses algorithms to assess credit card applications. The tool doesn't make decisions on its own and the application would not be rejected by the tool alone, but it helps our credit card officers find applications that might have a higher risk of not being repaid.

How the algorithm helped us make this decision:

We use information about you to see how your application compares to other people's data in our system. This helps us understand your situation better. The table below shows information about you that the algorithm considered for your application, where the information came from, how important that information is to the decision, and how it compares to the average of past applicants.

Information type	Your application	Average for past applicants	Importance of information	Effect of information	Source of information
Debt collection accounts opened against you in last 24 months	0	0.10	Most important	Increased your likelihood of approval	Credit reporting agencies
Late payments or overdue accounts in last 24 months	1	0.14	Very important	Decreased your likelihood of approval	Credit reporting agencies
Percentage of credit that you've already paid off	2%	47%	Very important	Decreased your likelihood of approval	Credit reporting agencies
Percentage of credit limit that you're using now	80%	38%	Important	Decreased your likelihood of approval	Credit reporting agencies

If you believe there is something wrong with how the algorithm assisted us in making the decision, **you can challenge the decision**. Please note that we can only reconsider the decision if you provide a **valid, appropriate reason** for why you'd like to challenge the decision.

Scenario 5 (Incorrect rejection – failure to consider relevant features), Combination + decision rule explanation

Message received from the credit provider:

Thank you for your application for our Regular Credit Card. We would like to inform you that the result of your application is: **Rejected**

We have a tool that uses algorithms to assess credit card applications. The tool doesn't make decisions on its own and the application would not be rejected by the tool alone, but it helps our credit card officers find applications that might have a higher risk of not being repaid.

How the algorithm helped us make this decision:

We use information about you to see how your application compares to other people's data in our system. This helps us understand your situation better. The table below shows information about you that the algorithm considered for your application, where the information came from, how important that information is to the decision, and how it compares to the average of past applicants. The decision rule explains the basis for this decision.

Information type	Your application	Average for past applicants	Importance of information	Effect of information	Source of information
Debt collection accounts opened against you in last 24 months	0	0.10	Most important	Increased your likelihood of approval	Credit reporting agencies
Late payments or overdue accounts in last 24 months	1	0.14	Very important	Decreased your likelihood of approval	Credit reporting agencies
Percentage of credit that you've already paid off	2%	47%	Very important	Decreased your likelihood of approval	Credit reporting agencies
Percentage of credit limit that you're using now	80%	38%	Important	Decreased your likelihood of approval	Credit reporting agencies

The following decision rule was applied:

If debt collection accounts opened against you in last 24 months is less than 0.5 then: If late payments or overdue accounts in last 24 months is greater than 0.5 and percentage of credit that you've already paid off is less than or equal to 2% and percentage of credit that you're using now is greater than 78% then Reject the application

If you believe there is something wrong with how the algorithm assisted us in making the decision, **you can challenge the decision**. Please note that we can only reconsider the decision if you provide a **valid, appropriate reason** for why

Experimental task and questions

In this section, we provide the judgement, comprehension, and attitudinal questions used in the experiment. The wording is as it appeared in the experiment.

Judgement Task

For this task, please assume that you have applied for a credit card and that the profile details describe you and your current situation. You will need to review your profile and the result of your application. Then you will be asked whether you'd like to accept or challenge the decision.

Please review your profile below – you should assume the information is true.

Applicant profile – 'Your Profile'

Now please review the message you have received from the credit provider in regard to your application.

Application outcome and explanation genre

Please select below what you would like to do.

- I accept the decision
- I challenge the decision
- I don't know

[If participant selects 'I accept the decision]

Please provide your reasoning for accepting the decision:

- The algorithm is using the right data and has made an appropriate decision
- I don't trust the algorithm but in this case I agree with the decision
- I trust the algorithm-assisted decision
- I don't know
- Other (write in)

[If participant selects 'I challenge the decision']

Please provide your reasoning for challenging the decision:

- I would like to challenge the decision because the algorithm is not considering a piece of information that is important for the decision
- I would like to challenge the decision because the algorithm is over-relying on one piece of information to be important for the decision and not considering other important pieces of information
- I would like to challenge the decision because my information has been entered into the algorithm incorrectly

- I don't know
- Other (write in)

[If participant selects 'I don't know']

Please provide your reasoning for saying you're not sure:

- I don't feel confident to answer this
- I don't understand the decision
- I don't know
- I'm undecided whether the decision is right or not
- I don't trust the algorithm, but I don't know which decision I would make

Comprehension questions

Participants were shown Scenario 2, where the algorithm correctly rejected the application. Participants were told to assume that the decision was correct and were asked to use this Scenario to answer the comprehension questions

CQ1:

Which of the following best describes how the credit application decisions is made?

Options:

- The algorithm compares the applicant's profile with similar profiles and automatically accepts or rejects based on that
- The algorithm looks at the applicant's profile only and automatically accepts or rejects based on that
- The algorithm compares the applicant's profile with similar profiles and flags any high risk profiles for manual review
- The algorithm looks at the applicant's profile and flags high risk profiles for manual review
- Don't know

CQ2:

Which of the following best describes how annual income influences this algorithm?

Options:

- Less annual income increases the likelihood of approval
- Annual income does not affect the likelihood of approval
- More annual income increases the likelihood of approval
- Don't know

CQ3:

If all other features of your profile are kept the same, which of the following changes on your profile is most likely to make your application get accepted?

Options:

- Having an above average annual income
- Having a below average number of debt collection accounts opened per last 24 months
- Having a below average percentage of credit that you've already paid off
- Having an above average percentage of credit that you've already paid off
- Don't know

Attitudinal questions

Your final task is to answer the following survey questions honestly and to your best ability

AQ1:

How important do you think it is to be provided this information about how the algorithm-assisted decision is made?

- Not at all important
- Slightly unimportant
- Neither unimportant nor important
- Slightly important
- Very important

AQ2:

How helpful do you think it is to have information about how the algorithm-assisted decision is made?

- Not at all helpful
- Slightly unhelpful
- Neither helpful nor unhelpful
- Slightly helpful
- Very helpful

AQ3:

Do you agree that you had enough information about how the algorithm-assisted decision was made?

- Strongly disagree
- Somewhat disagree
- Neither agree nor disagree

- Somewhat agree
- Strongly agree

AQ4:

How confident would you feel to disagree with an algorithm-assisted decision if you thought it was wrong?

- Not confident at all
- Somewhat unconfident
- Neither confident nor unconfident
- Somewhat confident
- Very confident

Regression results

Table 1. Primary Analysis. The effect of explainability genre on performance on judgement task.

	Judgement of AI-assisted decision:	
	Proportion of scenarios judged correctly	
	(1)	(2)
Feature-based explanation	-0.023*** (0.005)	-0.026*** (0.008)
Combination date centric/features based	-0.033*** (0.006)	-0.037*** (0.008)
Combination data centric + features based + rules-based explanation	-0.074***	-0.075*** (0.009)
Sex: Male		0.003 (0.010)
Sex: Prefer not to say		-0.345*** (0.021)
Age: 25-34		0.017 (0.073)
Age: 35-44		0.029*** (0.004)
Age: 45-54		0.033*** (0.078)
Age: 55-64		0.033*** (0.005)
Age: 65-74		0.030* (0.005)
Age: 75+		0.023 (0.005)
Age: Prefer not to say		-0.030
Constant: Data-centric explanation	0.821*** (0.005)	0.804*** (0.008)
Observations	8,860	8,860
R^2	0.021	0.106
Adjusted R^2	0.021	0.105
Residual Std. Error	0.182 (df = 8856)	0.174 (df = 8847)
F Statistic	63.887*** (df = 3; 8856)	87.351*** (df = 12; 8847)

Note: *p<0.05; **p<0.01; ***p<0.001

Models 1 displays the results of just the treatment variables impact on the outcome.

Models 2 displays the results of the model with covariates to increase statistical power. The purpose of covariate inclusion is not to interpret their coefficients.

Both models use the Bonferroni adjusted p-values.

Table 2. Secondary Analysis. The effect of explainability genre on performance on Scenario 3 (incorrect rejection – data input error).

	Likelihood of correctly judging that an algorithm-assisted decision is incorrect	
	Scenario 3: 'Incorrect prediction due to data input' error	
	(1)	(2)
Feature-based explanation	0.029*** (0.007)	0.024** (0.012)
Combination date centric/features based	0.028*** (0.008)	0.024** (0.012)
Combination data centric + features based + rules-based explanation	0.005	0.004 (0.013)
Sex: Male		-0.014* (0.015)
Sex: Prefer not to say		-0.579*** (0.028)
Age: 25-34		0.037*** (0.012)
Age: 35-44		0.040*** (0.005)
Age: 45-54		0.050*** (0.040)
Age: 55-64		0.048*** (0.007)
Age: 65-74		0.047** (0.007)
Age: 75+		0.036 (0.008)
Age: Prefer not to say		0.096
Constant: Data-centric explanation	0.925*** (0.007)	0.904*** (0.012)
Observations	8,860	8,860
R^2	0.003	0.088
Adjusted R^2	0.003	0.087
Residual Std. Error	0.236 (df = 8856)	0.226 (df = 8847)
F Statistic	8.896*** (df = 3; 8856)	71.104*** (df = 12; 8847)

Note: *p<0.05; **p<0.01; ***p<0.001

Models 1 displays the results of just the treatment variables impact on the outcome.

Models 2 displays the results of the model with covariates to increase statistical power. The purpose of covariate inclusion is not to interpret their coefficients.

Both models use the Bonferroni adjusted p-values.

Table 3. Secondary Analysis. The effect of explainability genre on performance on Scenario 4 (incorrect rejection – over-reliance on one feature).

	Likelihood of correctly judging that an algorithm-assisted decision is incorrect	
	Scenario 4: 'Overreliance on one feature' error	
	(1)	(2)
Feature-based explanation	-0.053*** (0.013)	-0.057*** (0.019)
Combination date centric/features based	-0.059*** (0.014)	-0.062*** (0.021)
Combination data centric + features based + rules-based explanation	-0.179***	-0.179*** (0.022)
Sex: Male		-0.016 (0.027)
Sex: Prefer not to say		-0.413 (0.052)
Age: 25-34		-0.018 (0.186)
Age: 35-44		-0.025 (0.010)
Age: 45-54		-0.033 (0.189)
Age: 55-64		-0.041 (0.013)
Age: 65-74		-0.026 (0.013)
Age: 75+		0.027 (0.014)
Age: Prefer not to say		0.009
Constant: Data-centric explanation	0.768*** (0.013)	0.808*** (0.019)
Observations	8,860	8,860
R^2	0.020	0.032
Adjusted R^2	0.020	0.031
Residual Std. Error	0.456 (df = 8856)	0.453 (df = 8847)
F Statistic	60.121*** (df = 3; 8856)	24.429*** (df = 12; 8847)

Note: *p<0.05; **p<0.01; ***p<0.001

Models 1 displays the results of just the treatment variables impact on the outcome.

Models 2 displays the results of the model with covariates to increase statistical power. The purpose of covariate inclusion is not to interpret their coefficients.

Both models use the Bonferroni adjusted p-values.

Table 4. The effect of explainability genre on performance on Scenario 5 (incorrect rejection - failure to consider relevant features).

	Likelihood of correctly judging that an algorithm-assisted decision is incorrect	
	Scenario 5: 'Failure to consider relevant feature' error	
	(1)	(2)
Feature-based explanation	-0.113*** (0.015)	-0.114*** (0.021)
Combination date centric/features based	-0.146*** (0.015)	-0.149*** (0.022)
Combination data centric + features based + rules-based explanation	-0.198***	-0.199*** (0.024)
Sex: Male		0.057*** (0.029)
Sex: Prefer not to say		-0.009 (0.060)
Age: 25-34		0.055* (0.208)
Age: 35-44		0.093*** (0.010)
Age: 45-54		0.093*** (0.210)
Age: 55-64		0.123*** (0.015)
Age: 65-74		0.079* (0.015)
Age: 75+		0.121 (0.014)
Age: Prefer not to say		-0.113
Constant: Data-centric explanation	0.513*** (0.015)	0.413*** (0.020)
Observations	8,860	8,860
R^2	0.022	0.033
Adjusted R^2	0.021	0.031
Residual Std. Error	0.484 (df = 8856)	0.482 (df = 8847)
F Statistic	65.567*** (df = 3; 8856)	24.846*** (df = 12; 8847)

Note: *p<0.05; **p<0.01; ***p<0.001

Models 1 displays the results of just the treatment variables impact on the outcome.

Models 2 displays the results of the model with covariates to increase statistical power. The purpose of covariate inclusion is not to interpret their coefficients.

Both models use the Bonferroni adjusted p-values.

Table 5. The effect of explainability genre on performance on Scenario 1 (correct acceptance).

	Likelihood of correctly judging that an algorithm-assisted decision is correct	
	Scenario 1: Correct Acceptance	
	(1)	(2)
Feature-based explanation	0.010 (0.005)	0.006 (0.008)
Combination date centric/features based	0.009 (0.006)	0.005 (0.008)
Combination data centric + features based + rules-based explanation	0.003	0.002 (0.008)
Sex: Male		-0.007 (0.010)
Sex: Prefer not to say		-0.495*** (0.032)
Age: 25-34		0.004 (0.007)
Age: 35-44		0.016 (0.003)
Age: 45-54		0.022* (0.040)
Age: 55-64		0.015 (0.005)
Age: 65-74		0.012 (0.005)
Age: 75+		-0.045 (0.005)
Age: Prefer not to say		0.036
Constant: Data-centric explanation	0.963*** (0.005)	0.965*** (0.008)
Observations	8,860	8,860
R^2	0.001	0.127
Adjusted R^2	0.0002	0.126
Residual Std. Error	0.174 (df = 8856)	0.163 (df = 8847)
F Statistic	1.628 (df = 3; 8856)	107.008*** (df = 12; 8847)

Note: *p<0.05; **p<0.01; ***p<0.001

Models 1 displays the results of just the treatment variables impact on the outcome.

Models 2 displays the results of the model with covariates to increase statistical power. The purpose of covariate inclusion is not to interpret their coefficients.

Both models use the Bonferroni adjusted p-values.

Table 6. The effect of explainability genre on performance on Scenario 2 (correct rejection).

	Likelihood of correctly judging that an algorithm-assisted decision is correct	
	Scenario 2: Correct Rejection	
	(1)	(2)
Feature-based explanation	0.014 (0.007)	0.010 (0.011)
Combination date centric/features based	0.001 (0.007)	-0.002 (0.011)
Combination data centric + features based + rules-based explanation	-0.001	-0.002 (0.012)
Sex: Male		-0.002 (0.013)
Sex: Prefer not to say		-0.227 (0.035)
Age: 25-34		0.009 (0.219)
Age: 35-44		0.021 (0.005)
Age: 45-54		0.031* (0.222)
Age: 55-64		0.022 (0.007)
Age: 65-74		0.040* (0.007)
Age: 75+		-0.022 (0.007)
Age: Prefer not to say		-0.179
Constant: Data-centric explanation	0.936*** (0.007)	0.928*** (0.011)
Observations	8,860	8,860
R^2	0.001	0.057
Adjusted R^2	0.0003	0.055
Residual Std. Error	0.239 (df = 8856)	0.233 (df = 8847)
F Statistic	1.837 (df = 3; 8856)	44.297*** (df = 12; 8847)

Note: *p<0.05; **p<0.01; ***p<0.001

Models 1 displays the results of just the treatment variables impact on the outcome.

Models 2 displays the results of the model with covariates to increase statistical power. The purpose of covariate inclusion is not to interpret their coefficients.

Both models use the Bonferroni adjusted p-values.

Table 7. The effect of explainability genre on comprehension of basic information about how the algorithm is used.

	Effect of AI explainability genre on comprehension	
	Comprehension of basic information about how the algorithm is used	
	(1)	(2)
Feature-based explanation	-0.044** (0.015)	-0.046** (0.022)
Combination date centric/features based	-0.009 (0.015)	-0.010 (0.023)
Combination data centric + features based + rules-based explanation	-0.119***	-0.119*** (0.025)
Sex: Male		-0.021 (0.030)
Sex: Prefer not to say		-0.595* (0.055)
Age: 25-34		-0.028 (0.214)
Age: 35-44		-0.030 (0.011)
Age: 45-54		-0.033 (0.216)
Age: 55-64		-0.010 (0.015)
Age: 65-74		-0.043 (0.015)
Age: 75+		-0.158* (0.015)
Age: Prefer not to say		0.300
Constant: Data-centric explanation	0.458*** (0.015)	0.501*** (0.021)
Observations	8,860	8,860
R^2	0.009	0.015
Adjusted R^2	0.009	0.014
Residual Std. Error	0.491 (df = 8856)	0.489 (df = 8847)
F Statistic	26.614*** (df = 3; 8856)	11.447*** (df = 12; 8847)
Note:		*p<0.05; **p<0.01; ***p<0.001
		Models 1 displays the results of just the treatment variables impact on the outcome.
		Models 2 displays the results of the model with covariates to increase statistical power. The purpose of covariate inclusion is not to interpret their coefficients.
		Both models use the Bonferroni adjusted p-values.

Table 8. The effect of explainability genre on comprehension of features directionality information.

	Effect of AI explainability genre on comprehension	
	Comprehension of features directionality information	
	(1)	(2)
Feature-based explanation	0.029* (0.010)	0.024* (0.014)
Combination date centric/features based	0.014 (0.010)	0.010 (0.015)
Combination data centric + features based + rules-based explanation	-0.006	-0.007 (0.016)
Sex: Male		-0.011 (0.021)
Sex: Prefer not to say		-0.619*** (0.041)
Age: 25-34		0.007 (0.014)
Age: 35-44		0.005 (0.007)
Age: 45-54		0.003 (0.040)
Age: 55-64		0.004 (0.010)
Age: 65-74		-0.017 (0.010)
Age: 75+		-0.013 (0.010)
Age: Prefer not to say		0.124
Constant: Data-centric explanation	0.865*** (0.010)	0.878*** (0.014)
Observations	8,860	8,860
R^2	0.002	0.040
Adjusted R^2	0.001	0.039
Residual Std. Error	0.331 (df = 8856)	0.325 (df = 8847)
F Statistic	4.762** (df = 3; 8856)	30.929*** (df = 12; 8847)

Note: *p<0.05; **p<0.01; ***p<0.001

Models 1 displays the results of just the treatment variables impact on the outcome.

Models 2 displays the results of the model with covariates to increase statistical power. The purpose of covariate inclusion is not to interpret their coefficients.

Both models use the Bonferroni adjusted p-values.

Table 9. The effect of explainability genre on comprehension of features importance information.

	Effect of AI explainability genre on comprehension	
	Comprehension of features importance information	
	(1)	(2)
Feature-based explanation	0.187*** (0.013)	0.183*** (0.018)
Combination date centric/features based	0.160*** (0.013)	0.156*** (0.019)
Combination data centric + features based + rules-based explanation	0.179***	0.178*** (0.020)
Sex: Male		-0.005 (0.024)
Sex: Prefer not to say		-0.696*** (0.049)
Age: 25-34		0.054** (0.048)
Age: 35-44		0.085*** (0.008)
Age: 45-54		0.084*** (0.059)
Age: 55-64		0.096*** (0.012)
Age: 65-74		0.092*** (0.013)
Age: 75+		0.059 (0.012)
Age: Prefer not to say		0.295
Constant: Data-centric explanation	0.675*** (0.012)	0.617*** (0.018)
Observations	8,860	8,860
R^2	0.037	0.066
Adjusted R^2	0.037	0.064
Residual Std. Error	0.388 (df = 8856)	0.382 (df = 8847)
F Statistic	113.186*** (df = 3; 8856)	51.686*** (df = 12; 8847)

Note: *p<0.05; **p<0.01; ***p<0.001

Models 1 displays the results of just the treatment variables impact on the outcome.

Models 2 displays the results of the model with covariates to increase statistical power. The purpose of covariate inclusion is not to interpret their coefficients.

Both models use the Bonferroni adjusted p-values.

Table 10. The effect of explainability genre on the likelihood of reporting that the information was important.

	Effect of AI explainability genre on attitudes	
	Likelihood of reporting that the information provided was important	
	(1)	(2)
Feature-based explanation	0.021* (0.008)	0.016 (0.011)
Combination date centric/features based	0.008 (0.007)	0.004 (0.012)
Combination data centric + features based + rules-based explanation	0.015	0.014 (0.012)
Sex: Male		-0.019*** (0.013)
Sex: Prefer not to say		-0.597*** (0.039)
Age: 25-34		0.027* (0.012)
Age: 35-44		0.024* (0.005)
Age: 45-54		0.036** (0.040)
Age: 55-64		0.035** (0.007)
Age: 65-74		0.047** (0.007)
Age: 75+		-0.042 (0.007)
Age: Prefer not to say		0.078
Constant: Data-centric explanation	0.929*** (0.007)	0.923*** (0.011)
Observations	8,860	8,860
R^2	0.001	0.093
Adjusted R^2	0.001	0.092
Residual Std. Error	0.237 (df = 8856)	0.226 (df = 8847)
F Statistic	3.151* (df = 3; 8856)	75.403*** (df = 12; 8847)

Note: *p<0.05; **p<0.01; ***p<0.001

Models 1 displays the results of just the treatment variables' impact on the outcome.

Model 2 displays the results of the model with covariates to increase statistical power. The purpose of covariate inclusion is not to interpret their coefficients.

Both models use the Bonferroni adjusted p-values.

Table 11. The effect of explainability genre on the likelihood of reporting that the information was sufficient.

	Effect of AI explainability genre on attitudes	
	Likelihood of reporting that there was enough information provided	
	(1)	(2)
Feature-based explanation	0.117*** (0.014)	0.114*** (0.020)
Combination date centric/features based	0.134*** (0.014)	0.131*** (0.021)
Combination data centric + features based + rules-based explanation	0.161***	0.162*** (0.022)
Sex: Male		-0.022 (0.026)
Sex: Prefer not to say		-0.728** (0.054)
Age: 25-34		0.053* (0.041)
Age: 35-44		0.095*** (0.009)
Age: 45-54		0.128*** (0.053)
Age: 55-64		0.133*** (0.014)
Age: 65-74		0.156*** (0.014)
Age: 75+		0.037 (0.013)
Age: Prefer not to say		0.386
Constant: Data-centric explanation	0.624*** (0.014)	0.556*** (0.020)
Observations	8,860	8,860
R^2	0.019	0.044
Adjusted R^2	0.019	0.042
Residual Std. Error	0.441 (df = 8856)	0.435 (df = 8847)
F Statistic	56.692*** (df = 3; 8856)	33.641*** (df = 12; 8847)

Note: *p<0.05; **p<0.01; ***p<0.001

Models 1 displays the results of just the treatment variables' impact on the outcome.

Model 2 displays the results of the model with covariates to increase statistical power. The purpose of covariate inclusion is not to interpret their coefficients.

Both models use the Bonferroni adjusted p-values.

Table 12. The effect of explainability genre on the likelihood of reporting that the information was helpful.

	Effect of AI explainability genre on attitudes	
	Likelihood of reporting that the information provided was helpful	
	(1)	(2)
Feature-based explanation	0.024** (0.008)	0.019* (0.012)
Combination date centric/features based	0.010 (0.008)	0.006 (0.012)
Combination data centric + features based + rules-based explanation	0.012	0.011 (0.013)
Sex: Male		-0.017** (0.014)
Sex: Prefer not to say		-0.609*** (0.039)
Age: 25-34		0.021 (0.012)
Age: 35-44		0.032** (0.005)
Age: 45-54		0.040*** (0.040)
Age: 55-64		0.039** (0.007)
Age: 65-74		0.049** (0.007)
Age: 75+		-0.035 (0.007)
Age: Prefer not to say		0.087
Constant: Data-centric explanation	0.923*** (0.007)	0.915*** (0.012)
Observations	8,860	8,860
R^2	0.001	0.088
Adjusted R^2	0.001	0.086
Residual Std. Error	0.248 (df = 8856)	0.237 (df = 8847)
F Statistic	3.397* (df = 3; 8856)	70.779*** (df = 12; 8847)

Note: *p<0.05; **p<0.01; ***p<0.001

Models 1 displays the results of just the treatment variables' impact on the outcome.

Model 2 displays the results of the model with covariates to increase statistical power. The purpose of covariate inclusion is not to interpret their coefficients.

Both models use the Bonferroni adjusted p-values.

Table 13. The effect of explainability genre on the likelihood of reporting confidence in ability to challenge decision.

	Effect of AI explainability genre on attitudes	
	Likelihood of reporting confidence in ability to agree/disagree with decision	
	(1)	(2)
Feature-based explanation	0.062*** (0.013)	0.059*** (0.019)
Combination date centric/features based	0.047*** (0.013)	0.043** (0.020)
Combination data centric + features based + rules-based explanation	0.042**	0.041** (0.021)
Sex: Male		0.028** (0.025)
Sex: Prefer not to say		-0.421 (0.040)
Age: 25-34		0.049* (0.231)
Age: 35-44		0.045* (0.009)
Age: 45-54		0.053* (0.234)
Age: 55-64		0.089*** (0.013)
Age: 65-74		0.090** (0.013)
Age: 75+		0.177** (0.013)
Age: Prefer not to say		0.045
Constant: Data-centric explanation	0.723*** (0.013)	0.667*** (0.019)
Observations	8,860	8,860
R^2	0.003	0.026
Adjusted R^2	0.003	0.024
Residual Std. Error	0.426 (df = 8856)	0.421 (df = 8847)
F Statistic	8.573*** (df = 3; 8856)	19.375*** (df = 12; 8847)
Note:		*p<0.05; **p<0.01; ***p<0.001
		Models 1 displays the results of just the treatment variables' impact on the outcome.
		Model 2 displays the results of the model with covariates to increase statistical power. The purpose of covariate inclusion is not to interpret their coefficients.

Both models use the Bonferroni adjusted p-values.

Table 14. The effect of adding a decision rule on performance on the judgement task.

	Judgement of AI-assisted decision	
	Proportion of judgements correctly accepted/challenged	
	(1)	(2)
Combination data centric + features based + rules-based explanation	-0.041***	-0.038*** (0.011)
Sex: Male		0.001 (0.011)
Sex: Prefer not to say		-0.245 (0.013)
Age: 25-34		0.023 (0.015)
Age: 35-44		0.026* (0.033)
Age: 45-54		0.038** (0.010)
Age: 55-64		0.031* (0.005)
Age: 65-74		0.037* (0.042)
Age: 75+		0.026 (0.005)
Age: Prefer not to say		-0.128
Constant: Combination data centric/features based	0.788*** (0.005)	0.766*** (0.011)
Observations	4,447	4,447
R^2	0.012	0.096
Adjusted R^2	0.012	0.094
Residual Std. Error	0.182 (df = 4445)	0.174 (df = 4436)
F Statistic	55.881*** (df = 1; 4445)	46.955*** (df = 10; 4436)
Note:		*p<0.05; **p<0.01; ***p<0.001
		Model 1 displays the results of just the treatment variables' impact on the outcome.
		Model 2 displays the results of the model with covariates to increase statistical power. The purpose of covariate inclusion is not to interpret their coefficients.
		Both models use the Bonferroni adjusted p-values.

Table 10. The effect of adding a decision rule on performance on Scenario 3 (incorrect rejection - data input error)

	Likelihood of correctly judging that an algorithm-assisted decision is incorrect	
	Scenario: 'Incorrect prediction due to data input' error	
	(1)	(2)
Combination data centric + features based + rules-based explanation	-0.023**	-0.020** (0.016)
Sex: Male		-0.009 (0.017)
Sex: Prefer not to say		-0.561* (0.018)
Age: 25-34		0.030 (0.021)
Age: 35-44		0.031 (0.033)
Age: 45-54		0.040* (0.016)
Age: 55-64		0.037 (0.007)
Age: 65-74		0.038 (0.057)
Age: 75+		0.052 (0.007)
Age: Prefer not to say		0.086
Constant: Combination data centric/features based	0.953*** (0.007)	0.933*** (0.016)
Observations	4,447	4,447
R^2	0.002	0.081
Adjusted R^2	0.002	0.079
Residual Std. Error	0.235 (df = 4445)	0.226 (df = 4436)
F Statistic	10.582** (df = 1; 4445)	39.000*** (df = 10; 4436)

Note: *p<0.05; **p<0.01; ***p<0.001

Model 1 displays the results of just the treatment variables' impact on the outcome.

Model 2 displays the results of the model with covariates to increase statistical power. The purpose of covariate inclusion is not to interpret their coefficients.

Both models use the Bonferroni adjusted p-values.

Table 16. The effect of adding a decision-rule on performance on Scenario 4 (incorrect rejection - overreliance on one feature).

	Likelihood of correctly judging that an algorithm-assisted decision is incorrect	
	Scenario: 'Overreliance on one feature' error	
	(1)	(2)
Combination data centric + features based + rules-based explanation	-0.119***	-0.117*** (0.029)
Sex: Male		-0.021 (0.031)
Sex: Prefer not to say		0.253 (0.033)
Age: 25-34		0.012 (0.041)
Age: 35-44		-0.013 (0.081)
Age: 45-54		-0.011 (0.028)
Age: 55-64		-0.048 (0.014)
Age: 65-74		0.012 (0.053)
Age: 75+		0.069 (0.014)
Age: Prefer not to say		-0.613
Constant: Combination data centric/features based	0.709*** (0.014)	0.731*** (0.029)
Observations	4,447	4,447
R^2	0.016	0.027
Adjusted R^2	0.015	0.025
Residual Std. Error	0.473 (df = 4445)	0.471 (df = 4436)
F Statistic	70.757*** (df = 1; 4445)	12.233*** (df = 10; 4436)

Note: *p<0.05; **p<0.01; ***p<0.001

Model 1 displays the results of just the treatment variables' impact on the outcome.

Model 2 displays the results of the model with covariates to increase statistical power. The purpose of covariate inclusion is not to interpret their coefficients.

Both models use the Bonferroni adjusted p-values.

Table 17. The effect of adding a decision rule on performance on Scenario 5 (incorrect rejection - failure to consider relevant features).

	Likelihood of correctly judging that an algorithm-assisted decision is incorrect	
	Scenario: 'Failure to consider relevant feature' error	
	(1)	(2)
Combination data centric + features based + rules-based explanation	-0.053***	-0.051*** (0.028)
Sex: Male		0.034 (0.030)
Sex: Prefer not to say		0.110 (0.033)
Age: 25-34		0.051 (0.041)
Age: 35-44		0.071* (0.086)
Age: 45-54		0.089* (0.027)
Age: 55-64		0.100** (0.014)
Age: 65-74		0.044 (0.039)
Age: 75+		0.066 (0.014)
Age: Prefer not to say		-0.237
Constant: Combination data centric/features based	0.367*** (0.014)	0.288*** (0.028)
Observations	4,447	4,447
R^2	0.003	0.010
Adjusted R^2	0.003	0.008
Residual Std. Error	0.474 (df = 4445)	0.472 (df = 4436)
F Statistic	13.666*** (df = 1; 4445)	4.673*** (df = 10; 4436)

Note: *p<0.05; **p<0.01; ***p<0.001

Model 1 displays the results of just the treatment variables' impact on the outcome.

Model 2 displays the results of the model with covariates to increase statistical power. The purpose of covariate inclusion is not to interpret their coefficients.

Both models use the Bonferroni adjusted p-values.

Table 18. The effect of adding a decision rule on performance on Scenario 1 (correct acceptance).

	Likelihood of correctly judging that an algorithm-assisted decision is correct	
	In instances where Scenario 1 has been correctly accepted	
	(1)	(2)
Combination data centric + features based + rules-based explanation	-0.006	-0.003 (0.011)
Sex: Male		-0.005 (0.011)
Sex: Prefer not to say		-0.515** (0.012)
Age: 25-34		0.002 (0.012)
Age: 35-44		0.014 (0.042)
Age: 45-54		0.022 (0.011)
Age: 55-64		0.014 (0.005)
Age: 65-74		0.024 (0.057)
Age: 75+		-0.025 (0.005)
Age: Prefer not to say		0.032
Constant: Combination data centric/features based	0.972*** (0.005)	0.971*** (0.011)
Observations	4,447	4,447
R^2	0.0003	0.138
Adjusted R^2	0.0001	0.136
Residual Std. Error	0.173 (df = 4445)	0.161 (df = 4436)
F Statistic	1.480 (df = 1; 4445)	71.213*** (df = 10; 4436)

Note: *p<0.05; **p<0.01; ***p<0.001

Model 1 displays the results of just the treatment variables' impact on the outcome.

Model 2 displays the results of the model with covariates to increase statistical power. The purpose of covariate inclusion is not to interpret their coefficients.

Both models use the Bonferroni adjusted p-values.

Table 19. The effect of adding a decision rule on performance on Scenario 2 (correct rejection).

	Likelihood of correctly judging that an algorithm-assisted decision is correct	
	In instances where Scenario 2 has been correctly rejected	
	(1)	(2)
Combination data centric + features based + rules-based explanation	-0.002	0.0004 (0.017)
Sex: Male		0.006 (0.017)
Sex: Prefer not to say		-0.513 (0.018)
Age: 25-34		0.020 (0.018)
Age: 35-44		0.029 (0.058)
Age: 45-54		0.049** (0.016)
Age: 55-64		0.052** (0.007)
Age: 65-74		0.069** (0.057)
Age: 75+		-0.030 (0.007)
Age: Prefer not to say		0.091
Constant: Combination data centric/features based	0.937*** (0.007)	0.908*** (0.017)
Observations	4,447	4,447
R^2	0.00002	0.063
Adjusted R^2	-0.0002	0.061
Residual Std. Error	0.246 (df = 4445)	0.238 (df = 4436)
F Statistic	0.098 (df = 1; 4445)	29.952*** (df = 10; 4436)

Note: *p<0.05; **p<0.01; ***p<0.001

Model 1 displays the results of just the treatment variables' impact on the outcome.

Model 2 displays the results of the model with covariates to increase statistical power. The purpose of covariate inclusion is not to interpret their coefficients.

Both models use the Bonferroni adjusted p-values.

Table 20. The effect of adding a decision rule on comprehension of basic information about how the algorithm is used.

	Effect of AI explainability genre on comprehension	
	Comprehension of basic information about how the algorithm is used	
	(1)	(2)
Combination data centric + features based + rules-based explanation	-0.110***	-0.110*** (0.031)
Sex: Male		-0.014 (0.033)
Sex: Prefer not to say		-0.823 (0.035)
Age: 25-34		-0.051 (0.043)
Age: 35-44		-0.065 (0.081)
Age: 45-54		-0.071 (0.030)
Age: 55-64		-0.028 (0.015)
Age: 65-74		-0.084 (0.047)
Age: 75+		-0.132 (0.015)
Age: Prefer not to say		0.597
Constant: Combination data centric/features based	0.450*** (0.015)	0.513*** (0.030)
Observations	4,447	4,447
R^2	0.013	0.017
Adjusted R^2	0.012	0.015
Residual Std. Error	0.486 (df = 4445)	0.485 (df = 4436)
F Statistic	56.692*** (df = 1; 4445)	7.694*** (df = 10; 4436)

Note: *p<0.05; **p<0.01; ***p<0.001

Model 1 displays the results of just the treatment variables' impact on the outcome.

Model 2 displays the results of the model with covariates to increase statistical power. The purpose of covariate inclusion is not to interpret their coefficients.

Both models use the Bonferroni adjusted p-values.

Table 21. The effect of adding a decision rule on comprehension of features directionality information.

	Effect of AI explainability genre on comprehension	
	Comprehension of features directionality information	
	(1)	(2)
Combination data centric + features based + rules-based explanation	-0.020	-0.018 (0.021)
Sex: Male		-0.020 (0.022)
Sex: Prefer not to say		-0.600 (0.024)
Age: 25-34		0.020 (0.029)
Age: 35-44		-0.011 (0.058)
Age: 45-54		-0.004 (0.019)
Age: 55-64		-0.001 (0.010)
Age: 65-74		0.012 (0.057)
Age: 75+		0.010 (0.010)
Age: Prefer not to say		0.125
Constant: Combination data centric/features based	0.879^{***} (0.010)	0.893^{***} (0.020)
Observations	4,447	4,447
R^2	0.001	0.035
Adjusted R^2	0.001	0.033
Residual Std. Error	0.337 (df = 4445)	0.332 (df = 4436)
F Statistic	3.885^{*} (df = 1; 4445)	16.314^{***} (df = 10; 4436)

Note: *p<0.05; **p<0.01; ***p<0.001

Model 1 displays the results of just the treatment variables' impact on the outcome.

Model 2 displays the results of the model with covariates to increase statistical power. The purpose of covariate inclusion is not to interpret their coefficients.

Both models use the Bonferroni adjusted p-values.

Table 22. The effect of adding a decision rule on comprehension of features importance information.

	Effect of AI explainability genre on comprehension	
	Comprehension of features importance information	
	(1)	(2)
Combination data centric + features based + rules-based explanation	0.019	0.022 (0.024)
Sex: Male		-0.007 (0.025)
Sex: Prefer not to say		-0.609 (0.026)
Age: 25-34		0.032 (0.033)
Age: 35-44		0.043 (0.064)
Age: 45-54		0.059* (0.023)
Age: 55-64		0.073* (0.011)
Age: 65-74		0.043 (0.056)
Age: 75+		0.046 (0.011)
Age: Prefer not to say		0.177
Constant: Combination data centric/features based	0.834*** (0.011)	0.802*** (0.024)
Observations	4,447	4,447
R^2	0.001	0.032
Adjusted R^2	0.0004	0.029
Residual Std. Error	0.363 (df = 4445)	0.358 (df = 4436)
F Statistic	2.987 (df = 1; 4445)	14.487*** (df = 10; 4436)

Note: *p<0.05; **p<0.01; ***p<0.001

Model 1 displays the results of just the treatment variables' impact on the outcome.

Model 2 displays the results of the model with covariates to increase statistical power. The purpose of covariate inclusion is not to interpret their coefficients.

Both models use the Bonferroni adjusted p-values.

Exploratory Analysis: Participants' ability to challenge an error and identify the error type

We explored whether participants who identified incorrect decisions could also identify the decision error. Performance varied by explanation genre and error type.

Table 23 gives an overview of participants' ability to identify the correct error type having accurately judged each incorrect decision. In comparison to the other error types, participants were least likely to identify the algorithm's overreliance on one feature as the relevant error, doing so only 19% of the time. To note, we found that a large majority (79% on average across all treatment groups) of participants instead selected that the error was due to the algorithm's failure to consider relevant features. We attribute the poor performance here to the conceptual similarities between these error types rather than a complete lack of understanding as both error types relate to how the algorithm weights the importance of features in its decision.

Table 23. Participants' ability to challenge errors and identify the correct error types.

Treatment	Error type: Data input (Scenario 3)	Error type: Overreliance on one feature (Scenario 4)	Error type: Failure to consider relevant features (Scenario 5)
Data-centric (control)	69%	19%	61%
Features-based (comparison to control)	80% **(+11pp)**	12% **(-7pp)**	41% **(-20pp)**
Combination (comparison to control)	78% **(+9pp)**	15% **(-4pp)**	45% **(-16pp)**
Combination + rules (comparison to control)	79% **(+10pp)**	16% **(-3pp)**	44% **(-17pp)**

© Financial Conduct Authority 2020

12 Endeavour Square, London E20 1JN
Telephone: +44 (0)20 7066 1000 Website: www.fca.org.uk

All rights reserved