

Supplementary information for ‘Trust in artificial
intelligence is agent-specific and
multidimensional’

Scott Claessens¹, Jim A.C. Everett¹

¹School of Psychology, University of Kent, Canterbury, UK

2026-02-18

Supplementary Methods

Model specifications

We fitted the following Bayesian multilevel cumulative-link ordinal models to the data using the R package *brms*.

Baseline models

For the baseline intercept-only model below, we denote trust T as the outcome variable, but we repeated the model across all measures as separate outcome variables, dealing with any missing data using listwise deletion. For observations i from participants j responding to AI types k , the model is as follows:

$$\begin{aligned}T_i &\sim \text{OrderedLogit}(\phi_i, \bar{\alpha}) \\ \phi_i &= \alpha_{\text{ID}[j]} + \alpha_{\text{TYPE}[k]} \\ \alpha_{\text{ID}[j]} &\sim \text{Normal}(0, \sigma_{\text{ID}}) \\ \alpha_{\text{TYPE}[k]} &\sim \text{Normal}(0, \sigma_{\text{TYPE}}) \\ \sigma_{\text{ID}}, \sigma_{\text{TYPE}} &\sim \text{Exponential}(2) \\ \bar{\alpha}_c &\sim \text{Normal}(0, 1)\end{aligned}$$

where α_{ID} and α_{TYPE} are varying intercepts over participants and AI types, and $\bar{\alpha}$ is a vector of six cutpoints c .

Regression models

In regression models, we focus on trust as the outcome variable and include regression slopes for performance (P , a composite variable of reliable + competent) and morality (M , a composite variable of ethical + genuine). We allowed these slopes to vary across AI types. Due to MCAR missing data from the experimental design, we iterated this model over 20 imputed datasets using the *mice* package and pooled the resulting posterior distributions. The model is as follows:

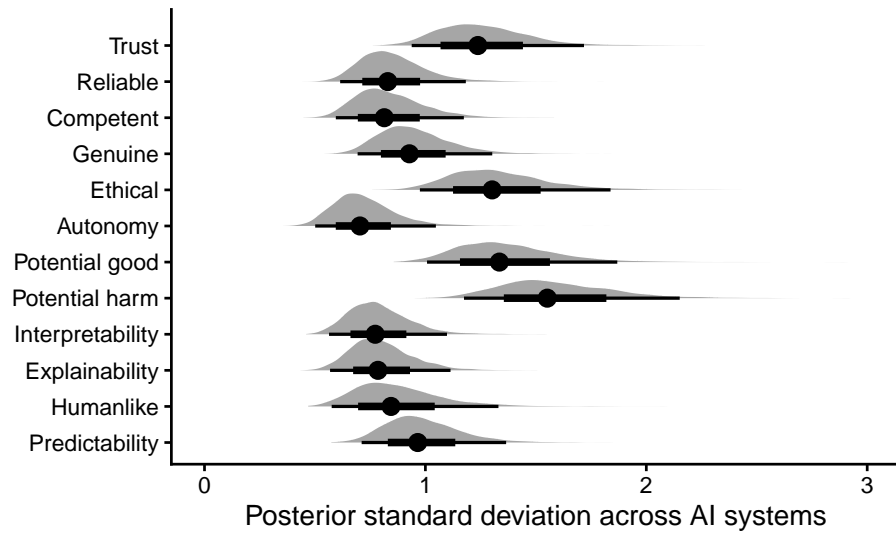
$$\begin{aligned}
T_i &\sim \text{OrderedLogit}(\phi_i, \bar{\alpha}) \\
\phi_i &= \alpha_{\text{ID}[j]} + \alpha_{\text{TYPE}[k]} + \beta_P P_i + \beta_M M_i \\
\beta_P &= \bar{\beta}_P + \beta_{P, \text{TYPE}[k]} \\
\beta_M &= \bar{\beta}_M + \beta_{M, \text{TYPE}[k]} \\
\alpha_{\text{ID}[j]} &\sim \text{Normal}(0, \sigma_{\text{ID}}) \\
\begin{bmatrix} \alpha_{\text{TYPE}[k]} \\ \beta_{P, \text{TYPE}[k]} \\ \beta_{M, \text{TYPE}[k]} \end{bmatrix} &\sim \text{MVNormal} \left(0, \text{R}, \begin{bmatrix} \sigma_{\alpha} \\ \sigma_{\beta_P} \\ \sigma_{\beta_M} \end{bmatrix} \right) \\
\sigma_{\text{ID}}, \sigma_{\alpha}, \sigma_{\beta_P}, \sigma_{\beta_M} &\sim \text{Exponential}(2) \\
\text{R} &\sim \text{LKJCorr}(2) \\
\bar{\beta}_P, \bar{\beta}_M &\sim \text{Normal}(0, 1) \\
\bar{\alpha}_c &\sim \text{Normal}(0, 1)
\end{aligned}$$

Moderation models

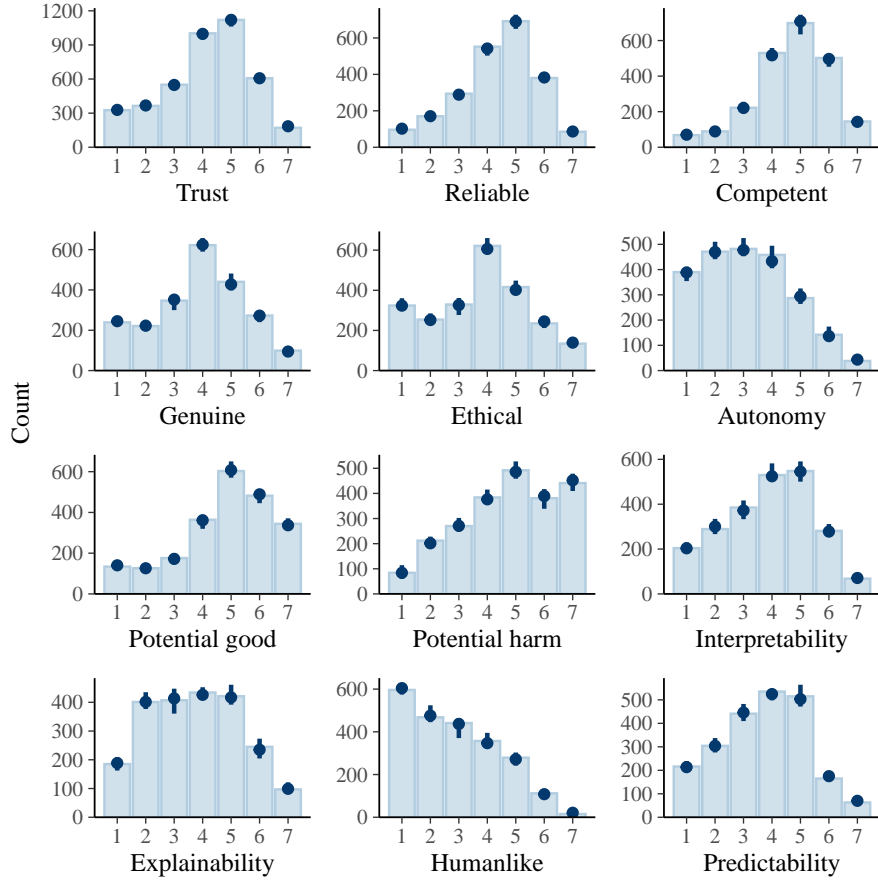
We then included additional variables as moderators of the effects of performance and morality on trust: autonomy, potential harm, interpretability, anthropomorphism, and predictability. We included these moderators in separate models, but below we denote the moderator variable generally as X . As before, we iterated the model over 20 imputed data sets.

$$\begin{aligned}
T_i &\sim \text{OrderedLogit}(\phi_i, \bar{\alpha}) \\
\phi_i &= \alpha_{\text{ID}[j]} + \alpha_{\text{TYPE}[k]} + \beta_P P_i + \beta_M M_i + \beta_X X_i + \beta_{PX} P_i X_i + \beta_{MX} M_i X_i \\
\beta_P &= \bar{\beta}_P + \beta_{P, \text{TYPE}[k]} \\
\beta_M &= \bar{\beta}_M + \beta_{M, \text{TYPE}[k]} \\
\beta_X &= \bar{\beta}_X + \beta_{X, \text{TYPE}[k]} \\
\beta_{PX} &= \bar{\beta}_{PX} + \beta_{PX, \text{TYPE}[k]} \\
\beta_{MX} &= \bar{\beta}_{MX} + \beta_{MX, \text{TYPE}[k]} \\
\alpha_{\text{ID}[j]} &\sim \text{Normal}(0, \sigma_{\text{ID}}) \\
\begin{bmatrix} \alpha_{\text{TYPE}[k]} \\ \beta_{P, \text{TYPE}[k]} \\ \beta_{M, \text{TYPE}[k]} \\ \beta_{X, \text{TYPE}[k]} \\ \beta_{PX, \text{TYPE}[k]} \\ \beta_{MX, \text{TYPE}[k]} \end{bmatrix} &\sim \text{MVNormal} \left(0, \text{R}, \begin{bmatrix} \sigma_{\alpha} \\ \sigma_{\beta_P} \\ \sigma_{\beta_M} \\ \sigma_{\beta_X} \\ \sigma_{\beta_{PX}} \\ \sigma_{\beta_{MX}} \end{bmatrix} \right) \\
\sigma_{\text{ID}}, \sigma_{\alpha}, \sigma_{\beta_P}, \sigma_{\beta_M}, \sigma_{\beta_X}, \sigma_{\beta_{PX}}, \sigma_{\beta_{MX}} &\sim \text{Exponential}(2) \\
\text{R} &\sim \text{LKJCorr}(2) \\
\bar{\beta}_P, \bar{\beta}_M, \bar{\beta}_X, \bar{\beta}_{PX}, \bar{\beta}_{MX} &\sim \text{Normal}(0, 1) \\
\bar{\alpha}_c &\sim \text{Normal}(0, 1)
\end{aligned}$$

Supplementary Figures



Supplementary Figure 1: Posterior distributions for standard deviation parameters across AI types. Greater standard deviation parameters mean that the measure varies more across different AI types. Parameters are on the log odds scale. Grey densities represent samples from posterior distributions, point and line ranges represent median estimates and 66% and 95% credible intervals.



Supplementary Figure 2: Posterior predictive checks for baseline models. Light blue bars are the observed data, dark blue point ranges are the posterior predictions (posterior means and 95% credible intervals).

Supplementary Tables

Supplementary Table 1: Descriptions of the AI types in the study.

AI Type	Description
General AI	Artificial intelligence (AI) refers to computer systems designed to perform tasks that typically require human intelligence. These systems can learn from data, recognize patterns, make decisions, and adapt over time—powering applications in fields like healthcare, finance, transportation, and more.
Skin cancer diagnosis apps	Skin cancer diagnosis apps use AI to assess skin lesions for signs of cancer. By analyzing images of moles or spots, these tools can detect patterns associated with conditions like melanoma and provide users with risk assessments, often recommending follow-up care with a dermatologist.
AI therapists	AI therapists are mental healthcare alternatives to human therapists. Offering personalized mental health support for users, AI therapists can tailor their support to match different specialties, including cognitive behavior therapy (CBT) and psychoanalysis.
Medical triage AI	Medical triage AI helps hospitals assess incoming patients and prioritize treatment based on the severity of their condition. It analyzes vital signs, symptoms, and medical history to support staff in deciding who needs immediate attention and how to allocate limited resources such as beds, equipment, or specialist care.
Facial recognition AI	Facial recognition AI analyzes images or video footage to identify individuals based on their facial features. In the justice system, it can be used to match suspects against criminal databases, assist in locating missing persons, or monitor public spaces for known individuals of interest.
Predictive policing algorithms	Predictive policing algorithms are used by police to predict crime. By drawing sophisticated inferences from past crime statistics, current weather, and recent news, the algorithms help police departments by informing them where crimes are likely to occur.
Predictive sentencing algorithms	Predictive sentencing algorithms are used in courts to determine what punishment to give convicts. Using facts about the convict, including their demographics, what crime they committed, and how they acted in court, predictive sentencing algorithms inform judges' sentencing by predicting the likelihood that the convicted person will commit another crime.
ChatGPT	ChatGPT is an advanced and flexible AI chatbot designed by the American company OpenAI. It is capable of solving complex logical problems, passing medical exams, producing functional computer code, and accurately identifying the contents of photos.
DeepSeek	DeepSeek is an advanced and flexible AI chatbot designed in China. It is capable of solving complex logical problems, passing medical exams, producing functional computer code, and accurately identifying the contents of photos.
Audio transcription AI	Audio transcription AI uses advanced algorithms to convert recorded audio into text. It can reliably transcribe recorded conversations and can create live captioning to television shows or videos that don't already have captions.

Instagram filters	Instagram filters automatically alter the images and videos taken on a phone camera. They can make a user appear younger, older, give the user a moustache, or make it look like they are a cartoon character.
DALL-E	DALL-E generates images based on user prompts. It can copy a wide variety of styles, such as photographs, oil paintings, and cartoons, and what it creates is largely limited by the imagination of the user's prompts.
Military cybersecurity AI	Military cybersecurity AI monitors military networks for signs of cyber threats, such as unauthorized access or unusual activity. It can detect and respond to attacks in real time, helping to protect sensitive data, communication systems, and critical infrastructure from hacking or sabotage.
Autonomous killer drones	Autonomous killer drones identify and kill targets on their own. They are able to identify potential targets, evaluate the likelihood they are in fact an enemy combatant, and if the likelihood is high enough, act on their own to kill the target without human intervention.
Robot soldiers	Robot soldiers are autonomous or semi-autonomous machines designed to perform combat roles on the battlefield. They can navigate complex environments, identify targets, and engage in military operations, either independently or under human control, reducing the need to place human soldiers in direct danger.
Robot vacuums	Robot vacuums automatically clean the floors of homes. They are able to keep track of where they have already vacuumed and where they need to vacuum, planning their route through and around rooms to make sure they clean everything.
Apple's Siri	Apple's Siri is an AI assistant on the iPhone. Siri responds to the user's verbal instructions by doing things like making calendar appointments, writing messages, and searching the internet for answers to questions.
Google Maps AI	Google Maps AI analyzes real-time traffic data to help drivers avoid congestion, accidents, or road closures. It can suggest faster or safer routes based on current conditions and may continuously update directions as the situation on the road changes.
Air traffic control AI	Air traffic control AI is used to help manage the movement of aircraft in busy airspace and at airports. It can monitor flight paths, predict potential conflicts, and suggest adjustments to improve safety and efficiency. In some systems, it may assist or automate decisions normally made by human air traffic controllers.
Self-driving cars	Self-driving cars are capable of driving themselves as if there was a human driver. Without human input, self-driving cars can figure out how to navigate to the destination, safely dealing with other traffic, pedestrians, and road signs.
AI superintelligence	AI superintelligence refers to a hypothetical future AI system that greatly surpasses human intelligence across all domains. Such a system would be capable of independent reasoning, long-term planning, and potentially reshaping entire aspects of society and science. It would operate with far greater speed, accuracy, and foresight than any human, and might act with limited or no human oversight.
