

Ethics application

Should AI Make Moral Decisions?

Ethics ID: **2025175881120310129**

Expiry date: 25-09-2027

Covers multiple related studies that will use the same methods:

No.

Summary

Artificial intelligence is increasingly advising humans on moral matters. Previous research has suggested that people are averse to machines making moral decisions. However, this work has focused on US participants, only a few domains (e.g., autonomous vehicles), and has not distinguished between AI in advisory and decision-making roles.

In this pilot study, we will trial an online survey in which we ask people whether they think AI should be used to advise on and make moral decisions across a wide-range of domains. The plan is to test whether this design would be suitable for deployment on an online data collection website.

Full description of study methodology and procedures

We will conduct a quantitative survey on the data collection platform Prolific (<https://www.prolific.com/>). The survey is written using Qualtrics. Prolific participants will be paid commensurate with an hourly wage of £10.00. The expected duration is approximately 5 minutes.

Participants will first see an information sheet and will be asked for their informed consent. To ensure participants are being attentive, they will be given an attention check at this stage. They will also be asked to provide their age and gender.

The main task is split into two sections. In each section, the participant will be shown 15 randomly chosen "cards" from the full set of 60. Each card asks the participant whether they think AI should be used to make different moral decisions in different domains (yes / no). The first section will focus on the AI "deciding" what to do without human input, and the second section will focus on the AI "advising" on what to do (randomly counterbalanced). The full set of cards is as follows:

Transportation:

- Should AI decide / advise on which pedestrian to hit in a car accident
- Should AI decide / advise on whether to prioritise the safety of passengers or pedestrians in a car accident
- Should AI decide / advise on which vehicles get priority at traffic lights and intersections
- Should AI decide / advise on how to allocate ridesharing vehicles to customers
- Should AI decide / advise on which areas should have more traffic enforcement
- Should AI decide / advise on whether to override a human driver's commands for safety reasons

Justice:

- Should AI decide / advise on whether to release convicted criminals from jail
- Should AI decide / advise on whether to grant parole to prisoners
- Should AI decide / advise on the length of jail sentences for criminals
- Should AI decide / advise on who police should stop and search
- Should AI decide / advise on which personalized rehabilitation programs to offer to offenders
- Should AI decide / advise on which neighborhoods should have more police patrols

Healthcare:

Should AI decide / advise on whether to perform a surgery

Should AI decide / advise on who to allocate ventilators to

Should AI decide / advise on which personalised treatments to recommend to patients

Should AI decide / advise on who should receive mental health support

Should AI decide / advise on who gets an organ transplant

Should AI decide / advise on when to withdraw life support

Business:

Should AI decide / advise on who to hire for a job

Should AI decide / advise on who to invite to a job interview

Should AI decide / advise on the salary to offer prospective employees

Should AI decide / advise on who to lay off during company downsizing

Should AI decide / advise on who gets promoted within a company

Should AI decide / advise on whether to extend an employee's contract

Military:

Should AI decide / advise on where to fire missiles in military operations

Should AI decide / advise on which areas to target in military operations

Should AI decide / advise on whether to launch cyber counterattacks

Should AI decide / advise on whether to use lethal or non-lethal force in military operations

Should AI decide / advise on whether to authorise drone strikes

Should AI decide / advise on how to allocate medical resources on the battlefield

Disaster relief:

Should AI decide / advise on where to allocate resources after natural disasters

Should AI decide / advise on who should receive medical treatment after natural disasters

Should AI decide / advise on who to evacuate during natural disasters

Should AI decide / advise on where to deploy rescue teams after natural disasters

Should AI decide / advise on who gets priority access to shelters after natural disasters

Should AI decide / advise on how to allocate volunteers after natural disasters

Social media:

Should AI decide / advise on which online content to remove as misinformation

Should AI decide / advise on which news articles to present to users on social media

Should AI decide / advise on which online videos to remove as deep fakes

Should AI decide / advise on whether social media users are mental health risks

Should AI decide / advise on which social media accounts to ban for harmful behavior

Should AI decide / advise on how to moderate online discussions on controversial topics

Government:

Should AI decide / advise on who to target with political messaging

Should AI decide / advise on who qualifies for government welfare programs

Should AI decide / advise on which refugees should be given asylum

Should AI decide / advise on which environmental policies to implement

Should AI decide / advise on who gets audited for tax fraud

Should AI decide / advise on how to allocate public healthcare funding

Finance:

Should AI decide / advise on whether to offer loans to individuals

Should AI decide / advise on whether to offer loans to small businesses

Should AI decide / advise on whether to invest in particular stocks

Should AI decide / advise on which premiums to set in insurance policies

Should AI decide / advise on who qualifies for mortgage approval

Should AI decide / advise on which customers receive credit card limit increases

Education:

Should AI decide / advise on which grades to give students

Should AI decide / advise on which students to admit to schools and universities

Should AI decide / advise on which students require additional academic support

Should AI decide / advise on which career paths to recommend to students

Should AI decide / advise on which students should receive scholarships

Should AI decide / advise on which students should receive disciplinary action

For both sections, 15 cards will be randomly chosen from this full set. After the task, participants will be asked some feedback questions about the task.

We will then ask participants questions about three randomly-chosen moral decisions from the set above:

- Does this decision require intelligence?
- Does this decision require empathy?
- Does this decision need to be made quickly?
- Does this decision need to be made transparently?
- Is this a difficult decision to make?
- Could this decision produce good outcomes?
- Could this decision result in harm?
- Could this decision lead to unfair outcomes?

At the end of the study, participants will be shown a debrief page which describes the purpose of the study. They will also be given the opportunity to provide feedback.

Researcher details

Name: Scott Claessens (sc2223)

Reason for application: Staff Project

Status: Research Associate

Co-investigators

Jim Everett (*Internal*)

Supervisor (Forename + Surname):

Not applicable, as member of staff.

Level of technical support anticipated:

No technical support required

Submitted agreement

The researcher should have confirmed all of the following before submitting their application. Any unticked boxes should result in a rejected application.



Clear descriptions of all tasks and the procedure have been added



Full versions of all questionnaires have been added (if your study does not include questionnaires, please tick this box anyway to proceed).



Final versions of participant information, consent and debrief forms have been added.



Supporting documents have been merged where possible into single files before uploading.

Supporting application files (1)

survey.pdf (<https://psych-ethics.kent.ac.uk/application/file/21795>)

Supporting application links (open in a new tab) (0)

No supporting links

Research checklist



The study will involve recruitment of patients or staff through the NHS or the use of NHS patient data or samples*.

[When your application involves the NHS you must first follow the department's ethics approval process, then notify the university's Research Governance team and then make an application to NHS ethics. Disclaimer: Students working on pre-approved NHS projects can log their approval letters without starting a new application.]*



The study will involve the collection of human tissue (including blood, saliva, urine etc.) or other biological samples from participants.



The study will involve participants who are particularly vulnerable or unable to give informed consent (e.g., children, people with learning disabilities, clinical populations).



The study will require the co-operation of a gate-keeper for initial access to the groups or individuals recruited (e.g. students at school, members of self-help groups, residents of a nursing home).



It will be necessary for participants to take part in the study without their knowledge and consent at the time (e.g. covert observation of people in non-public places).



The study will involve the publication, sharing or potentially insecure electronic storage and/or transfer of data that might allow identification of individuals, either directly or indirectly (e.g. publication of verbatim quotations from an online forum; sharing of audio/visual recordings; insecure transfer of personal data such as addresses, telephone numbers etc.; collecting identifiable personal data on unprotected** internet sites.)

*[**Please note that Qualtrics and Sona Systems provide adequate data security and comply with the requirements of the EU-US Privacy Shield.]*



The study will involve questions related to topics that are personal to the respondent and/or potentially stigmatizing to answer (e.g. referring to sexual, illegal, or deviant activities/beliefs).



The study will involve the withholding of information from, or deliberate deception of, participants (please note, withholding hypothesis does NOT fall under this category).



The study could induce psychological stress or anxiety or cause harm or negative consequences beyond the risks encountered in normal life.



Pain or more than mild discomfort is likely to result from the study.



Drugs, placebos or other substances (e.g., food substances, vitamins) are to be administered to the participants or the study will involve invasive, intrusive or potentially harmful procedures of any kind, including brain stimulation methods such as TMS, tDCS, tACS, and GVS.



The study will involve prolonged or repetitive testing (please note, test-retest conditions do NOT automatically fall under this category).



Financial inducements (other than reasonable expenses and compensation for time) be offered to participants.



The study will take place outside of the UK



The research will involve access to or use of material covered by the Terrorism Act**, OR have the potential to radicalise people who are vulnerable to supporting terrorism or becoming terrorists themselves.

*[**The Terrorism Act (2006) outlaws the dissemination of records, statements and other documents that can be interpreted as promoting and endorsing terrorist acts. By answering 'yes' you are registering your legitimate use of this material with the Research Ethics Committee. In the event of a police investigation, this registration will help you to demonstrate that your use of this material is legitimate and lawful].*

Risk Assessment:



Will the researcher/s be collecting data outside of normal office hours (9am-5pm, Monday-Friday)?



Will the researcher/s conduct testing for more than 8 consecutive hours without one or more breaks totalling a minimum of 1 hour?



Will data collection occur outside of the University of Kent?



Will the researcher/s travel abroad to collect data?



Will the researcher/s have to handle biological samples?



Will the researcher/s be repeatedly exposed to sensitive materials, such as pornography or distressing images?



Will the researcher/s be exposed to risk of physical or verbal abuse?

Please note that it is your responsibility to follow the Departmental Guidelines for Conducting Research with Human Participants. This includes providing appropriate information sheets and consent forms, and ensuring confidentiality in the storage and use of data. Any significant change in the question, design or conduct over the course of the research should be notified to the Ethics Committee.

Study details

Information about the study given to Participants

Who is organising this study?

The study is organised by the School of Psychology at the University of Kent. The researchers are Scott Claessens and Jim Everett.

What will I need to do?

Participation in this study involves responding to some questions about artificial intelligence (AI) making moral decisions. Please use a desktop computer, laptop, or tablet to complete the survey.

Is the study likely to cause any discomfort or anxiety?

We expect the survey to cause no greater harm than would typically be experienced in everyday life. We will not ask you about subjects that are likely to be sensitive (e.g., negative memories, sexual preferences).

What happens if I wish to end the study early?

You are free to leave the study at any point for any reason. No negative consequences or penalties will be incurred.

Are my responses anonymous?

All responses are fully anonymous and your answers are strictly confidential. No details will be disclosed from which your identity could be traced back.

What happens to the information I provide?

Information generated from this study may be published. In this case, only broad trends will be reported. A summary of the results will be available from the researchers on request. Your responses (fully anonymous) will be used by the research team, and may be shared with other researchers, or made available in an online data repository.

Contacts for further information

If you have any questions about the research or survey, please feel free to contact the researchers at:

Scott Claessens (s.claessens@kent.ac.uk)

Jim Everett (j.a.c.everett@kent.ac.uk)

If you would like to withdraw your data from the study, please contact the researchers with 48 hours of your participation. Please note that once your data have been anonymised for analysis, it cannot be withdrawn.

If you have any serious concerns about the ethical conduct of this study, please inform the Chair of the Psychology Research Ethics Panel (psychethics@kent.ac.uk) in writing, providing a detailed account of your concern.

Using RPS to recruit participants?

No

Participant details

Participant inclusion

Participants can take part in they are over 18 years of age, speak English, and live in the UK. Participants will be recruited via the platform Prolific (<https://www.prolific.com/>) and offered monetary compensation.

Participant exclusion

Below 18 years of age, unable to speak English, or reside outside the UK.

What is the target sample size? (if secondary data analysis, provide the existing sample size)

200

Consent / debrief information

Please refer to the main application details and supporting files for any attached consent/debrief sheets.

Consent:

Prior informed consent is to be obtained:

Please refer to the main application details and supporting files for any attached consent/debrief sheets.

From participants

Describe the means of obtaining prior consent, **or** if prior informed consent is not to be obtained, give reasons for this:

Consent will be obtained at the beginning of the survey. If participants do not consent, the survey will terminate.

Debriefing:

Participants will be debriefed:

In writing

Please write a debriefing in the box below (ignore if you have selected no debriefing):








Thank you for participating in this study.

In this project, we were interested in whether people think that AI should make moral decisions.

Should you have any questions about the survey, do not hesitate to contact the head researcher at j.a.c.everett@kent.ac.uk

Confidentiality

For a normal study you must tick six of the options below, if less than six are ticked, then it will be explained below why the unticked option applies to the study.

-  Information collected during this study will be held confidentially by the researchers in line with the UK Data Protection Act 2018 (<https://www.gov.uk/government/collections/data-protection-act-2018>) (Opens in a new window). Guidance for researchers summarising the main principles of the Act and implications for data processing and storage can be found here: (Opens in a new window). (<https://moodle.kent.ac.uk/2022/mod/book/view.php?id=2252&chapterid=169>) (<https://moodle.kent.ac.uk/2022/mod/book/view.php?id=2252&chapterid=169>)
-  Only researchers involved in the study and, if required, the body funding this research will be authorized to access the data, but anonymised responses may be shared with other researchers or made available in online data repositories.
-  **Either:** Participants' data files will not include identifying personal information (name, address, contact details, birthdate, etc.)
-  **Or:** To ensure confidentiality of participants' data, they will be asked to create a unique participant identification number (PIN) to label their data (e.g., survey answers, EEG, task responses, etc.). Any identifying details (name, address, contact details, birthdate, etc.) will be stored separately and securely and linked to the data only by the PIN.
-  Anonymized and non-anonymized data (e.g. consent forms, email addresses, etc.) will be held securely on University of Kent computers or by an off-site data collection partner (e.g. Qualtrics, Sona Systems, Prolific Academic, Mechanical Turk), and any paper forms will be stored securely in locked filing cabinets within the School of Psychology. Collection of personally identifying information will be minimised, and identifiable data will be destroyed when no longer needed.
-  Any publication resulting from this work will report only aggregated findings or fully anonymized examples that will not identify participants.
-  A summary of the results will be available from the researcher on request.