

Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our [Editorial Policies](#) and the [Editorial Policy Checklist](#).
Please do not complete any field with "not applicable" or n/a. Refer to the help text for what text to use if an item is not relevant to your study.
For final submission: please carefully check your responses for accuracy; you will not be able to make changes later.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed
☐	☒ The exact sample size (<i>n</i>) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒ A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒ The statistical test(s) used AND whether they are one- or two-sided <i>Only common tests should be described solely by name; describe more complex techniques in the Methods section.</i>
☒	☐ A description of all covariates tested
☐	☒ A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒ A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒ For null hypothesis testing, the test statistic (e.g. <i>F</i> , <i>t</i> , <i>r</i>) with confidence intervals, effect sizes, degrees of freedom and <i>P</i> value noted <i>Give P values as exact values whenever suitable.</i>
☒	☐ For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐ For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☒	☐ Estimates of effect sizes (e.g. Cohen's <i>d</i> , Pearson's <i>r</i>), indicating how they were calculated

Our web collection on [statistics for biologists](#) contains articles on many of the points above.

Software and code

Policy information about [availability of computer code](#)

Data collection	Microsoft Excel (version 16.79.2) was used for tabulation of data.
Data analysis	Relative Preference variables were extracted using Matlab 7.1 with the following toolboxes: Curve Fitting Toolbox 1.1.4, Image Processing Toolbox 5.1, Matlab Builder for Excel 1.2.5, Statistics Toolbox 5.1 and Symbolic Math Toolbox 3.1.3. Machine learning analyses were implemented in Python 3.9 using the packages imblearn 0.0, sklearn 1.2.2, pandas 2. 0.2, and pandas 2. 0.2. Feature importance (Gini scores and mutual information scores) were obtained with sklearn 1.2.2. Figures relating to variable importance were plotted with seaborn 0.12.2. Mediation/Moderation analyses were conducted in R 4.2. 0 with libraries readxl 1. 4. 0, MASS 7. 3-56 and stats 4.2. 0.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio [guidelines for submitting code & software](#) for further information.

Data

Policy information about [availability of data](#)

All manuscripts must include a [data availability statement](#). This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our [policy](#)

Anonymized data (relative preference variables, demographics and survey variables including suicidality variables) will be made publicly available on Open Science Framework pending acceptance of the manuscript. The data will be provided in Microsoft Excel.

Research involving human participants, their data, or biological material

Policy information about studies with [human participants or human data](#). See also policy information about [sex, gender \(identity/presentation\), and sexual orientation](#) and [race, ethnicity and racism](#).

Reporting on sex and gender

Gender was reported as assigned at birth. 38.09% of participants (1324) were male, 61.51 % of participants (2138) were female and 0.4% of participants (14) did not disclose their gender assigned at birth. 5.9% of participants (205) were between the ages of 18-24, 11.05% of participants (384) were between the ages of 25-34, 15.22% of participants (529) were between the ages of 35-44, 17.52% of participants (609) were between the ages of 45-54, 24.54% of participants (853) were between the ages of 55-64, and 25.78% of participants (896) were greater than 65 years of age.

Reporting on race, ethnicity, or other socially relevant groupings

2.91 % of participants (101) reported some high school education, 21.12% of participants (734) reported high school graduate education, 29.63% of participants (1030) reported some college education, 22.35% of participants (777) reported bachelor's degree education, 5.06% of participants (176) reported some graduate school education, 16.23% of participants (564) reported graduate degree education and 2.7% of participants (94) reported post-doctoral training education. 83.66% of participants (2908) were White/Caucasian, 6.53% of participants (227) were African American, 3.54% of participants (123) were Hispanic/Latino, 3.51 % of participants (122) were Asian or Pacific Islander, 0.75% of participants (26) were Native American or Alaskan Native, 0.86% of participants (30) were Mixed racial background, 0.46% of participants (16) other socially relevant were another race and 0.69% did not disclose their race or ethnicity. 86.02% of participants (2990) were right handed, 11.71 % of participants (407) were left handed and 2.27% of participants (79) were ambidextrous.

Population characteristics

Suicidal thought and behavior was reported on a five-point Likert scale. For passive suicidal ideation 79.17% of participants (2752) reported a score of 1, 6.88% of participants (239) reported a score of 2, 8.29% of participants (288) reported a score of 3, 3.60% of participants (125) reported a score of 4, and 2.07% of participants (72) reported a score of 5. For active suicidal ideation, 85.76% of participants (2981) reported a score of 1, 5.01 % of participants (174) reported a score of 2, 4.83% of participants (168) reported a score of 3, 2.76% of participants (96) reported a score of 4 and 1.64% of participants (57) reported a score of 5. For planning for suicide, 88.55% of participants (3078) reported a score of 1, 3.77% of participants (131) reported a score of 2, 3.48% of participants (121) reported a score of 3, 2.7% of participants (94) reported a score of 4, and 1.50% of participants (52) reported a score of 5. For planning for safety, 82.8% of participants (2878) reported a score of 1, 3.62% of participants (126) reported a score of 2, 4.69% of participants (163) reported a score of 3, 3.68% of participants (128) reported a score of 4, and 5.21 % of participants (181) reported a score of 5. 5.90% of participants (205) were between the ages of 18-24, 11.05% of participants (384) were between the ages of 25-34, 15.22% of participants (529) were between the ages of 35-44, 17.52% of participants (609) were between the ages of 45-54, 24.54% of participants (853) were between the ages of 55-64 and 25.78% of participants (896) were at least 65 years of age. No participants under of the age of 18 were included in the cohort. 86.02% of participants (2990) were right handed, 11.71% (407) were left handed and 2.27% of participants (79) were ambidextrous.

Recruitment

Recruitment was conducted by a third party vendor, Gold Research (San Antonio, Texas) in December, 2021, who oversampled 15% of an initial recruitment pool for mental health conditions. Gold Research then surveyed subjects, assessed Recruitment for response completeness, and performed an initial quality screen of subjects. They released deidentified data for 4,019 subjects to the authors who performed a separate quality assurance check following published methods (see references 37,39,49,57) to leave 3,476 subjects for data analysis. Potential self-selection bias relevant to the recruitment involved sampling individuals who used technology (a personal device was necessary to complete the survey and behavioral task) -- this bias may have confounded population characteristics.

Ethics oversight

All participants provided informed consent which included their primary participation in the study as well as the secondary use of their anonymized, de-identified data (i.e., all identifying information removed by Gold Research Inc. prior to retrieval by Ethics oversight the research group) in secondary analyses. Informed consent was obtained for all participants, as approved by the Institutional Review Board of Northwestern University (NU) (approval number STU00213665) for initial project start and later also approved by the University of Cincinnati (UC) Institutional Review Board (approval number 2023-0164) as some NU investigators moved to UC, to be in accordance with the Declaration of Helsinki.

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☐ Life sciences

☐ Behavioural & social sciences

☐ Ecological, evolutionary & environmental sciences

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	A third-party vendor, Gold Research Inc. (San Antonio, Texas), recruited adults across the united states of ages 18-70. To ensure adequate samples of participants with mental health conditions, Gold Research oversampled 15% of the sample for mental health conditions. Participant demographics were matched to the U.S. Census Bureau at the time of sampling in December, 2021. 4019 adults decided to participate and completed all procedures; after applying data exclusion criteria, data from 3476 participants was retained. State-of-the-art machine learning algorithms were used to predict suicidal thoughts and behaviors. No sample size calculations were performed in advance as no standard methods exist to calculate the required sample size for machine learning models. We sought a sample size in the upper range of what has been published to date (please see Table S5), namely in the range of 4000 subjects, knowing there would be sample reduction from quality assurance. Such a sample would also be adequate for the number of partitions we were using with the cross-validation approach to prediction.
Data exclusions	Of the 4019 deidentified participants released to authors for data analysis, participants meeting at least one of the following six criteria were omitted from the cohort: (1) participants with ten or more clinician-diagnosed illnesses, (2) participants that selected the same response for at least one section of the survey, (3) participants that rated all images in the behavioral task the same or with a variance of one (meaning only two of seven Likert points were used), (4) participants whose relative preference analysis yielded extreme outliers > 3 IQRS or incomplete measurements, (5) participants that had mis-matching responses to years of education and education level in survey, (6) participants that completed the questionnaire in less than 800 seconds. Data exclusion reduced the sample from 4019 participants to 3476 participants for analysis.
Replication	The results were generated with 10-fold cross validation in Python to ensure reproducibility.
Randomization	Gold Research randomized participant selection from their database that fell within the demographic criteria for ages 18-70. Machine learning fits were further also randomized using <i>sklearn</i> in <i>python</i> with 10-fold cross validation.
Blinding	Participants were anonymized by the third party vendor, Gold Research before being released to the authors for data for data analysis. Investigators were blinded to group allocation during data collection.

Behavioural & social sciences study design

All studies must disclose on these points even when the disclosure is negative.

Study description	
Research sample	
Sampling strategy	
Data collection	
Timing	
Data exclusions	
Non-participation	
Randomization	

Ecological, evolutionary & environmental sciences study design

All studies must disclose on these points even when the disclosure is negative.

Study description	
Research sample	
Sampling strategy	
Data collection	
Timing and spatial scale	
Data exclusions	
Reproducibility	
Randomization	
Blinding	

Did the study involve field work? ☐ Yes ☐ No

Field work, collection and transport

Field conditions	
Location	
Access & import/export	
Disturbance	

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☒	☐ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology and archaeology
☒	☐ Animals and other organisms
☒	☐ Clinical data
☒	☐ Dual use research of concern
☒	☐ Plants

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging

Antibodies

Antibodies used	
Validation	

Eukaryotic cell lines

Policy information about [cell lines and Sex and Gender in Research](#)

Cell line source(s)

Authentication

Mycoplasma contamination

Commonly misidentified lines
(See [ICLAC](#) register)

Palaeontology and Archaeology

Specimen provenance

Specimen deposition

Dating methods

☐ Tick this box to confirm that the raw and calibrated dates are available in the paper or in Supplementary Information.

Ethics oversight

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Animals and other research organisms

Policy information about [studies involving animals](#); [ARRIVE guidelines](#) recommended for reporting animal research, and [Sex and Gender in Research](#)

Laboratory animals

Wild animals

Reporting on sex

Field-collected samples

Ethics oversight

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Clinical data

Policy information about [clinical studies](#)

All manuscripts should comply with the ICMJE [guidelines for publication of clinical research](#) and a completed [CONSORT checklist](#) must be included with all submissions.

Clinical trial registration

Study protocol

Data collection

Outcomes

Dual use research of concern

Policy information about [dual use research of concern](#)

Hazards

Could the accidental, deliberate or reckless misuse of agents or technologies generated in the work, or the application of information presented in the manuscript, pose a threat to:

No	Yes
☐	☐ Public health
☐	☐ National security
☐	☐ Crops and/or livestock
☐	☐ Ecosystems
☐	☐ Any other significant area

Experiments of concern

Does the work involve any of these experiments of concern:

No	Yes
☐	☐ Demonstrate how to render a vaccine ineffective
☐	☐ Confer resistance to therapeutically useful antibiotics or antiviral agents
☐	☐ Enhance the virulence of a pathogen or render a nonpathogen virulent
☐	☐ Increase transmissibility of a pathogen
☐	☐ Alter the host range of a pathogen
☐	☐ Enable evasion of diagnostic/detection modalities
☐	☐ Enable the weaponization of a biological agent or toxin
☐	☐ Any other potentially harmful combination of experiments and agents

Plants

Seed stocks	
Novel plant genotypes	
Authentication	

ChIP-seq

Data deposition

- ☐ Confirm that both raw and final processed data have been deposited in a public database such as [GEO](#).
- ☐ Confirm that you have deposited or provided access to graph files (e.g. BED files) for the called peaks.

Data access links <i>May remain private before publication.</i>	
Files in database submission	
Genome browser session (e.g. UCSC)	

Methodology

Replicates	
Sequencing depth	
Antibodies	
Peak calling parameters	
Data quality	
Software	

Flow Cytometry

Plots

Confirm that:

- ☐ The axis labels state the marker and fluorochrome used (e.g. CD4-FITC).
- ☐ The axis scales are clearly visible. Include numbers along axes only for bottom left plot of group (a 'group' is an analysis of identical markers).
- ☐ All plots are contour plots with outliers or pseudocolor plots.
- ☐ A numerical value for number of cells or percentage (with statistics) is provided.

Methodology

Sample preparation

Instrument

Software

Cell population abundance

Gating strategy

☐ Tick this box to confirm that a figure exemplifying the gating strategy is provided in the Supplementary Information.

Magnetic resonance imaging

Experimental design

Design type

Design specifications

Behavioral performance measures

Imaging type(s)

Field strength

Sequence & imaging parameters

Area of acquisition

Diffusion MRI

☐ Used

☐ Not used

Preprocessing

Preprocessing software

Normalization

Normalization template

Noise and artifact removal

Volume censoring

Statistical modeling & inference

Model type and settings

Effect(s) tested

Specify type of analysis:

☐ Whole brain

☐ ROI-based

☐ Both

Statistic type for inference

(See [Eklund et al. 2016](#))

Correction

Models & analysis

n/a | Involved in the study

- | | | |
|--------------------------|--------------------------|--|
| ☐ | ☐ | Functional and/or effective connectivity |
| ☐ | ☐ | Graph analysis |
| ☐ | ☐ | Multivariate modeling or predictive analysis |

Functional and/or effective connectivity

Graph analysis

Multivariate modeling and predictive analysis

