

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](#).

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	Participants were recruited via Prolific and completed self-report surveys through Qualtrics. Chat data were donated using a custom-designed Chrome extension that allowed users to locally export their Character.AI transcripts, redact private content, and upload selected files via Qualtrics. The extension did not collect or transmit any data. Informed consent was obtained for all data collection. Extension code is available at: https://github.com/SALT-NLP/AI-companionship-well-being .
Data analysis	Statistical analyses were conducted in R (version 4.5.0) using linear regressions and Heckman selection models. Language model analyses were performed using locally hosted LLaMA 3, GPT-4o, and TopicGPT. Custom R and Python code was used for preprocessing, classification, and analysis. Prompts used in language model pipelines are provided in the Supplementary Information, and the full analysis code is available at: https://github.com/SALT-NLP/AI-companionship-well-being .

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](#)

The de-identified quantitative data are available at <https://github.com/SALT-NLP/AI-companionship-well-being>. For the qualitative data, raw chat history data are not publicly available because they contain potentially sensitive personal information and were collected under consent agreements that do not permit data sharing. Even after de-identification, the conversational nature of the data poses a substantial risk of participant re-identification. To support interpretation and verification of the qualitative findings while preserving participant privacy, we provide de-identified and processed materials as well as the prompts we used in the Supplementary Information, including all the prompts (Sections 1.11-1.15 in Supplementary Information), evaluation procedures for the topic modeling pipeline (Section 1.4.1 in Supplementary Information), chatbot-user conversation topics (Section 1.5 in Supplementary Information), and self-disclosure topics (Section 1.6 in Supplementary Information), as well as positive and negative influence topics derived from topic modeling (Section 1.7 in Supplementary Information), illustrated with paraphrased excerpts from participant self-report survey responses. In addition, we include the topic model figure (Figure 2) to show the categories derived from the data and their relative volumes.

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	Participants self-reported their gender identity in the survey. Gender was included as a control variable in all regression analyses. No analyses were stratified by gender. Among the final survey sample (n = 1,131), 48.5% of participants identified as male, 46.2% as female, and 5.1% as non-binary.
Reporting on race, ethnicity, or other socially relevant groupings	Data on race and ethnicity were not collected in this study. As part of our control variables, we included participants' relationship status. Regarding relationship status, 40.3% of participants reported being single, while 59.5% were in a romantic relationship.
Population characteristics	Participants were U.S.-based Character.AI users recruited through Prolific. All participants were 18 years or older. The sample skewed younger, with 41.8% aged 18-25, 32.7% aged 26-35, 18.7% aged 36-50, and 6.8% over 50.
Recruitment	Participants were recruited via Prolific using prescreening criteria to target U.S.-based, native English-speaking users of Character.AI. Eligibility required having used the platform for over one month and having interacted with at least three distinct chatbots. Of the 2,836 individuals who initially enrolled, 1,417 met eligibility criteria, and 1,131 completed the full survey. The survey was administered using Qualtrics and took approximately 10 minutes to complete. Participants received \$4 for survey completion. A subsample of 237 participants additionally contributed their chatbot chat histories, yielding 4,664 conversational sessions and 464,687 individual messages. Participants who submitted chat logs received an additional \$20 bonus.
Ethics oversight	This study was approved by Stanford University Institutional Review Board (Protocol ID: 76072).

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

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Behavioural & social sciences study design

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

Study description	This study examined how users form relationships with character-based AI chatbots and how these interactions relate to psychological well-being. Using a mixed -methods design, we combined survey responses with large language models (LLMs)-based content analysis of participant-donated chat transcripts.
Research sample	The sample included 1,131 U.S.-based participants recruited through Prolific who had used Character.AI for over one month and interacted with at least three distinct chatbots. A subset of 237 participants also donated chatbot conversation logs.
Sampling strategy	No statistical methods were used to pre- determine sample size. Based on prior works on technology use and well-being, we

Sampling strategy	specified $ \beta = .10$ as the smallest effect size of interest. Sensitivity analyses indicated that the present sample size provided 80% power at $\alpha = .05$ to detect effects of approximately $\beta = .08$ ($f^2 = .007$).
Data collection	The survey was administered using Qualtrics and took approximately 10 minutes to complete. To collect chat history data, we developed a custom Google Chrome extension that allowed participants to export their Character.AI conversation logs in JSON format. The tool enabled participants to review and optionally redact sensitive content prior to submission. All processing occurred locally on the participant's device; no raw chat data were transmitted to external servers or APIs. Participants uploaded the final JSON file at the end of the survey, and all data were stored on secure institutional servers accessible only to study authors.
Timing	Data collection took place between October 2024 and January 2025. Recruitment and survey administration were conducted continuously during this period. There were no distinct cohorts or gaps in data collection.
Data exclusions	Of the 2,836 individuals who initially enrolled in the study, 1,417 met the predefined eligibility criteria, which required participants to have used Character.AI for more than one month and interacted with at least three distinct chatbots. These criteria were established in advance to ensure inclusion of active users capable of providing meaningful data. Of those eligible, 1,131 participants completed the full survey and were included in the final analysis. Participants were excluded if they provided incomplete responses (e.g., missing data on key variables such as gender or relationship status) or failed to complete the entire survey. All exclusion criteria were pre-established prior to data analysis.
Non-participation	Of the 2,836 individuals who initially enrolled in the study, 1,417 met the eligibility criteria. Among those eligible, 286 participants did not complete the full survey and were therefore excluded from the final analysis. As this was a one-time survey with no follow-up, reasons for non-completion were not collected. These individuals met the inclusion criteria but chose not to finish the survey. No participants explicitly declined participation after enrollment.
Randomization	This study did not involve experimental manipulation or assignment to experimental groups.

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

Methods

n/a	Involved in the study	n/a	Involved in the study
☐	☐ Antibodies	☐	☐ ChIP-seq
☐	☐ Eukaryotic cell lines	☐	☐ Flow cytometry
☐	☐ Palaeontology and archaeology	☐	☐ MRI-based neuroimaging
☐	☐ Animals and other organisms		
☐	☐ Clinical data		
☐	☐ Dual use research of concern		
☐	☐ Plants		

Antibodies

Antibodies used	<i>Describe all antibodies used in the study; as applicable, provide supplier name, catalog number, clone name, and lot number.</i>
Validation	<i>Describe the validation of each primary antibody for the species and application, noting any validation statements on the manufacturer's website, relevant citations, antibody profiles in online databases, or data provided in the manuscript.</i>

Eukaryotic cell lines

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

Cell line source(s)	<i>State the source of each cell line used and the sex of all primary cell lines and cells derived from human participants or vertebrate models.</i>
Authentication	<i>Describe the authentication procedures for each cell line used OR declare that none of the cell lines used were authenticated.</i>
Mycoplasma contamination	<i>Confirm that all cell lines tested negative for mycoplasma contamination OR describe the results of the testing for mycoplasma contamination OR declare that the cell lines were not tested for mycoplasma contamination.</i>
Commonly misidentified lines (See ICLAC register)	<i>Name any commonly misidentified cell lines used in the study and provide a rationale for their use.</i>

Palaeontology and Archaeology

Specimen provenance	<i>Provide provenance information for specimens and describe permits that were obtained for the work (including the name of the issuing authority, the date of issue, and any identifying information). Permits should encompass collection and, where applicable, export.</i>
Specimen deposition	<i>Indicate where the specimens have been deposited to permit free access by other researchers.</i>
Dating methods	<i>If new dates are provided, describe how they were obtained (e.g. collection, storage, sample pretreatment and measurement), where they were obtained (i.e. lab name), the calibration program and the protocol for quality assurance OR state that no new dates are provided.</i>
☐ Tick this box to confirm that the raw and calibrated dates are available in the paper or in Supplementary Information.	
Ethics oversight	<i>Identify the organization(s) that approved or provided guidance on the study protocol, OR state that no ethical approval or guidance was required and explain why not.</i>

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	<i>For laboratory animals, report species, strain and age OR state that the study did not involve laboratory animals.</i>
Wild animals	<i>Provide details on animals observed in or captured in the field; report species and age where possible. Describe how animals were caught and transported and what happened to captive animals after the study (if killed, explain why and describe method; if released, say where and when) OR state that the study did not involve wild animals.</i>
Reporting on sex	<i>Indicate if findings apply to only one sex; describe whether sex was considered in study design, methods used for assigning sex. Provide data disaggregated for sex where this information has been collected in the source data as appropriate; provide overall numbers in this Reporting Summary. Please state if this information has not been collected. Report sex-based analyses where performed, justify reasons for lack of sex-based analysis.</i>
Field-collected samples	<i>For laboratory work with field-collected samples, describe all relevant parameters such as housing, maintenance, temperature, photoperiod and end-of-experiment protocol OR state that the study did not involve samples collected from the field.</i>
Ethics oversight	<i>Identify the organization(s) that approved or provided guidance on the study protocol, OR state that no ethical approval or guidance was required and explain why not.</i>

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	<i>Provide the trial registration number from ClinicalTrials.gov or an equivalent agency.</i>
Study protocol	<i>Note where the full trial protocol can be accessed OR if not available, explain why.</i>
Data collection	<i>Describe the settings and locales of data collection, noting the time periods of recruitment and data collection.</i>
Outcomes	<i>Describe how you pre-defined primary and secondary outcome measures and how you assessed these measures.</i>

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 | <i>Report on the source of all seed stocks or other plant material used. If applicable, state the seed stock centre and catalogue number. If plant specimens were collected from the field, describe the collection location, date and sampling procedures.</i> |
| Novel plant genotypes | <i>Describe the methods by which all novel plant genotypes were produced. This includes those generated by transgenic approaches, gene editing, chemical/radiation-based mutagenesis and hybridization. For transgenic lines, describe the transformation method, the number of independent lines analyzed and the generation upon which experiments were performed. For gene-edited lines, describe the editor used, the endogenous sequence targeted for editing, the targeting guide RNA sequence (if applicable) and how the editor was applied.</i> |
| Authentication | <i>Describe any authentication procedures for each seed stock used or novel genotype generated. Describe any experiments used to assess the effect of a mutation and, where applicable, how potential secondary effects (e.g. second site T-DNA insertions, mosaicism, off-target gene editing) were examined.</i> |

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> | <i>For "Initial submission" or "Revised version" documents, provide reviewer access links. For your "Final submission" document, provide a link to the deposited data.</i> |
| Files in database submission | <i>Provide a list of all files available in the database submission.</i> |
| Genome browser session
(e.g. UCSC) | <i>Provide a link to an anonymized genome browser session for "Initial submission" and "Revised version" documents only, to enable peer review. Write "no longer applicable" for "Final submission" documents.</i> |

Methodology

- | | |
|-------------------------|--|
| Replicates | <i>Describe the experimental replicates, specifying number, type and replicate agreement.</i> |
| Sequencing depth | <i>Describe the sequencing depth for each experiment, providing the total number of reads, uniquely mapped reads, length of reads and whether they were paired- or single-end.</i> |
| Antibodies | <i>Describe the antibodies used for the ChIP-seq experiments; as applicable, provide supplier name, catalog number, clone name, and lot number.</i> |
| Peak calling parameters | <i>Specify the command line program and parameters used for read mapping and peak calling, including the ChIP, control and index files used.</i> |

Data quality

Describe the methods used to ensure data quality in full detail, including how many peaks are at FDR 5% and above 5-fold enrichment.

Software

Describe the software used to collect and analyze the ChIP-seq data. For custom code that has been deposited into a community repository, provide accession details.

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

Describe the sample preparation, detailing the biological source of the cells and any tissue processing steps used.

Instrument

Identify the instrument used for data collection, specifying make and model number.

Software

Describe the software used to collect and analyze the flow cytometry data. For custom code that has been deposited into a community repository, provide accession details.

Cell population abundance

Describe the abundance of the relevant cell populations within post-sort fractions, providing details on the purity of the samples and how it was determined.

Gating strategy

Describe the gating strategy used for all relevant experiments, specifying the preliminary FSC/SSC gates of the starting cell population, indicating where boundaries between "positive" and "negative" staining cell populations are defined.

- ☐ 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

Indicate task or resting state; event-related or block design.

Design specifications

Specify the number of blocks, trials or experimental units per session and/or subject, and specify the length of each trial or block (if trials are blocked) and interval between trials.

Behavioral performance measures

State number and/or type of variables recorded (e.g. correct button press, response time) and what statistics were used to establish that the subjects were performing the task as expected (e.g. mean, range, and/or standard deviation across subjects).

Acquisition

Imaging type(s)

Specify: functional, structural, diffusion, perfusion.

Field strength

Specify in Tesla

Sequence & imaging parameters

Specify the pulse sequence type (gradient echo, spin echo, etc.), imaging type (EPI, spiral, etc.), field of view, matrix size, slice thickness, orientation and TE/TR/flip angle.

Area of acquisition

State whether a whole brain scan was used OR define the area of acquisition, describing how the region was determined.

Diffusion MRI

☐ Used☐ Not used

Preprocessing

Preprocessing software

Provide detail on software version and revision number and on specific parameters (model/functions, brain extraction, segmentation, smoothing kernel size, etc.).

Normalization

If data were normalized/standardized, describe the approach(es): specify linear or non-linear and define image types used for transformation OR indicate that data were not normalized and explain rationale for lack of normalization.

Normalization template	Describe the template used for normalization/transformation, specifying subject space or group standardized space (e.g. original Talairach, MNI305, ICBM152) OR indicate that the data were not normalized.
Noise and artifact removal	Describe your procedure(s) for artifact and structured noise removal, specifying motion parameters, tissue signals and physiological signals (heart rate, respiration).
Volume censoring	Define your software and/or method and criteria for volume censoring, and state the extent of such censoring.

Statistical modeling & inference

Model type and settings	Specify type (mass univariate, multivariate, RSA, predictive, etc.) and describe essential details of the model at the first and second levels (e.g. fixed, random or mixed effects; drift or auto-correlation).
Effect(s) tested	Define precise effect in terms of the task or stimulus conditions instead of psychological concepts and indicate whether ANOVA or factorial designs were used.
Specify type of analysis:	☐ Whole brain ☐ ROI-based ☐ Both
Statistic type for inference	Specify voxel-wise or cluster-wise and report all relevant parameters for cluster-wise methods.
(See Eklund et al. 2016)	
Correction	Describe the type of correction and how it is obtained for multiple comparisons (e.g. FWE, FDR, permutation or Monte Carlo).

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	Report the measures of dependence used and the model details (e.g. Pearson correlation, partial correlation, mutual information).
Graph analysis	Report the dependent variable and connectivity measure, specifying weighted graph or binarized graph, subject- or group-level, and the global and/or node summaries used (e.g. clustering coefficient, efficiency, etc.).
Multivariate modeling and predictive analysis	Specify independent variables, features extraction and dimension reduction, model, training and evaluation metrics.