ECAS7013

Portfolio: Ethics Application, Research Proposal, and Progress Report

Empathic Design and Working Alliance in Human-AI Wellbeing Interactions: A Randomised Experimental Study

PROGRESS REPORT

MSc Applied Psychology (MSC12MSC-APY)

ABSTRACT

Mental health services face a critical access gap, driving interest in conversational AI as low-intensity interim support. Yet the relational quality of these systems remains poorly understood. This portfolio documents the development of a randomised experimental study examining whether deliberate empathic design in a conversational agent, RAIYU, produces stronger perceived working alliance and greater short-term affective improvement compared with a standard large language model. It includes the research proposal, ethics submission, and a critical progress report reflecting on the methodological and ethical decisions made in bringing this study to the point of data collection.

Reia Sigrist (2219662)

Applied Psychology Project (ECAS7013)

Supervisor: Dr Paul Hutchings

Word count: 1503

Date submitted: 25 July 2026

Table of Contents

AI Statement 2
Section 1: Research proposal 3
Section 2: Ethics submission 15
Section 3: Progress report 60
3.1 Introduction 60
3.2 Planned vs Actual Project Timeline 60
3.3 Key Milestones Achieved 61
3.4 Challenges Encountered 62
3.5 Adaptations and Decision-Making 64
3.6 Preliminary Data Engagement and Emerging Patterns 65
3.7 Reflection on Researcher Development 65
3.8 Next Steps and Project Completion Plan 66
References 68
Appendix 70

AI Statement

I acknowledge the use of Claude (https://claude.ai/), Anthropic, throughout July 2026 to assist with the preparation of this progress report and its appendix materials. I used the tool for:

- Talking me through the process of completing an OSF pre-registration, including how to complete the registration form fields based on my existing study design and analysis plan
- Proofreading drafted text for grammar, clarity, and readability, and checking my reference list for formatting errors against APA 7th edition conventions

I confirm that I have not presented any AI-generated text as my own original thinking or research judgement. All research design decisions, project management judgements, and critical reflections in this report reflect my own understanding of the RAIYU study. All factual claims, dates, and project details were provided by me and verified against my own records. I confirm I have copies of my drafts, notes, and other resources used before and after the use of AI tools, which I may be asked to provide in evidence.

Section 1: Research proposal

Prifysgol Cymru
Y Drindod Dewi Sant
University of Wales
Trinity Saint David

MSC APPLIED PSYCHOLOGY: ECAS7013 PART II DISSERTATION RESEARCH PROPOSAL

Student Details

Student Family / Surname:	Sigrist	
Student Given / First name:	Rhiannon	
Student Number	2219662	
Proposed supervisor	Dr Paul Hutchings	
Has your proposed supervisor approved your Expression of Interest?	YES ☒	NO ☐

Proposed Research (this section must be no longer than 2,500 words)

Current Proposed Title (14 / 30 words):
Empathic Design and Working Alliance in Human–AI Wellbeing Interactions: A Randomised Experimental Study.
Background & Context *(to include rationale and theoretical framework, leading to a clear research question)*
Digital mental health technologies are increasingly explored as potential responses to the global treatment gap in psychological care. The World Health Organization estimates that approximately one in seven people experience a mental disorder annually, yet access to timely psychological support remains limited (GBD 2019 Mental Disorders Collaborators, 2022; WHO, 2022). In the United Kingdom, waiting times for psychological services frequently exceed recommended targets, leaving many without timely support (Mind, 2024; NHS, 2022). As a result, scalable digital interventions such as conversational agents and large language model (LLM)-based systems have gained attention as potential forms of low-intensity interim support, with emerging meta-analytic evidence indicating that such systems can produce small but meaningful short-term improvements in depression and anxiety symptoms, though effects appear limited to brief treatment windows and do not persist at three-month follow-up (Zhong et al., 2024). Despite rapid technological advancement, many digital mental health tools face persistent engagement challenges. Meta-analytic evidence suggests that smartphone-based mental health interventions experience high attrition and limited sustained user engagement (Baumel et al., 2019; Firth et al., 2017). A key explanation is that digital systems struggle to reproduce the relational dynamics that underpin effective psychological support. Within psychotherapy research, the working alliance (defined as the collaborative and affective

bond between helper and recipient) is one of the most robust predictors of therapeutic outcomes across modalities (Bordin, 1979; Flückiger et al., 2018), comprising three components: agreement on goals, agreement on tasks, and development of a trusting bond. Systematic review evidence supports this relational explanation, finding that empathic and relational process qualities consistently emerge as the primary drivers of positive chatbot engagement and acceptability above content or technical features (Limpanopparat et al., 2024).

Research in human-computer interaction suggests that individuals naturally apply social expectations to interactive technologies, and that users may experience perceived relational bonds with computer agents designed to use socially attuned behaviours (Bickmore & Picard, 2005). This provides a theoretical basis for examining whether Bordin's (1979) alliance framework can meaningfully extend to human–AI interaction. Preliminary studies examining mental health chatbots indicate that users can indeed report alliance-like perceptions when interacting with conversational agents (Beatty et al., 2022), and qualitative pilot evidence suggests users can form spontaneous relational perceptions, describing chatbots as living characters and interactions as genuine relationships (Ly et al., 2017). Nevertheless, integrative reviews suggest that alliance formation in digital contexts remains substantially weaker than that observed in human-delivered therapy, raising questions about whether current AI systems can meaningfully replicate the relational processes associated with effective psychological support (Malouin-Lachance et al., 2025). Users frequently exhibit systematic resistance to AI in healthcare contexts, perceiving AI-delivered support as less trustworthy than equivalent human support regardless of technical accuracy (Longoni et al., 2019), presenting a further barrier to relational engagement.

One proposed explanation for this limitation relates to the difficulty AI systems face in producing emotionally appropriate responses. LLMs generate text probabilistically rather than through genuine emotional understanding, which can lead to responses that appear superficially empathic but lack contextual attunement or psychological safety. Meta-analytic evidence from human therapy consistently identifies empathy as one of the most robust predictors of client outcomes across therapeutic modalities (Elliott et al., 2018), yet recent evaluations of generative AI in mental health–adjacent contexts suggest this quality remains largely unachieved, with concerns raised regarding inconsistent emotional responsiveness, hallucinated information, and inadequate responses to expressions of distress (Hua et al., 2025; Torous et al., 2025). Empirical work has also identified a potential "safety–empathy gap", whereby systems designed to be supportive may struggle to balance empathic responsiveness with appropriate safeguarding behaviours when users disclose distress (Pichowicz et al., 2025). Research examining healthcare chatbots further suggests that AI-generated empathy frequently fails to feel authentic, producing discomfort rather than connection, with perceived authenticity emerging as a significant mediator of user experience quality (Seitz, 2024), highlighting the need to move beyond surface-level linguistic mimicry.

Addressing this challenge requires approaches that integrate psychological theory into the design of conversational systems. One emerging strategy involves embedding structured empathic communication techniques within system prompts and response constraints, operationalising core micro-skills from counselling psychology, specifically behavioural-

empathic strategies such as active listening, validation, and collaborative goal orientation rather than affective mirroring, which evidence suggests risks reducing perceived authenticity in AI interactions (Seitz, 2024), while maintaining clear non-clinical boundaries. In principle, such empathy-engineered conversational agents may enhance perceived relational quality without presenting themselves as substitutes for professional psychological care.

The present study evaluates RAIYU, an empathy-engineered conversational agent integrating structured empathic communication strategies with retrieval-augmented grounding and automated safety filtering. To support participant safety, the system incorporates a real-time filtering mechanism based on Nelson et al. (2025), which screens user inputs for indicators of psychological crisis (including expressions of suicidal ideation) and redirects flagged content to pre-specified safeguarding responses rather than allowing the generative model to respond freely to high-risk disclosures. By comparing participant interactions with RAIYU against a baseline LLM under controlled conditions, the study examines whether empathic design features produce stronger perceived working alliance and greater short-term reductions in negative affect, and explores whether baseline psychological distress moderates these relational and affective outcomes.

Accordingly, the study addresses the following research questions:
RQ1: Does interaction with an empathy-engineered conversational agent lead to higher perceived working alliance than interaction with a standard large language model?
RQ2: Is perceived working alliance associated with short-term affective change following interaction with the conversational agent?
RQ3: Does baseline psychological distress moderate the relationship between AI agent type and user-reported alliance or affective outcomes?

Aims and Objectives

Aim
The primary aim of this research is to experimentally evaluate whether an empathy-engineered conversational agent can enhance perceived working alliance and short-term affective outcomes during human–AI wellbeing interactions. Drawing on the working alliance framework (Bordin, 1979), the study seeks to determine whether embedding structured empathic communication strategies within a conversational agent improves users' relational perceptions and emotional responses compared with a standard LLM.

Objectives
1. To experimentally compare perceived working alliance between participants interacting with an empathy-engineered conversational agent (RAIYU) and those interacting with a baseline large language model.
2. To examine whether interaction with an empathy-engineered agent produces greater short-term reductions in negative affect.
3. To investigate whether stronger alliance perceptions are associated with greater affective improvement.

4. To explore whether baseline psychological distress moderates user experiences of alliance and affective outcomes, thereby identifying individual differences that may influence responses to AI-mediated wellbeing interactions.

Methodology

Research Design
The study will employ a randomised mixed design, combining a between-subjects factor (AI agent type: empathy-engineered vs baseline LLM) with a within-subjects factor (time: pre- vs post-interaction measurement). This structure enables both direct comparison of relational and affective outcomes between conditions and examination of change across the interaction within participants.

The primary dependent variables are perceived working alliance and affective change following interaction with the conversational agent. Working alliance will be measured using the Working Alliance Inventory – Short Revised (WAI-SR; Hatcher & Gillaspy, 2006), while affective change will be assessed using the Positive and Negative Affect Schedule – Short Form (PANAS-SF; Thompson, 2007). Baseline psychological distress will be measured using the Kessler Psychological Distress Scale (K10; Kessler et al., 2002) and explored as a potential moderating variable.

In addition to quantitative measures, qualitative insights will be collected using a concurrent think-aloud protocol, in which participants verbalise their thoughts and reactions during the interaction (Ericsson & Simon, 1993). This embedded qualitative component will provide contextual insight into participants' perceptions of empathic responsiveness, relational engagement, and conversational dynamics. Because concurrent verbalisation may influence naturalistic engagement with the conversational agent, the think-aloud component is optional and intended primarily to provide exploratory contextual insight rather than serve as a primary outcome measure.

Participants
Participants will be undergraduate students aged 18 years or older recruited from the University of Wales Trinity Saint David through the departmental Research Participation Scheme and university mailing lists. Eligibility criteria include fluency in English and the absence of current suicidal ideation, which will be screened prior to participation.

A target sample of 60–80 participants (approximately 30–40 per condition) is proposed. As prior effect size estimates for empathy-engineered conversational agents are limited, the proposed sample size is intended to provide sufficient sensitivity for detecting medium-sized effects while remaining feasible within the scope of an MSc dissertation. Given the novel nature of the intervention, the study is appropriately scoped as an initial investigation.

Participation will be entirely voluntary. Undergraduate students may receive course credit through the Research Participation Scheme. Participants may withdraw at any point without penalty and without providing a reason.

Materials and Measures

Working Alliance Inventory – Short Revised (WAI-SR)

The WAI-SR is a 12-item self-report measure assessing perceived working alliance across three components: bond, task agreement, and goal agreement (Hatcher & Gillaspy, 2006). Items are rated on a seven-point Likert scale ranging from strongly disagree to strongly agree. In the present study, the WAI-SR will be administered immediately before and after the interaction with the conversational agent. Baseline administration captures participants' relational expectations prior to engagement, while post-interaction administration reflects perceived alliance following the exchange. This adaptation of the measure's standard post-interaction use enables direct comparison of alliance perceptions and provides a more sensitive index of relational change.

Positive and Negative Affect Schedule – Short Form (PANAS-SF)

Affective state will be measured using the PANAS-SF, a validated short-form measure of positive and negative affect (Thompson, 2007). Participants rate ten emotion descriptors on a five-point scale indicating how strongly they feel each emotion at the present moment. The scale will be administered before and after the interaction to assess short-term affective change.

Kessler Psychological Distress Scale (K10)

Baseline psychological distress will be assessed using the K10, a widely used screening measure of non-specific psychological distress (Kessler et al., 2002). The scale consists of ten items assessing frequency of distress-related symptoms over the past four weeks. K10 scores will be used as a continuous individual-differences variable for exploratory moderation analyses.

System Usability Scale (SUS)

To control for potential usability confounds, participants will complete the System Usability Scale (SUS) following the interaction (Brooke, 1996). The SUS is a widely used ten-item instrument for evaluating perceived system usability.

Conversational AI Platform

Participants will interact with a locally hosted conversational platform containing two experimental conditions. The empathy-engineered condition incorporates structured behavioural-empathic communication strategies such as active listening, validation, and collaborative goal orientation. The baseline condition utilises the same underlying LLM architecture but without empathic prompting or relational guidance. Using the same underlying model across conditions ensures that differences in outcomes are attributable to empathic design features rather than differences in model capability.

Procedure

Participants will first complete an online eligibility screening before booking a laboratory session. Screening will include age verification and an assessment of suicidal ideation using Item 9 of the PHQ-9 (Kroenke et al., 2001). Participants indicating current suicidal ideation suggestive of elevated risk will be excluded and provided with support resources.

Eligible participants will attend a laboratory session where they will provide informed consent before completing baseline questionnaires, including the K10, anticipated WAI-SR, and pre-interaction PANAS-SF, followed by demographic questions.

Participants will then be randomly assigned to interact for approximately 10 minutes with either the empathy-engineered conversational agent (RAIYU) or the baseline LLM. A 10-minute interaction window aligns with established chatbot research practice (Xu et al., 2025) and evidence that core aspects of alliance can emerge within a single session (Río Olvera et al., 2022). On a voluntary basis, participants will engage in a concurrent think-aloud procedure, verbally describing their thoughts and reactions while using the system.

Following the interaction, participants will complete the post-interaction WAI-SR, post-interaction PANAS-SF, and the System Usability Scale before receiving a full debrief including information about the study aims and access to psychological support resources if required.

Data Analysis
Quantitative analyses will be conducted in R. Independent-samples t-tests will compare post-interaction working alliance scores between conditions, and mixed-design analyses will examine pre-to-post PANAS-SF change across conditions. Regression analyses will explore associations between perceived alliance and affective change. To examine moderation, a path analysis using the lavaan package will model the moderating role of baseline psychological distress (K10) across both the between-subjects (AI agent type) and within-subjects (affective change) components of the design; given the sample size, this will be treated as exploratory. Qualitative data from the think-aloud protocol will be analysed using reflexive thematic analysis (Braun & Clarke, 2006) to identify patterns relating to perceived empathy, relational engagement, and interactional breakdowns.

Proposed plan of work including timetable or Gantt Chart

The project follows the MSc dissertation timeline, aligning key stages with required course milestones. During January–March 2026, a systematic literature review will refine research questions and identify theoretical and methodological foundations. In April 2026, the research proposal and methodological design will be finalised.

Following proposal approval in May 2026, the ethics application will be submitted in June 2026, alongside finalisation and pilot testing of the conversational platform.

Data collection is scheduled for July 2026 and will involve recruiting participants through the University Research Participation Scheme and conducting laboratory-based interaction sessions. Quantitative and qualitative analyses will be conducted during August 2026, alongside drafting of the dissertation chapters. The final stage (September 2026) will involve completing the dissertation, integrating supervisor feedback, and preparing the visual showcase assessment for submission.

Jan-Mar 2026	Literature review and refinement of research questions
April 2026	Finalise research design and proposal
May 2026	Submission of assessed research proposal

June 2026	Ethics approval and system preparation/pilot testing
July 2026	Participant recruitment and data collection
August 2026	Data analysis and dissertation drafting
September 2026	Final dissertation writing and submission

Anticipated Outcomes and Real World Impact

The study is expected to generate empirical evidence regarding whether empathy-engineered conversational agents can enhance perceived working alliance and short-term emotional outcomes, clarifying whether structured empathic communication strategies meaningfully improve relational perceptions compared with standard LLM interactions.

From a theoretical perspective, the research will contribute to cyberpsychology by examining whether core relational constructs from psychotherapy (particularly the working alliance framework) can meaningfully translate to interactions with artificial agents, and may provide insight into how individual differences in psychological distress influence engagement with AI-mediated support systems.

From an applied perspective, the findings have potential implications for the responsible design of digital mental health technologies. As healthcare systems increasingly explore AI-supported tools to supplement overstretched mental health services, understanding how empathic design features influence user experience is critical. Evidence generated by this research may inform the development of safer, ethically grounded conversational agents that prioritise emotional validation, appropriate boundaries, and responsible signposting rather than attempting to replace professional care.

More broadly, the project may contribute to ongoing discussions regarding the role of AI in preventative and low-intensity psychological support, offering design insights relevant to digital wellbeing tools used by universities, healthcare providers, and organisations seeking scalable approaches to supporting mental health.

Reference list *(not included in word count)*

References

Baumel, A., Muench, F., Edan, S., & Kane, J. M. (2019). Objective user engagement with mental health apps: systematic search and panel-based usage analysis. *Journal of Medical Internet Research, 21*(9). https://doi.org/10.2196/14567

Beatty, C., Malik, T., Meheli, S., & Sinha, C. (2022). Evaluating the therapeutic alliance with a free-text CBT conversational agent (Wysa): a mixed-methods study. *Frontiers in Digital Health, 4*, 847991. https://doi.org/10.3389/fdgth.2022.847991

Bickmore, T. W., & Picard, R. W. (2005). Establishing and maintaining long-term human-computer relationships. *ACM Transactions on Computer-Human Interaction, 12*(2), 293–327. https://doi.org/10.1145/1067860.1067867

Bordin, E. S. (1979). The generalizability of the psychoanalytic concept of the working alliance. *Psychotherapy: Theory, Research & Practice, 16*(3), 252–260. https://doi.org/10.1037/h0085885

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology, 3*(2), 77–101. https://doi.org/10.1191/1478088706qp063oa

Brooke, J. (1996). SUS: A quick and dirty usability scale. In P. W. Jordan, B. Thomas, B. A. Weerdmeester, & I. L. McClelland, *Usability evaluation in industry* (pp. 189–194). Taylor & Francis. https://doi.org/10.1201/9781498710411

Elliott, R., Bohart, A. C., Watson, J. C., & Murphy, D. (2018). Therapist empathy and client outcome: An updated meta-analysis. *Psychotherapy, 55*(4), 399–429. https://doi.org/10.1037/pst0000175

Ericsson, K. A., & Simon, H. A. (1993). *Protocol analysis: Verbal reports as data (rev. ed.).* MIT Press.

Firth, J., Torous, J., Nicholas, J., Carney, R., Pratap, A., Rosenbaum, S., & Sarris, J. (2017). The efficacy of smartphone-based mental health interventions for depressive symptoms: a meta-analysis of randomized controlled trials. *World Psychiatry, 16*(3), 287-298. https://doi.org/10.1002/wps.20472

Flückiger, C., Del Re, A. C., Wampold, B. E., & Horvath, A. O. (2018). The alliance in adult psychotherapy: A meta-analytic synthesis. *Psychotherapy, 55*(4), 316. https://doi.org/10.1037/pst0000172

GBD 2019 Mental Disorders Collaborators. (2022). Global, regional, and national burden of 12 mental disorders in 204 countries and territories, 1990–2019: a systematic analysis for the Global Burden of Disease Study 2019. *The Lancet Psychiatry, 9*(2), 137-150. https://doi.org/10.1016/S2215-0366(21)00395-3

Hatcher, R. L., & Gillaspy, J. A. (2006). Development and validation of a revised short version of the Working Alliance Inventory. *Psychotherapy Research, 16*(1), 12-25. https://doi.org/10.1080/10503300500352500

Hua, Y., Na, H., Li, Z., Liu, F., Fang, X., Clifton, D., & Torous, J. (2025). A scoping review of large language models for generative tasks in mental health care. *npj Digital Medicine, 8*(1), 230. https://doi.org/10.1038/s41746-025-01611-4

Kessler, R. C., Andrews, G., Colpe, L. J., Hiripi, E., Mroczek, D. K., Normand, S. L., . . . Zaslavsky, A. M. (2002). Short screening scales to monitor population prevalences and trends in non-specific psychological distress. *Psychological Medicine, 32*(6), 959-976. https://doi.org/10.1017/S0033291702006074

Kroenke, K., Spitzer, R. L., & Williams, J. B. (2001). The PHQ-9: validity of a brief depression severity measure. *Journal of General Internal Medicine, 16*(9), 606-613. https://doi.org/10.1046/j.1525-1497.2001.016009606.x

Limpanopparat, S., Gibson, E., & Harris, A. (2024). User engagement, attitudes, and the effectiveness of chatbots as a mental health intervention: A systematic review. *Computers in Human Behavior: Artificial Humans, 2*(2), 100081. https://doi.org/10.1016/j.chbah.2024.100081

Longoni, C., Bonezzi, A., & Morewedge, C. K. (2019). Resistance to medical artificial intelligence. *Journal of Consumer Research, 46*(4), 629–650. https://doi.org/10.1093/jcr/ucz013

Ly, K. H., Ly, A. M., & Andersson, G. (2017). A fully automated conversational agent for promoting mental well-being: A pilot RCT using mixed methods. *Internet Interventions, 10*, 39–46. https://doi.org/10.1016/j.invent.2017.10.002

Malouin-Lachance, A., Capolupo, J., Laplante, C., & Hudon, A. (2025). Does the digital therapeutic alliance exist? Integrative review. *JMIR Mental Health, 12*, e69294. https://doi.org/10.2196/69294

Mind. (2024). *Thousands of people are still waiting six months or more to access specialist psychological support in Wales.* Mind Cymru. Retrieved from https://www.mind.org.uk/news-campaigns/news/thousands-of-people-are-still-waiting-six-months-or-more-to-access-specialist-psychological-support-in-wales/

Nelson, B. W., Wong, C., Silvestrini, M. T., Shin, S., Robinson, A., Lee, J., . . . Trister, A. (2025). An AI-based behavioral health safety filter and dataset for identifying mental health crises in text-based conversations. *arXiv preprint*. https://doi.org/10.48550/arXiv.2510.12083

NHS. (2022). *Mental health access and waiting time resources*. NHS England. Retrieved from https://www.england.nhs.uk/mental-health/resources/access-waiting-time/

Pichowicz, W., Kotas, M., & Piotrowski, P. (2025). Performance of mental health chatbot agents in detecting and managing suicidal ideation. *Scientific Reports, 15*(1), 31652. https://doi.org/10.1038/s41598-025-17242-4

Río Olvera, F. J., Rodríguez Mora, Á., Senín Calderón, M. C., & Rodríguez-Testal, J. F. (2022). The first session is the one that counts: An exploratory study of therapeutic alliance. *Frontiers in Psychology, 13*. https://doi.org/10.3389/fpsyg.2022.1016963

Seitz, L. (2024). Artificial empathy in healthcare chatbots: Does it feel authentic? *Computers in Human Behavior: Artificial Humans, 2*(1), 100067. https://doi.org/10.1016/j.chbah.2024.100067

Thompson, E. R. (2007). Development and validation of an internationally reliable short-form of the positive and negative affect schedule (PANAS). *Journal of Cross-Cultural Psychology, 38*(2), 227-242. https://doi.org/10.1177/0022022106297301

Torous, J., Linardon, J., Goldberg, S. B., Sun, S., Bell, I., Nicholas, J., . . . Firth, J. (2025). The evolving field of digital mental health: current evidence and implementation issues for smartphone apps, generative artificial intelligence, and virtual reality. *World Psychiatry, 24*(2), 156-174. https://doi.org/10.1002/wps.21299

WHO. (2022). *World Health Organization*. Retrieved from Mental disorders: https://www.who.int/news-room/fact-sheets/detail/mental-disorders

Xu, Z., Lee, Y. C., Stasiak, K., Warren, J., & Lottridge, D. (2025). The Digital Therapeutic Alliance With Mental Health Chatbots: Diary Study and Thematic Analysis. *JMIR Mental Health, 12*. https://doi.org/10.2196/76642

Zhong, W., Luo, J., & Zhang, H. (2024). The therapeutic effectiveness of artificial intelligence-based chatbots in alleviation of depressive and anxiety symptoms in short-course treatments: a systematic review and meta-analysis. *Journal of Affective Disorders, 356*, 459–469. https://doi.org/10.1016/j.jad.2024.04.057

Facilities and resources available for the investigation: please indicate which resources you will require for your project *(not included in word count)*

The following facilities and resources will be required for the successful completion of the project:

Laboratory Facilities
Access to a university computer laboratory or designated research room for conducting controlled participant interaction sessions.

Computer Hardware
A password-protected personal laptop used by the researcher to host and run the conversational AI system during the interaction sessions.

Conversational AI Platform
A locally hosted conversational interface enabling interaction with the empathy-engineered agent (RAIYU) and the baseline large language model condition. The system will operate offline or within a locally controlled environment to ensure that participant data are not transmitted to external servers.

Survey and Data Collection Tools
Online questionnaire software (e.g., Qualtrics or equivalent) for administering psychological measures including the Working Alliance Inventory – Short Revised (WAI-

SR), Positive and Negative Affect Schedule – Short Form (PANAS-SF), Kessler Psychological Distress Scale (K10), System Usability Scale (SUS), and demographic questions.

Audio Recording
Digital audio recording for capturing participants' think-aloud verbalisation during the interaction sessions.

Data Analysis Software
R statistical software for quantitative analysis.
Standard qualitative analysis procedures for analysing think-aloud data using reflexive thematic analysis.

Secure Data Storage
All study data will be stored on encrypted, password-protected university-approved storage systems in accordance with UK GDPR and institutional data protection policies.

Special Issues

Tick √

	YES	NO
Proposal requires additional external approvals	☐	☒
Proposal includes vulnerable groups	☐	☒
Research is part of a group project	☐	☒
Research is to be undertaken outside of the UK	☐	☒
Is a bar on access being sought?	☐	☒

Training Needs

Please include any specific training needs that have been identified for this project:
The researcher will undertake additional familiarisation with R statistical software to support quantitative data analysis. This will involve independent learning and consultation of statistical guidance materials during the data analysis phase. No other specialist training is required beyond the research training provided within the MSc Applied Psychology programme.

Ethical Issues

Following approval of the Full Research Proposal an application must be made by ALL research students to the Institute Ethics Committee through the form PG2 prior to any data collection taking place.	
I have understood the above statement and will make an application to the Ethics Committee via my supervisor following approval of the Full Research Proposal	☒

Student Signature

Student Signature	R.Sigrist	Date: 19/05/2026

Supervisory Team

Comments on the proposal by the supervisor (or programme director if no supervisor agreed)(approx. 300 words):
Comments on the progress made so far by the supervisor (approx. 300 - 500 words)

Authorisation

To be completed by Supervisor and Programme Manager

Prifysgol Cymru
Y Drindod Dewi Sant
University of Wales
Trinity Saint David

Supported: ☐	**Not Supported:** ☐

Supervisor	Signature or Print Name:	Date:
Programme Manager	Signature or Print Name:	Date:

Section 2: Ethics submission

Prifysgol Cymru
Y Drindod Dewi Sant
University of Wales
Trinity Saint David

PG2 FORM

APPLICATION FOR ETHICAL APPROVAL – INTRODUCTORY NOTES

In order for research to result in benefit and minimise risk of harm, it must be conducted ethically. A researcher may not be covered by the University's insurance if ethical approval has not been obtained prior to commencement.

The University follows the OECD Frascati manual definition of **research activity**:

> *"[C]reative work undertaken on a systematic basis in order to increase the stock of knowledge, including knowledge of man, culture and society, and the use of this stock of knowledge to devise new applications".*

As such this covers activities undertaken by:

- Members of staff (academic staff or support staff);
- Individuals holding honorary title from the University;
- Postgraduate research students;

In all of these cases ethical review is required regardless of discipline.

The individual undertaking the research activity is known as the **lead researcher**.

Ethical approval is not required for routine audits, performance reviews, quality assurance studies, testing within normal educational requirements, and literary or artistic criticism.

For most projects you will only need to complete sections in the main part of the form. A few projects will need more detailed ethical consideration and will need you to fill in sections in an appendix. At some points in this form you will be asked questions about your research in the following way:

1	Does your research involve <<something specific>>?	**YES**	**Click to continue to appendix <<N>>**	**NO**	**Continue to the next section**
2	Have you completed appendix <<N>>?	**YES**	☐	**NO**	☐

Clicking in the appropriate box will take you to the appropriate location in the text. At the end of each appendix is a box. Clicking on this will return you to the question that originally took you to the appendix.

Return to the main form

NB These clickable links will only work if you are editing the document in the desktop version of Word, not the browser version. Please download and save the form before editing it.

Guidance notes are provided as comments or as text IN GREEN throughout this form. Please ensure that the comments are removed and that the green text is replaced by your own text or removed as appropriate BEFORE submitting this form.

Please ensure that you have read the UWTSD Research Ethics and Integrity Policy and the UWTSD Research Data Management Policy BEFORE submitting this form.

Prifysgol Cymru
Y Drindod Dewi Sant
University of Wales
Trinity Saint David

PG2 FORM

APPLICATION FOR ETHICAL APPROVAL

SECTION A: About You (Lead Researcher)

1	Full Name:		Rhiannon Sigrist			
2	Tick all boxes that apply:		Member of staff:	☐	Honorary research fellow:	☐
	Undergraduate Student	☐	Taught Postgraduate Student	☒	Postgraduate Research Student	☐
3	Institute/Academic Discipline/Centre:		Institute of Education and Humanities / Psychology and Counselling			
4	Campus:		University of Wales Trinity Saint David – IQ Swansea (Yfforwm)			
5	E-mail address:		2219662@student.uwtsd.ac.uk			
6	Contact Telephone Number:		07889210292			
7	Project Title		Empathic Design and Working Alliance in Human–AI Wellbeing Interactions: A Randomised Experimental Study.			
8	Proposed Start Date		July 2026	Proposed End Date	September 2026	
For students:						
9	Student Number:		2219662			
10	Programme of Study:		MSc Applied Psychology			
11	Director of Studies/Supervisor:		Dr Paul Hutchings			

SECTION B: Grant or Contract Funding

1	Has your research been awarded, applied for, or intend to apply for external funding?	**YES**	**Click to continue to appendix 1**	**NO**	**Continue to the next section**
2	Have you completed appendix 1?	**YES**	☐	**NO**	☐

SECTION C: Internal and External Ethical Guidance Materials

Ethical Guidance Documents Please list the core ethical guidance documents that have been referred to during the completion of this form (including any discipline-specific codes of research ethics, location-specific codes of research ethics, and also any specific ethical guidance relating to the proposed methodology). Please tick to confirm that your research proposal adheres to these codes and guidelines. You may add rows to this table if needed.		
1	UWTSD Research Ethics & Integrity Policy	☒
2	UWTSD Research Data Management Policy	☒
3	UWTSD Open Access Publications Policy	☒
4	BPS Code of Ethics and Conduct (Dec 2021)	☒

5	BPS Code of Human Research Ethics (April 2021)	☒
6	BPS Data Protection Regulation – Guidance for Researchers (2018)	☒
7	ESRC Framework for Research Ethics (2015)	☒
8	General Data Protection Regulation (GDPR) 2016	☒
9	Data Protection Act 2018	☒
10	Nuffield Council on Bioethics: AI and Human Rights (2018)	☒
11	European Commission: Ethics Guidelines for Trustworthy AI (2019)	☒
12	WHO Ethics and Governance of AI for Health (2021)	☒

SECTION D: External Collaborative Research Activity

1	Does your research involve collaborating partners at another institution or organisation?	**YES**	**Click to continue to appendix 2**	**NO**	**Continue to the next section**
2	Have you completed appendix 2?	**YES**	☐	**NO**	☐

SECTION E: Details of Research Activity

Aims, Vision and Approach of the Research (maximum 200 words per section) Please outline what are you hoping to achieve with your proposed work and how will you deliver it?	
1	**Research aims and objectives**
The primary aim of this research is to experimentally evaluate whether an empathy-engineered conversational AI agent produces a stronger perceived working alliance and greater short-term improvement in mood during human–AI wellbeing interactions compared with a standard AI agent. Drawing on Bordin's (1979) working alliance framework, the study examines whether deliberately designed empathic communication strategies improve users' relational perceptions and emotional responses. The objectives are as follows. First, to compare perceived working alliance between participants interacting with RAIYU, an empathy-engineered conversational agent, and those interacting with a standard baseline agent, using the Working Alliance Inventory – Short Revised (WAI-SR). Second, to examine whether interaction with RAIYU produces greater short-term improvement in mood, specifically a greater reduction in negative affect, as measured by the Positive and Negative Affect Schedule – Short Form (PANAS-SF). Third, to investigate whether stronger alliance perceptions are associated with greater mood improvement across both conditions.	
2	**Vision**

There is growing interest in conversational AI as a tool for everyday wellbeing support, yet the relational quality of these systems remains poorly understood. How a user experiences an AI interaction (whether they feel heard, understood, and engaged) is likely to influence how useful they find it. The working alliance, comprising agreement on goals and tasks and the development of a felt bond between helper and recipient, is one of the most robust predictors of positive outcomes in human-delivered support (Flückiger et al., 2018). Whether this relational quality can be reproduced in human–AI interaction through deliberate design is an open empirical question.

Existing evidence suggests that alliance-like perceptions can form with AI agents but are substantially weaker than in human-delivered contexts (Malouin-Lachance et al., 2025). No controlled experimental study has directly manipulated empathic design features within a conversational agent to isolate their effect on alliance perceptions and mood. This research addresses that gap directly.

The applied impact lies in providing an evidence base for the responsible design of conversational AI tools used in university wellbeing, digital health, and public-facing support contexts, contributing a replicable design framework relevant to policy and practice in Wales and beyond.

3	**Approach**

The study employs a randomised between-subjects design comparing two conversational AI agents: RAIYU, an empathy-engineered agent, and a standard baseline agent. Both use an identical chat window interface so that any differences in outcome are attributable to the empathic design features.

RAIYU is built on a large language model and given specific instructions to communicate using empathic listening strategies, drawing on an additional bank of psychological information to guide its responses toward person-centred, supportive communication. The baseline agent uses the same underlying technology without these instructions or additional information, functioning as a general-purpose helpful assistant.

Validated measures are administered at two time points: the WAI-SR (Hatcher & Gillaspy, 2006) assesses perceived working alliance and the PANAS-SF (Thompson, 2007) assesses mood, both pre- and post-interaction. The SUS (Brooke, 1996) controls for usability confounds. An optional think-aloud protocol (Ericsson & Simon, 1993) generates qualitative data analysed via reflexive thematic analysis (Braun & Clarke, 2006). Quantitative analyses in R include independent-samples t-tests, mixed-design ANOVAs, and regression analyses.

A target sample of 60–80 undergraduate participants will be recruited via the Research Participation Scheme and university mailing lists. An automated behavioural health safety filter (Nelson et al., 2025) monitors all interactions in real time.

4	**Location of research activity** Identify all locations where research activity will take place.

All research activity will take place within a designated university laboratory or research room at the University of Wales Trinity Saint David, IQ Swansea campus. The conversational platform will

be hosted locally on the researcher's password-protected laptop within the laboratory environment. No research activity will take place outside the university setting.

5	Will any of your research be undertaken outside of the UK?	**YES**	Click to continue to appendix 3	**NO**	**Continue to the next section**
6	Have you completed any of appendix 3?	**YES**	☐	**NO**	☐

SECTION F: Research Methods

Will the research activity include:		**YES**	**NO**
1	Use of a questionnaire or similar research instrument?	☒	☐
2	Use of interviews?	☐	☒
3	Use of focus groups?	☐	☒
4	Use of participant diaries?	☐	☒
5	Use of video or audio recording?	☒	☐
6	Use of computer-generated log files?	☒	☐
7	Participant observation with their knowledge?	☐	☒
8	Participant observation without their knowledge?	☐	☒
9	Access to personal or confidential information without the participants' specific consent?	☐	☒
10	Administration of any questions, test stimuli, presentation that may be experienced as physically, mentally or emotionally harmful / offensive?	☐	☒
11	Performance of any acts which may cause embarrassment or affect self-esteem?	☐	☒
12	Investigation of participants involved in illegal activities?	☐	☒
13	Use of procedures that involve deception?	☒	☐
14	Administration of any substance, agent or placebo?	☐	☒
15	Working with live vertebrate animals?	☐	☒
16	Procedures that may have a negative impact on the environment?	☐	☒
17	Other primary data collection methods. Please indicate the type of data collection method(s) below.	☐	☒
Details of any other primary data collection method: **Audio recording**: A concurrent think-aloud protocol will be offered as an optional component of the interaction session. Participants may decline this element without affecting their participation in the main study. This will be clearly described in the participant information sheet. Audio recordings will be made on a password-protected device within the laboratory, transcribed by the researcher, and stored in accordance with UWTSD data protection policies. Recordings will be deleted following transcription and verification. **Deception:** The study employs partial concealment. Participants will be informed that the study evaluates an AI assistant for wellbeing but will not be			

told which condition they have been allocated to (empathy-engineered or baseline) or that the primary focus of the study is the comparison of empathic design features. No false information will be provided; participants will interact with the system exactly as described. Partial concealment is methodologically necessary to prevent demand characteristics and expectancy effects that could bias self-reported alliance and affect scores. A full debrief will be provided immediately following data collection, disclosing the study aims, randomisation process, and condition allocation. Participants will retain a one-month window to withdraw their data following debrief.		

SECTION G: Research Participants

Research participants include people who will be taking part in your research as providers of research data (e.g. being interviewed, answering questionnaires, or doing user testing of hardware or software). This includes animal participants.					
1	Does your research involve human or non-human participants?	**YES**	**Continue this section**	**NO**	**Click to continue to section N**

Participant numbers and source How will you ensure an appropriately convened sample group in order to meet the aims of the research? Give details for subgroups separately, if appropriate. How will any potential pitfalls, for example dual roles or potential for coercion, be addressed?		
1	How many participants will be recruited and how was the number decided upon?	A target sample of 60–80 participants is proposed, with approximately 30–40 allocated to each condition via computer-generated random assignment. Sample size was determined on the basis of power analysis: a minimum of 60 participants provides adequate power (0.80) to detect a medium effect size (Cohen's d = 0.50) at an alpha level of .05 in an independent-samples t-test, appropriate for the primary between-group comparison of working alliance scores. This target is feasible within the MSc dissertation timeline and proportionate to the study's aims. Recruitment will continue toward 80 participants where time permits to maximise statistical power.
2	Who will the participants be?	Participants will be undergraduate students aged 18 years or older, enrolled at the University of Wales Trinity Saint David. All participants must be fluent in English, as the conversational platform and all study materials are in English.
3	How will you identify the participants?	Participants will be recruited through the departmental Research Participation Scheme and via university mailing lists, with course credit available upon completion. To maintain anonymity, participants are assigned a unique

		numeric identifier at the point of consent; no names or student numbers appear in the research dataset. A separate encrypted data sheet, stored exclusively on the researcher's password-protected device, links each identifier to the participant's registration details solely to enable data withdrawal requests within the one-month post-participation window. The participant's identifier is recorded on their debrief sheet for their own reference. Once the withdrawal window closes, the encrypted data sheet will be permanently deleted and all remaining data fully anonymised.
4	Will you be excluding any groups of people, and if so what is the rationale for that?	The following groups will be excluded: individuals under the age of 18 and individuals who are not fluent in English, as all study materials and the conversational platform are in English only. No other groups are excluded, as the study is designed to be accessible to a general undergraduate student population.
5	Will the research involve any element of deception? If yes, please describe why this is necessary and whether participants will be informed at the end of the study	The study uses partial concealment as described in Section F. Participants are informed they are evaluating an AI wellbeing assistant but are not informed of the specific comparison being made or the primary outcome variables being measured. This is necessary to minimise demand characteristics. No false information is provided. A full debrief will be administered immediately after data collection, and participants will be given a one-month window to withdraw their data once fully informed.
6	Describe whether participants will be able to withdraw from the study, and up to what point (eg if data is to be anonymised). If withdrawal is not possible, after a certain point, explain why not.	Participants may withdraw from the study at any point during the laboratory session without providing a reason and without academic or personal penalty. Course credit will be awarded regardless of whether a participant completes or withdraws from the study. Following the debrief, participants may request withdrawal of their data for up to one month after participation by contacting the researcher directly. After this period, data will have been incorporated into anonymised analyses and individual withdrawal will no longer be possible. Participants will be clearly informed of this withdrawal window during the consent process and in the debrief.
7	How long will the participant have to decide whether to take part in the research?	Participants will have the time between receiving the study advertisement through the Research Participation Scheme or mailing lists and the point of booking their laboratory session to decide whether to take part. The participant information sheet will be available prior to booking. At the start of the laboratory session, participants will be given

		additional time to read the information sheet and ask questions before providing written consent, and they may choose not to proceed at that point without penalty.
8	What arrangements have been made for participants who might have difficulties understanding verbal explanations or written information, or who have particular communication needs that should be taken into account to facilitate their involvement in the research?	All study materials (including the participant information sheet, consent form, questionnaires, and debrief) are provided in clear written English. The researcher will be present throughout the laboratory session and available to answer questions or clarify instructions verbally. Participants who require additional support to engage with the written materials are invited to inform the researcher, who will read materials aloud or provide additional explanation as needed. The laboratory computer interface has been designed to be straightforward in use. Participants with accessibility needs that cannot be accommodated within the current setup will be advised that the study may not be suitable for them and will be signposted to other available studies.

SECTION H: Research with Vulnerable People

1	Are your participants classed as vulnerable people?	**YES**	**Click to continue to appendix 4**	**NO**	**Continue to the next section**
2	Have you completed appendix 4?	**YES**	☐	**NO**	☐

SECTION I: Research Involving Animals

1	Are your participants non-human animals?	**YES**	**Click to continue to appendix 5**	**NO**	**Continue to the next section**
2	Have you completed appendix 5?	**YES**	☐	**NO**	☐

SECTION J: Research with the Prison Service or the Probation Service

1	Are your participants staff or offenders within the Prison Service or the Probation service?	**YES**	**Click to continue to appendix 6**	**NO**	**Continue to the next section**
2	Have you completed appendix 6?	**YES**	☐	**NO**	☐

SECTION K: Health and Social Care Research

1	Are your participants selected due to their health or social care needs?	**YES**	Click to continue to appendix 7	**NO**	**Continue to the next section**
2	Have you completed appendix 7?	**YES**	☐	**NO**	☐

SECTION L: Types of Personal Data

Personal Data Does the research activity involve personal data (as defined by the General Data Protection Regulation 2016 and the Data Protection Act 2018)?		**YES**	**NO**
1	Personal data means any information relating to an identified or identifiable natural person ('data subject'). An identifiable natural person is one who can be identified, directly or indirectly, in particular by reference to an identifier such as a name, an identification number, location data, an online identifier or to one or more factors specific to the physical, physiological, genetic, mental, economic, cultural or social identity of that natural person. Any video or audio recordings of participants and transcription of such are considered to be personal data. **If you ticked YES, provide a description of the data and explain why this data needs to be collected:**	☒	☐
The study collects the following personal data: demographic information (age and gender identity); psychometric scale responses (WAI-SR, PANAS-SF, SUS); and, where participants choose to engage in the optional think-aloud procedure, audio recordings and derived transcriptions. Demographic data is required to describe the sample and assess generalisability. Psychometric data constitutes the primary outcome variables for the study. Audio and transcript data are required for the embedded qualitative analysis. All data is pseudonymised at the point of collection using a unique numeric identifier; no names or student numbers appear in the research dataset.			

Special Category Data Does the research involve the processing of special category data (as defined by the General Data Protection Regulation 2016 and the Data Protection Act 2018)?		**YES**	**NO**
2	Racial or ethnic origin	☐	☒
3	Political opinions	☐	☒
4	Religious or philosophical beliefs	☐	☒
5	Genetic, biometric or health data	☒	☐
6	Sex life or sexual orientation	☐	☒
7	Trade union membership	☐	☒

If you ticked **YES** to any of questions L2-L7, provide a description of the data and explain why this special category data needs to be collected to support the research and that it is in the public interest to do so:	
8	The study collects self-reported mood ratings (PANAS-SF) and perceived working alliance ratings (WAI-SR). These data relate to participants' current emotional state and are necessary to measure the study's primary outcomes. Audio recordings and transcriptions generated through the optional think-aloud procedure may contain incidental disclosures of personal or health-related information, which is strictly necessary for the qualitative component of the study. Collection of this data is in the public interest as it contributes to the evidence base for the responsible design of conversational AI tools used in wellbeing contexts. Data is collected under the lawful basis of public task (GDPR Article 6(1)(e)) and scientific research purposes (GDPR Article 9(2)(j); DPA 2018 Schedule 1, Part 1, Paragraph 4).
9	Please describe how the research has been designed to gather the minimum amount of personal data necessary for the specified research purpose:
Only the data necessary to address the stated research questions is collected. No contact details, names, or student numbers are retained in the research dataset. Demographic data collection is limited to age and gender identity. To enable participants to exercise their right to withdraw data within one month of participation, each participant is assigned a unique numeric code at the point of consent. A separate encrypted lookup table, stored on a password-protected device and accessible only to the researcher, links each participant code to their registration details. This lookup table is used solely for the purpose of data withdrawal and will be permanently deleted once the withdrawal window has closed. Participation in the think-aloud protocol is at the participant's discretion and clearly presented as a separate voluntary element that does not affect participation in the core procedure. Audio recordings will be deleted following transcription and verification, with transcripts pseudonymised throughout. The conversation log generated by the AI platform is limited to the text of the exchange and does not record identifying metadata.	
10	Demonstrate that the processing is not likely to cause substantial damage or distress to individuals and that their interests will be safeguarded:
The risk of harm to participants is low. The interaction topic is guided by a standardised prompt inviting discussion of a recent source of stress or a current goal, which is a low-intensity disclosure context appropriate for a general undergraduate population. The researcher is present throughout the laboratory session and available to respond if a participant shows signs of distress. The automated safety filter (Nelson et al., 2025) integrated within the conversational platform monitors interactions in real time and is designed to respond appropriately to any high-risk language. Participants retain the right to withdraw from the interaction at any point without penalty.	
11	Demonstrate that the data will not be used to take any action or make decisions in relation to the individuals concerned:
The data will not be used to make decisions about individual participants. All data will be reported in aggregated, anonymised form. To facilitate the right to data withdrawal, each participant is assigned a unique numeric code at the point of consent. This code appears on the participant's consent record and is provided to them on their debrief sheet to retain. A separate	

encrypted lookup table, stored exclusively on the researcher's password-protected device, links each participant code to their registration details. This table is used solely to locate and delete a specific participant's data in the event of a withdrawal request and will be permanently deleted at the close of the one-month withdrawal window. Beyond this function, pseudonymous identifiers serve no analytical or administrative purpose and do not appear in any dataset, report, or publication.				
12	Data should be anonymised wherever possible, either at point of capture or once collated. If the data cannot be anonymised, and if pseudonymisation is not possible, demonstrate why the that is possible. If special category data cannot be anonymised, explicit consent to share that data must normally be obtained.			
Data will be pseudonymised at the point of collection using a unique numeric identifier. The lookup table linking this identifier to the participant's registration details will be stored separately, in an encrypted file accessible only to the principal researcher. Transcripts generated from the think-aloud procedure will be pseudonymised prior to analysis. Quantitative data will be reported only in aggregated form (group means, confidence intervals). Illustrative quotations from qualitative analysis will be attributed only to non-traceable pseudonyms (e.g., Participant 12, Empathy Condition) and will be checked for accidental personal disclosures before inclusion in any publication. Full anonymisation of the dataset will be completed at the end of the analysis phase, at which point the lookup table will be securely deleted.				
13	If special category data cannot be anonymised, will you obtain explicit consent to share that data?	YES	NO	N/A
		☒	☐	☐

SECTION M: Participant Information and Consent

Information for participants It is normally expected that those participating in research should do so having given their informed consent. If this is not appropriate or practical, explicit justification must be provided, in line with prevailing disciplinary conventions. Please note however, that informed consent to participate in the research is separate from consent to process (e.g. collect, analyse, store, reuse or share) personal data, as the University's lawful basis for processing personal data in research activities is part of its 'public task'.		**YES**	**NO**	**N/A**
1	Will you obtain written consent for participation?	☒	☐	☐
2	Will you describe the main research procedures to participants in advance, so that they are informed about what to expect?	☒	☐	☐
3	Will you tell participants that their participation is voluntary?	☒	☐	☐
4	Will you explain to participants that refusal to participate in the research will not affect their treatment or education (if relevant)?	☒	☐	☐
5	If the research is observational, will you ask participants for their consent to being observed?	☐	☐	☒

6	Will you tell participants that they may withdraw from the research at any time and for any reason?	☒	☐	☐
7	Will participants be made aware of how you will treat any information that may reasonably by considered as confidential?	☒	☐	☐
8	If your research requires your participants to sign up for a service provided by an external company or organisation will you inform them that their data may be stored and processed outside UWTSD?	☐	☐	☒
9	Will you give participants the UWTSD Privacy Notice for Research?	☒	☐	☐

If you have answered **NO** to any of questions M1-M8, please provide justification detailing why this is necessary for the research.

10	M8 is marked N/A because the study uses a locally hosted platform on the researcher's laptop within the university laboratory. No participant data is transmitted to or stored on external company servers.

Informed consent: Describe the arrangements to inform potential participants, before providing consent, of what is involved in participating. Describe the arrangements for participants to provide full consent before the research begins. If gaining consent in this way is inappropriate, explain how consent will be obtained and recorded.

11	Participants will be recruited through the Research Participation Scheme or university mailing lists and directed to a full Participant Information Sheet (PIS) prior to booking a laboratory session. The PIS describes the study duration, the nature of the interaction (evaluating a conversational AI agent), the types of data collected, the optional think-aloud component, the use of partial concealment, and the post-session debrief process. It also describes participants' rights, including the right to withdraw at any point and to request data deletion within one month of participation. At the start of the laboratory session, participants will be given time to read the PIS and ask questions before providing informed consent digitally via Qualtrics. The consent form confirms that the participant is aged 18 or over, has read and understood the PIS, understands that certain details about the study have been withheld to prevent bias, consents to the optional audio recording if they wish to participate in the think-aloud procedure, and understands the limits of confidentiality. Consistent with ethical standards for psychological research, partial concealment of the specific hypotheses and condition allocation is disclosed upfront in the PIS as a methodological necessity, ensuring consent is as informed as possible without compromising scientific validity. Immediately following the post-interaction measures, participants receive a full written debrief disclosing the study's aims, the nature of the comparison, and their condition allocation. The debrief details the one-month data withdrawal window and provides wellbeing signposting where appropriate.

Confidentiality: Confidential information refers to information that can be related to an identifiable individual, whether living or deceased, which is not in the public domain and which is given with the expectation that it will be kept confidential. Under Common Law when an individual entrusts a researcher or research team with confidential information, the team must handle this in line with 'reasonable expectations'. In other words, confidential information should only normally be shared when there would be 'no surprises' for the individuals concerned. Where participants would not expect the researcher to be sharing their confidential information with others, researchers can manage their expectations by informing them of their intentions (e.g. in the participant information sheet or during discussions about participation) and asking them if they are happy with these plans. They should understand what is being proposed and what this might mean for them, before they decide whether the researcher can share their confidential information with others.	
Please detail how you will treat confidential information:	
12	Participants will be informed through the PIS and consent form that their data will be treated with full confidentiality. All data will be pseudonymised using a unique numeric identifier, and no names or student numbers will appear in any research dataset, report, or publication. The lookup table linking the identifier to the participant's registration details will be stored in a separately encrypted file accessible only to the principal researcher, used solely for processing data withdrawal requests. The only exception to confidentiality arises if a participant discloses information during the session that indicates imminent serious risk of harm to themselves or others. This limit is clearly described in the PIS and consent form. In such circumstances the researcher would consult with the supervisor and, if appropriate, the University Safeguarding Officer. The automated safety filter integrated within the conversational platform provides an additional layer of monitoring throughout the interaction. Think-aloud audio recordings will be transcribed by the researcher and the audio files deleted following verification. Transcripts will be pseudonymised and stored on an encrypted device accessible only to the researcher.
Copies of any written consent form, written information and all other explanatory material should accompany this application. If you have included these then please list the documents included here.	
13	Study information sheet Consent form (including optional think-aloud audio recording consent item) Debrief form Recruitment advertisement

14	Consent form attached?	**YES** ☒	**NO** ☐	**N/A** ☐

SECTION N: Research with Potential Military Application

1	Does your research have any potential military application?	**YES**	Click to continue to appendix 8	**NO**	**Continue to the next section**

2	Have you completed appendix 8?	**YES**	☐	**NO**	☐

SECTION O: Use of Artificial Intelligence Tools

Use of generative AI and other AI-based tools There are a number of ethical issues relating to the use of generative AI and other AI-based tools, including data security, output bias, and output errors. In general you should assume that any generative AI or other AI-based tool that you pay to use is more secure and accurate than one which is free to use. Will the research use AI tools to do any of the following:		**YES**	**NO**	**N/A**
1	Capture data?	☒	☐	☐
2	Generate and/or edit text?	☒	☐	☐
3	Generate graphics?	☐	☒	☐
4	Generate video?	☐	☒	☐
5	Generate sound/music?	☐	☒	☐
6	Generate code?	☐	☒	☐
7	Generate synthetic data?	☐	☒	☐
8	Analyse data?	☐	☒	☐
9	Analyse personal or sensitive data?	☒	☐	☐
10	Classify data?	☒	☐	☐
11	Do you intend to use the free version of any AI tools?	☐	☒	☐
If you have answered **YES** to any of questions O1-O11, please provide justification detailing why this is necessary for the research.				
12	This study employs two locally hosted conversational AI agents as the experimental stimuli. Both agents generate text responses during the interaction session (O2) — this is the object of study rather than a research tool used by the investigator. Conversational data, comprising participant text input and AI responses, is captured as computer-generated log files (O1). The AI processes participant text input in real time to generate contextually relevant responses; this constitutes processing of personal data (O9) in the sense that the model responds to what the participant writes, though no data is transmitted externally at any point. The integrated behavioural health safety filter (Nelson et al., 2025) classifies conversational content for indicators of distress in real time (O10), which is essential for participant safety. No free AI tools are used (O11); both agents are locally hosted using a specific model configuration within the researcher's controlled laboratory environment.			

Section P: Research Data Management Plan

1	**Data Summary**
The study will generate the following data types: pseudonymised psychometric questionnaire data (WAI-SR pre and post, PANAS-SF pre and post, SUS, demographic variables) collected via Qualtrics; computer-generated interaction logs comprising the text of conversational exchanges from the locally hosted AI platform; and, where participants opt in, audio recordings and derived transcriptions from the think-aloud procedure.	
2	**Responsibilities**
The principal researcher (Rhiannon Sigrist) holds primary responsibility for data management throughout the project, including collection, storage, analysis, and archiving. The primary supervisor (Dr Paul Hutchings) will provide oversight. No other individuals will have access to identifiable or pseudonymised research data. All team members will adhere to UWTSD data management policies.	
3	**Assessment of existing data**
No existing datasets will be used. All data will be collected prospectively. Validated psychometric instruments (WAI-SR, PANAS-SF, SUS) are established measures whose properties are well documented in the existing literature; no adaptations to these instruments require additional data sourcing.	
4	**Management and curation of data**
Questionnaire data will be exported from Qualtrics in anonymised CSV format, with participant identifiers replaced by unique numeric codes prior to export. Interaction logs will be extracted from the locally hosted platform in pseudonymised form. Transcripts will be typed by the researcher and checked against the audio recording before the recording is deleted. All data files will be named using a consistent convention referencing the condition, time point, and participant code. An accompanying codebook will document variable names, scale items, and scoring procedures.	
5	**Quality assurance of data**
Questionnaire data will be checked for completeness and validity upon export. Incomplete or invalid responses (e.g., all items rated identically) will be noted and handled according to pre-specified criteria. Interaction log data will be reviewed for technical anomalies (e.g., truncated sessions) prior to analysis. Transcripts will be verified against audio recordings. Quantitative analysis scripts will be pre-registered on the Open Science Framework (OSF) prior to data collection to ensure analytical transparency.	
6	**Short-term data storage**

During the study, questionnaire data will be stored on the Qualtrics platform under the university account, protected by the platform's security infrastructure. Interaction logs will be stored on the researcher's password-protected, encrypted laptop within an encrypted container (VeraCrypt or equivalent). Audio recordings will be stored on the same encrypted device and deleted following transcription. No data will be stored on portable media or shared cloud services outside university-approved systems.

7	**Long-term data storage and preservation**

Following analysis, the fully anonymised dataset (pseudonymised identifiers replaced with sequential codes, lookup table deleted) will be retained for five years in accordance with UWTSD research data policy. The dataset will be archived in the UWTSD Research Data Repository or an approved equivalent (such as the OSF) upon project completion. Only aggregated statistics and anonymised illustrative quotations will be included in any public-facing outputs.

8	**Documentation and metadata**

The dataset will be accompanied by a README file documenting the study design, variable descriptions, scoring procedures, and data collection dates. The thematic codebook for qualitative analysis will be included in the archival package. Pre-registration documentation from the OSF will be linked to the dataset to ensure methodological transparency.

9	**Data sharing**

An anonymised archival package will be made available through the UWTSD Research Data Repository or OSF upon completion of the project. This package will contain aggregated quantitative results and the AI-generated side of the conversational logs only; all participant utterances will be redacted to prevent re-identification. The data will be released under an open access licence consistent with UWTSD policy. No charges will be made for access. No participant-level psychometric data or identifiable transcripts will be shared publicly.

10	**Copyright and intellectual property ownership**

Copyright in all data generated by the research will be held by the University of Wales Trinity Saint David in accordance with the standard terms governing student research. The conversational AI platform and system prompt design developed for this study will be documented in sufficient detail for replication, and findings will be disseminated through open access outputs wherever possible.

11	**Resources**

The study requires secure data storage on the university Qualtrics platform (existing institutional licence), a personal encrypted laptop for local data processing, and an OSF account for pre-registration and archiving (free). No additional financial resources are required for data

management. Costs are entirely covered by existing university infrastructure and the researcher's personal equipment.

SECTION Q: Risk Assessment

Outline any anticipated risks that may adversely affect any of the participants, the researchers and/or the University, and the steps that will be taken to address them. If you have completed a full risk assessment (for example as required by a laboratory, or external research collaborator) you may append that to this form.			
1	Full risk assessment completed and appended?	**YES**	**NO**
		☐	☒
Risks to participants, researchers, and the University			
2	**Social risks:** *disclosures that could affect participants standing in the community, in their family and their job.*	**YES**	**NO**
		☒	☐
Risk: Participants may disclose personal or sensitive information during the interaction session (such as relationship difficulties, academic stress, or mental health concerns) that could, if not properly handled, affect their standing or wellbeing.	*How will you address any risks?* All data is pseudonymised and no identifiable information is included in any publication or report. Participation in the think-aloud protocol is at the participant's discretion and clearly presented as a separate voluntary element that does not affect participation in the core procedure. The participant information sheet and debrief both make clear that disclosures within the interaction are confidential, subject to the limits described. The researcher is present throughout the session and available to respond supportively if a participant becomes distressed.		
3	**Legal risks:** *activities that could result in the participant, researchers and / or University committing an offence; activities that might lead to a participant disclosing criminal activity to a researcher which would necessitate reporting to enforcement authorities; activities that could result in a civil claim for compensation*	**YES**	**NO**
		☒	☐
Risk: A participant may disclose information during the interaction session indicating imminent risk of serious harm to themselves or others, which could necessitate a safeguarding referral.	*How will you address any risks?* The participant information sheet clearly states the limits of confidentiality. The integrated safety filter (Nelson et al., 2025) monitors conversational content in real time and redirects participants to crisis resources if high-risk language is detected. In the event of a disclosure indicating imminent serious harm, the researcher will consult the supervisor and, where appropriate, the University Safeguarding Officer. All decisions to break confidentiality will be made in line with University safeguarding procedures.		

4	**Economic harm:** *financial harm to participant, researcher and / or University through disclosure or other event.*	YES ☐	NO ☒
Risk: *(this box should expand as you type)*	*How will you address any risks?*		
5	**Reputational risk**: *damage to public perception of University or the University / researchers' reputation in the eyes of funders, the research community and / or the general public.*	YES ☒	NO ☐
Risk: If the conversational platform generates unsafe, misleading, or inappropriate responses, this could reflect negatively on the researcher and the University.	*How will you address any risks?* The platform will undergo pilot testing prior to data collection to identify and address any problematic response patterns. The system prompt for the empathy-engineered condition enforces clear non-clinical boundaries and explicitly refuses diagnostic or medication-related queries. The automated safety filter (Nelson et al., 2025) provides an additional layer of oversight. All study materials clearly describe the platform as a research prototype and not a clinical service. Findings will be reported transparently, including any limitations of the platform's performance.		
6	**Safeguarding risks:** *Risk to young people, vulnerable adults and / or researcher from improper behaviour, abuse or exploitation. Risk to researcher of being in a comprising situation, in which there might be accusations of improper behaviour.*	YES ☒	NO ☐
Risk: Participants may disclose distress or crisis-level ideation during the interaction. There is also a general risk of researcher secondary trauma from reviewing transcripts involving distressing content.	*How will you address any risks?* The study invites participants to discuss a recent source of stress or a current goal, which is a low-intensity disclosure context appropriate for a general undergraduate population. While acute distress is not anticipated, the researcher is present throughout the laboratory session and trained to respond supportively if a participant shows signs of distress during or after the interaction. The automated behavioural health safety filter (Nelson et al., 2025) monitors all conversational content in real time. If language indicating serious distress or risk is detected, the interaction is immediately terminated and the participant is redirected to crisis support resources. General wellbeing signposting — specifically the UWTSD Student Wellbeing Service and Student Assistance Programme — is provided to all participants at the end of the session as routine		

	information. Targeted crisis resources are provided where the safety filter has been triggered or where a participant has visibly become distressed during the session. The researcher will schedule regular supervisor check-ins during the data collection and analysis phases to monitor their own wellbeing when reviewing transcripts. A no-DBS requirement applies as participants are adults aged 18 or over and not a designated vulnerable group.

		YES	NO
7	**Sensitive, embarrassing or upsetting topics**. *If research activity may include sensitive, embarrassing or upsetting topics or issues likely to disclose information requiring further action (e.g. criminal activity), give details of the procedures to deal with these issues, including any support/advice (e.g. helpline numbers) to be offered to participants. Note that where applicable, consent procedures should make it clear that if something potentially or actually illegal is discovered in the course of a project, it may need to be disclosed to the proper authorities.*	☒	☐

The study invites participants to discuss a recent source of stress or a current goal, which may surface mildly upsetting content for some participants. The following procedures are in place:

- The researcher remains present throughout the session and is available to respond if distress occurs
- The automated safety filter monitors conversational content in real time and redirects to crisis resources where indicated
- Participants may exit the interaction or the study at any time without penalty
- General wellbeing signposting is provided to all participants at the end of the session
- Targeted crisis resources are provided where the safety filter has been triggered or where a participant has become visibly distressed, including University Counselling (wellbeingsupport@uwtsd.ac.uk), Samaritans (116 123), Shout (text SHOUT to 85258), NHS 111 (option 2), and PAPYRUS HOPELINEUK (0800 068 4141)

	Health and safety risks		
8	**Location hazards**: *for example: fire; visiting or working in participant's homes; working in remote locations and in high crime areas; overseas travel; hot, cold or extreme weather conditions; working on or by water. Also hazardous work locations, such as construction sites, confined spaces, roofs or laboratories. For overseas travel, researchers will need to check country / city specific information, travel health requirements and consider emergency arrangements as part of their research planning. In all cases the University's International Travel Policy must be followed.*	YES ☒	NO ☐

Risk: The study takes place in a university laboratory. Standard location hazards apply (fire, trip hazards, access).	*How will you address any risks?* The laboratory will be a designated university room compliant with standard health and safety requirements. The researcher will be familiar with fire evacuation procedures and emergency contacts for the building. No additional location-specific hazards are anticipated.

9	**Activity hazards:** *for example: potentially mentally harmful activities; distressing and stressful work and content; driving; tripping, or slipping; falling from height; physically demanding work; lifting, carrying, pushing and pulling loads; night time and weekend working.*	YES ☒	NO ☐
Risk: Discussing personal stressors may cause mild to moderate emotional distress for some participants.	*How will you address any risks?* Discussing personal stressors may cause mild emotional discomfort for some participants. Participants are informed in advance that the interaction involves discussing a source of stress or a current goal and retain the right to exit the interaction or the study at any point without penalty. The researcher is present and able to provide an immediate supportive response and signposting if distress is observed.		
10	**Machinery and equipment:** *for example: ergonomic hazards, including computer workstations and equipment; contact with electricity; contact with moving, rotating, ejecting or cutting parts in machinery and instruments; accidental release of energy from machines and instruments.*	YES ☒	NO ☐
Risk: Ergonomic hazards from computer use (posture, screen time) during a 10-minute interaction session.	*How will you address any risks?* Sessions are brief (approximately 10 minutes of interaction plus questionnaire completion). Standard laboratory workstation ergonomic provision applies. No specific additional mitigation is required.		
11	**Chemicals and other hazardous substances**: *the use, production, storage, waste, transportation and accidental release of chemicals and hazardous substances; flammable, dangerous and explosive substances; asphyxiating gases; allergens; biological agents, blood and blood products.*	YES ☐	NO ☒
Risk: *(this box should expand as you type)*	*How will you address any risks?*		
12	**Physical agents**: *for example: excessive noise exposure, hand-arm vibration and whole body vibration; ionising radiation; lasers; artificial optical radiation and electromagnetic fields*	YES ☐	NO ☒
Risk: *(this box should expand as you type)*	*How will you address any risks?*		
13	**Environmental risks:** *for example: accidental spillage of pollutants, damage to local ecosystems*	YES ☐	NO ☒
Risk: *(this box should expand as you type)*	*How will you address any risks?*		

SECTION R: Disclosure and Barring Service

Disclosure and Barring Service				
1	If the research activity involves children or vulnerable adults, a Disclosure and Barring Service (DBS) certificate must be obtained before any contact with such participants.	**YES**	**NO**	**N/A**
	Does your research require you to hold a current DBS Certificate?	☐	☐	☒
2	If YES, please give the certificate number. If the certificate number is not available please write "Pending"; in this case any ethical approval will be subject to providing the appropriate certificate number.			

SECTION S: Declaration

The information which I have provided is correct and complete to the best of my knowledge. I have attempted to identify any risks and issues related to the research activity and acknowledge my obligations and the rights of the participants.

In submitting this application I hereby confirm that I undertake to ensure that the above named research activity will meet the University's Research Ethics and Integrity Policy which is published on the website: https://www.uwtsd.ac.uk/research/research-ethics/

1	**Name of applicant**	Rhiannon Sigrist
2	**Electronic or written signature of applicant**:	*Rhiannon Sigrist*
3	**Date**	15/07/2026
4	**Name of Director of Studies**	Assoc Prof Paul B. Hutchings
5	**Electronic or written signature of Director of Studies**:	Paul B. Hutchings
6	**Date**	15/06/26

Checklist: Please complete the checklist below to ensure that you have completed the form according to the guidelines and attached any required documentation:

1	☒	I have read the guidance notes supplied before completing the form.
2	☒	I have completed all relevant sections of the form in full.
3	☒	I have attached a copy of the Participant Information Sheet (PIS) or Consent Form

4	☐	I have attached a completed copy of the Trusted Research and Innovation Checklist (where applicable)
5	☐	I have attached a full risk assessment (where appropriate)
6	☒	I understand that it is my responsibility to ensure that the above named research activity will meet the University's Research Ethics and Integrity Code of Practice
7	☒	I understand that it is my responsibility to ensure that the above named research activity will meet the University's Research Data Management Policy
8	☒	I understand that it is my responsibility to ensure that the above named research activity will meet the University's Research Export Control Policy
9	☒	I understand that it is my responsibility to ensure that the above named research activity will meet the University's Open Access Research Policy

10	I have completed the following appendices:	1	☐	2	☐	3	☐	4	☐
		5	☐	6	☐	7	☐	8	☐

All communications relating to this application during its processing must be in writing and emailed to pgresearch@uwtsd.ac.uk , with the title 'Ethical Approval' followed by your name.

Prifysgol Cymru
Y Drindod Dewi Sant
University of Wales
Trinity Saint David

PG2 FORM

APPENDIX 1 – Grant or Contract Funding

A1.1	Has your research been awarded external funding?	**YES**	☐	**NO**	☐
A1.2	Have you applied for funding / awaiting decision?	**YES**	☐	**NO**	☐
A1.3	Are you intending to apply for funding?	**YES**	☐	**NO**	☐
A1.4	Funding Body				
A1.5	Lead institution (grant holder)				
A1.6	If awarded, grant reference				
A1.7	If awarded, grant value				
A1.8	If awarded, start date and duration				

Return to the main form

APPENDIX 2: External Collaborative Research Activity

INSTITUTION 1					
A2.1	Name				
A2.2	Researcher name and position				
A2.3	Role in project (e.g. co-investigator)				
A2.4	Contact e-mail address				
A2.5	**External Ethical Approval** *Please detail the ethical approval required and arrangements for this partner. See Section 14 of the UWTSD Research Ethics and Integrity Policy. If approved, please supply confirmation.*	**Obtained?**	**YES**	**NO**	**PENDING**
			☐	☐	☐

INSTITUTION 2					
A2.6	Name				
A2.7	Researcher name and position				
A2.8	Role in project (e.g. co-investigator)				
A2.9	Contact e-mail address				
A2.10	**External Ethical Approval** *Please detail the ethical approval required and arrangements for this partner. See Section 14 of the UWTSD Research Ethics and Integrity Policy. If approved, please supply confirmation.*	**Obtained?**	**YES**	**NO**	**PENDING**
			☐	☐	☐

INSTITUTION 3					
A2.11	Name				
A2.12	Researcher name and position				
A2.13	Role in project (e.g. co-investigator)				
A2.14	Contact e-mail address				
A2.15	**External Ethical Approval** *Please detail the ethical approval required and arrangements for this partner. See Section 14 of the UWTSD Research Ethics and Integrity Policy. If approved, please supply confirmation.*	**Obtained?**	**YES**	**NO**	**PENDING**
			☐	☐	☐

Add additional collaborators as needed

Return to the main form

APPENDIX 3: Research Undertaken Outside of the UK

A3.1	**Research activity outside of the UK** If research activity will take place overseas, you are responsible for ensuring that local ethical considerations are complied with and that the relevant permissions are sought. Specify any local guidelines (e.g. from local professional associations/learned societies/universities) that exist and whether these involve any ethical stipulations beyond those usual in the UK (provide details of any licenses or permissions required). Also specify whether there are any specific ethical issues raised by the local context in which the research activity is taking place, for example, particular cultural and/or legal sensitivities or vulnerabilities of participants. Describe the measures you have taken to comply with these and provide evidence of approvals etc. If you live in the country where you will do the research then please state this. If you have a local research supervisor or research advisor then please give their contact details here.
(this box should expand as you type)	
A3.2	**Responsible Research and Innovation** If research is taking place overseas please refer to the UWTSD Trusted Research and Innovation Policy and complete the Trusted Research and Innovation Checklist and submit it with this form.
A3.3	Confirm that you have read the Trusted Research and Innovation Policy ☐

If you are a postgraduate research student working with a research supervisor or research advisor at your local university then you do not need to fill in any more sections of this appendix.
If you are working in a research collaboration with an overseas partner please complete the remaining sections of this appendix.

Return to the main form

Organisational Details	
A3.4	Name of Organisation
A3.5	Country where organisation is based:
A3.6	Registration number and name of registration body (examples of relevant registration bodies are government, tax, corporation, charity, educational authority):
(this box should expand as you type)	
A3.7	What is the legal status of the organisation? (for example: public university, private university, charity, private company, non-profit making company, government research organisation, independent research organisation) Please include the date of foundation. Note: We may be required to ask for evidence of this at a later stage.

(this box should expand as you type)	
A3.8	Is your organisation affiliated to any other organisation? (i.e. is the organisation legally part of a larger organisation or a group of larger organisations).
(this box should expand as you type)	
A3.9	Name(s) of investigators in the organisation who will be responsible for the collaborative research and innovation
(this box should expand as you type)	

About New Partners			
A3.10	Why does a partner want to work with you?		
(this box should expand as you type)			
A3.11	What are they expecting in return for their financial support or involvement?		
(this box should expand as you type)			
A3.12	Is the organisation associated with a country which may be viewed as hostile to the UK or one which has different democratic and ethical values from our own? If you have ticked YES, please briefly describe below how you will maintain the UK's democratic and ethical values within the research.	**YES** ☐	**NO** ☐
(this box should expand as you type)			
A3.13	Has due diligence into the partner identified any involvement in research on behalf of the military or police with links to a hostile state? If you have ticked YES, please summarise below how you will prevent information from this research from being released to these organisations.	**YES** ☐	**NO** ☐
(this box should expand as you type)			
A3.14	Set within the context of any information gained from due diligence, could your research be misused or have unintended applications which would be negative? If you have ticked YES, please summarise below how this might occur.	**YES** ☐	**NO** ☐

(this box should expand as you type)			
A3.15	Are there any legal, regulatory or university policy constraints on undertaking your research with this partner? If you have ticked YES, please summarise these below.	YES	NO
		☐	☐
(this box should expand as you type)			
A3.16	Having considered the answers to questions A3.10-A3.15, are there potential reputational or ethical risks to you or the university? If you have ticked YES, please summarise these risks below.	YES	NO
		☐	☐
(this box should expand as you type)			

About Research Relationships		YES	NO
A3.17	Are the terms of any proposed Memorandum of Understanding (MoU) in keeping with the expectations of your department and university?	☐	☐
A3.18	Are you providing existing intellectual property (IP), research data, confidential or personally identifiable data to the project? If so, how is this going to be protected?	☐	☐
(this box should expand as you type)			
A3.19	Who will own any IP that is generated?		
(this box should expand as you type)			
A3.20	What plans do you have in place for protecting the resulting IP?		
(this box should expand as you type)			
A3.21	What contractual requirements are you able to put in place to protect the interests of your academic institutions?		
(this box should expand as you type)			
A3.22	What access will the research partner have to your IT network? If they do have access, what broader visibility might this provide?		

(this box should expand as you type)	
A3.23	Is there any physical separation or protection required between research in similar fields?
(this box should expand as you type)	

About Existing Partners			
A3.24	Would proceeding with the research raise potential conflicts of interest with existing research partners? If you have ticked YES, please explain these conflicts below.	YES	NO
		☐	☐
(this box should expand as you type)			
A3.25	Have you spoken with your existing partners about any potential conflict of interest? If you have ticked YES, please briefly summarise their response(s) below.	YES	NO
		☐	☐
(this box should expand as you type)			
A3.26	Does any non-disclosure agreement for this project include an expectation that you will need to provide visibility to existing partners? If you have ticked YES, please explain below how this will be managed.	YES	NO
		☐	☐
(this box should expand as you type)			
A3.27	Will this research breach any existing contractual agreements that you, your department or university already have? If you have ticked YES, please explain below how this will be managed.	YES	NO
		☐	☐
(this box should expand as you type)			

Return to the main form

APPENDIX 4: Research with Vulnerable People

Will the research involve data processing involving the following data subjects:		YES	NO
A4.1	Children under 16	☐	☐
A4.2	Adults with learning disabilities	☐	☐
A4.3	Adults with other forms of mental incapacity or mental illness	☐	☐
A4.4	Adults in emergency situations	☐	☐
A4.5	Prisoners or young offenders	☐	☐
A4.6	Those who could be considered to have a particularly dependent relationship with the investigator, e.g. members of staff, students	☐	☐
A4.7	People who have not given their explicit consent to participate in the project	☐	☐
A4.8	People engaged in illegal activities	☐	☐
A4.9	Any other people who can be considered vulnerable in the context of this project	☐	☐
Please justify the inclusion of the above people (if applicable), explaining why the research cannot be conducted on non-vulnerable people.			
A4.10	*(this box should expand as you type)*		
Please give details of extra steps taken to assure the protection of vulnerable people participating in the research. You should describe any arrangements to be made for obtaining consent from a legal representative.			
A4.11	*(this box should expand as you type)*		
It is the researcher's responsibility to check whether a DBS check (or equivalent) is required and to obtain one if it is needed. Please confirm:			

		YES	NO	
A4.12	DBS Check required?	☐	☐	
		YES	NO	N/A
A4.13	DBS certificate obtained?	☐	☐	☐

Return to the main form

APPENDIX 5: Research Involving Animals

The use of animals in research is tightly governed and monitored by law and by the Home Office, specifically under the Animals (Scientific Procedures) Act 1986 and its accompanying codes of practice and processes. The Act regulates procedures that are carried out on protected animals for scientific or educational purposes that may cause pain, suffering, distress or lasting harm. Under the Act, a protected animal is 'any living vertebrate, other than man, and any living cephalopod'. The University does not hold a Home Office issued Establishment Licence and therefore researchers must ensure that any proposed work involving animals is either exempt from the Act or falls outside of its scope. Exempted research, for example involving routine veterinary practice or clinical veterinary research, may still require external ethical approval from the Royal Society of Veterinary Surgeons. When undertaking collaborative research involving animals, the UWTSD Research Ethics Committee must be assured that the partner institution(s) conform to the appropriate regulatory framework or legislation, and have approval from their local Research Ethics Committee. Any such work must follow the guidance of the National Centre for the Replacement, Reduction and Refinement of Animals in Research, known as the 3Rs. In all cases, advise must be sought from INSPIRE for any proposed research involving animals, either at UWTSD, or by research collaborators in partnership with UWTSD. This should be at the very early stages of research design, so that an appropriate due diligence exercise can be undertaken.

Will the research involve any work with any of the following animals (including observation in their natural habitat)		**YES**	**NO**
A5.1	Any living vertebrates, other than humans?	☐	☐
A5.2	Any living cephalopods?	☐	☐
A5.3	Has INSPIRE undertaken an initial due diligence exercise?	☐	☐
A5.4	Due diligence assessment appended to this application?	☐	☐

Please provide a summary of the due diligence exercise, noting arrangements for compliance with the necessary regulatory frameworks and legislation. Please append the full due diligence assessment from INSPIRE.

A5.5	*(this box should expand as you type)*

Return to the main form

APPENDIX 6: Research with Staff or Offenders in Prison Establishments and Probation Services

Research with staff and/or offenders in prison establishments, the Probation Service or within His Majesty's Prison and Probation Services requires additional approvals.			
Will the research involve any of the following?		**YES**	**NO**
A6.1	Staff in prison establishments	☐	☐
A6.2	Offenders in prison establishments	☐	☐
A6.3	The Probation Service	☐	☐
A6.4	His Majesty's Prison and Probation Services	☐	☐
All researchers wanting to conduct research with staff and/or offenders in prison establishments, the Probation Service regions or within HM Prison and Probation Service (HMPPS) Headquarters are required to formally apply for research approval to the HMPPS National Research Committee (NRC). Before doing so, INSPIRE must undertake an initial due diligence exercise. If you have not contacted INSPIRE, please do so immediately. If YES to any of the above, please confirm:			
A6.5	Has INSPIRE undertaken an initial due diligence exercise?	☐	☐
A6.6	Have you applied to the HMPPS National Research Committee (NRC)?	☐	☐
A6.7	Has HMPPS National Research Committee approval been granted? If yes, please append to this application.	☐	☐
A6.8	Due diligence assessment appended to this application?	☐	☐
Please provide a summary of the due diligence exercise, noting arrangements for compliance with the necessary regulatory frameworks and legislation. Please append the full due diligence assessment from INSPIRE.			
A6.9	*(this box should expand as you type)*		

Return to the main form

APPENDIX 7: Health and Social Care Research

All health and social care research must have a sponsor as defined by the Health Research Authority UK Policy Framework for Health & Social Care. The University will under no circumstances act as the sponsor of health and social care research (i.e. that which requires approval by a NHS Research Ethics Committee). In all cases, therefore, an appropriate collaborating partner must lead the research. Any health and social care research therefore requires an initial due diligence exercise, which will be undertaken by INSPIRE. The purpose of the due diligence exercise is to ensure that the University has the capability to undertake the research, either independently or in collaboration with research partners. This will include advice on suitable collaborative arrangements and the appropriate involvement of UWTSD researchers in the proposed research study. This is likely to apply to the following areas:

Will the data processing involve:		**YES**	**NO**
A7.1	Participants who have been recruited on the basis of medical and health conditions, life stage or other physical or mental attribute.	☐	☐
A7.2	Participants who will engage in physical exercise, activity or exertion of any kind, or who use exercise equipment, assistive devices or other apparatus for the purposes of the research study. This includes interventions that are minimally invasive.	☐	☐
A7.3	Patients and users of the NHS (including NHS patients treated in the private sector).	☐	☐
A7.4	Individuals identified as potential participants because of their status as relatives or carers of patients and users of the NHS.	☐	☐
A7.5	Patients who are cared for in private and voluntary sector nursing homes.	☐	☐
A7.6	Confidential patient information.	☐	☐
A7.7	Clinical trial of a medicinal product or medical device.	☐	☐
A7.8	Access to organs or other bodily material of past and present NHS patients, or any material consisting of or containing human cells.	☐	☐
A7.9	Use of human tissue (including non-NHS sources) where the collection is not covered by a Human Tissue Authority licence.	☐	☐
A7.10	Foetal material and IVF involving NHS patients.	☐	☐
A7.11	The recently deceased under NHS care.	☐	☐
A7.12	NHS staff or premises.	☐	☐
A7.13	Practising midwives conducting a clinical trial.	☐	☐
A7.14	Investigational medicinal product.	☐	☐
A7.15	Medical devices that are not CE-marked or CE-marked medical devices that have been modified or are being used for a new purpose.	☐	☐
A7.16	Exposure to ionising radiation.	☐	☐
A7.17	Other health and social care related research.		
Please specify *(this box should expand as you type)*		☐	☐

Researchers proposing to work on projects involving health and social care research must check whether their study meets the HRA definition of research and whether it should undergo ethics review by a NHS Research Ethics Committee. If you have answered **YES** to any of questions A7.1-A7.17 then this is likely to be the case and sponsorship by a third party will be required. This can be checked via the Health Research Authority's Decision Tool : https://www.hra-decisiontools.org.uk/research/				
A7.18	Does the research meet the HRA definition of research?	**YES**	**NO**	**N/A**
		☐	☐	☐
A7.19	Is the Research being sponsored by a collaborating partner?	**YES**	**NO**	**N/A**
		☐	☐	☐
A7.20	Has NHS ethics approval been granted?	**YES**	**NO**	**N/A**
		☐	☐	☐
A7.21	Has INSPIRE undertaken an initial due diligence exercise?	**YES**	**NO**	
		☐	☐	
Please provide a summary of the due diligence exercise, noting arrangements for compliance with the necessary regulatory frameworks and legislation. Please append the full due diligence assessment from INSPIRE. If you have not contacted INSPIRE, please do immediately.				
A7.22	*(this box should expand as you type)*			

A7.23	Have you been granted NHS ethical approval?	**YES**	Click to continue to Section R	**NO**	Return to the main form

APPENDIX 8: Export Control for Research with Potential Military End Uses

Export controls apply to the physical, electronic or oral transmission outside the UK of applied research, sensitive technology, software and equipment that could potentially be misused, either directly or indirectly, for military purposes. These legal controls are a serious obligation and UWTSD researchers must actively manage the risks of these being misused to fuel conflict, threaten national security, support terrorism and crime, violate human rights or proliferate Weapons of Mass Destruction. UWTSD's Export Control Policy provides a clear six step process to determine if your research is covered by export controls. Please complete the following screening to judge if you need to compete a full export control assessment. If you are in any doubt, then please contact INSPIRE. Any research in UWTSD involving the following high-risk areas must be screened. **Will the research involve:**		
Defence	**YES**	**NO**

A8.1	Military equipment and munitions	☐	☐
A8.2	Defence equipment, including satellites	☐	☐
A8.3	Technical data, including some border and port information	☐	☐
A8.4	Dual use research	☐	☐
Materials and Technologies		**YES**	**NO**
A8.5	Nuclear materials, facilities and equipment	☐	☐
A8.6	Chemicals, microorganisms and toxins	☐	☐
A8.7	Electronics design, development and production	☐	☐
A8.8	Computers	☐	☐
A8.9	Telecommunications	☐	☐
A8.10	Sensors and lasers	☐	☐
A8.11	Navigation and avionics	☐	☐
A8.12	Marine technology	☐	☐
A8.13	Composite materials	☐	☐
A8.14	Aerospace and propulsion	☐	☐
Sensitive Information		**YES**	**NO**
A8.15	Cryptology	☐	☐
A8.16	Counterterrorism	☐	☐
A8.17	Counternarcotics	☐	☐
A8.18	Cyber-related	☐	☐
A8.19	Non-proliferation	☐	☐
A8.20	Transnational criminal organisations	☐	☐

Genetic Resources		**YES**	**NO**
A8.21	Genetic material of actual or potential value	☐	☐
A8.22	Organisms, or parts thereof	☐	☐
A8.23	Populations (or organisms)	☐	☐
A8.24	Other biotic components of ecosystems	☐	☐
A8.25	Domesticated or cultivated species in which the evolutionary process has been influenced by humans	☐	☐

Will you be exporting military or dual-use technology, now or possibly in the future? This information may take many forms, including:	**YES**	**NO**

A8.26	Physical transfer of goods, technologies, material or equipment	☐	☐
A8.27	Blueprints	☐	☐
A8.28	Plans	☐	☐
A8.29	Diagrams	☐	☐
A8.30	Models	☐	☐
A8.31	Formulae	☐	☐
A8.32	Tables	☐	☐
A8.33	Engineering designs and specifications	☐	☐
A8.34	Manuals and instructions	☐	☐
A8.35	Software	☐	☐
A8.36	Know-how	☐	☐

Will your project transfer technology using any of the following physical methods?		**YES**	**NO**
A8.37	USB flash drives	☐	☐
A8.38	Portable hard drives	☐	☐
A8.39	Laptops	☐	☐
A8.40	Tablets	☐	☐
A8.41	Intangible forms by using electronic media, such as email, phone storage and cloud file sharing	☐	☐
A8.42	Physical exports	☐	☐

Will your project involve any of the following?		**YES**	**NO**
A8.43	Oral transfer by phone or videoconferencing or presentations	☐	☐
A8.44	Travelling overseas for any purpose, while unintendingly carrying any of the above.	☐	☐
A8.45	Will you be hosting foreign visitors?	☐	☐

If you answer YES to any of questions A8,1-A8.45 then you should complete a full export control assessment and append that to this application. Before doing this please contact INSPIRE who will advise further.		**YES**	**NO**
A8.46	Full export control assessment required?	☐	☐
A8.47	Full export control assessment completed?	☐	☐
A8.48	Full export control assessment appended?	☐	☐

Prifysgol Cymru
Y Drindod Dewi Sant
University of Wales
Trinity Saint David

PG2 FORM

Return to the main form

Prifysgol Cymru
Y Drindod Dewi Sant
University of Wales
Trinity Saint David

PG2 FORM

Prifysgol Cymru
Y Drindod Dewi Sant
University of Wales
Trinity Saint David

Centre for Psychology and Counselling
MSc Applied Psychology

Participant Advertisement

Volunteers Needed — AI Wellbeing Study
University of Wales Trinity Saint David — MSc Applied Psychology Research

What is the study about?
We are looking for volunteers to take part in a research study examining how people experience interacting with a conversational AI assistant in a wellbeing context.

What will I have to do?
You will attend a single laboratory session at UWTSD lasting approximately 30–40 minutes. During the session you will:

- Complete a short set of questionnaires
- Have a ten-minute text-based conversation with an AI assistant about a recent source of stress or a goal you are working toward
- Complete a further short set of questionnaires about your experience

You will also have the option to speak your thoughts aloud during the conversation as part of an optional element of the study — this is entirely voluntary.

Who can take part?
This study is open to UWTSD undergraduate students aged 18 or over who are fluent in English. The interaction involves discussing a recent source of stress or a personal goal. If you are currently experiencing significant mental health difficulties or are in crisis, this study may not be suitable for you at this time — if you are unsure, please contact the researcher before signing up.

What do I get?
Participation counts toward your Research Participation Scheme (RPS) course credit. Credit will be awarded upon attendance regardless of whether you choose to complete or withdraw from the study.

Where and when?
Sessions take place in a computer laboratory at UWTSD IQ Swansea. Available dates and times will be provided upon sign-up.

How do I sign up?
To register your interest or for more information, please contact:
Rhiannon Sigrist — 2219662@student.uwtsd.ac.uk
Or sign up directly through the Research Participation Scheme: [Link here]

This study has been submitted for ethical approval through the University of Wales Trinity Saint David Institute Ethics Committee. It is conducted under the supervision of Dr Paul Hutchings (paul.hutchings@uwtsd.ac.uk).

Prifysgol Cymru
Y Drindod Dewi Sant
University of Wales
Trinity Saint David

PG2 FORM

Prifysgol Cymru
Y Drindod Dewi Sant
University of Wales
Trinity Saint David

Centre for Psychology and Counselling
MSc Applied Psychology

Study Information Sheet

Title of Study
Empathic Design and Working Alliance in Human–AI Wellbeing Interactions: A Randomised Experimental Study

Principal Researcher: Rhiannon Sigrist (2219662@student.uwtsd.ac.uk)
Supervisor: Dr Paul Hutchings (paul.hutchings@uwtsd.ac.uk)

Introduction
You are invited to take part in a research study being conducted as part of an MSc Applied Psychology dissertation at the University of Wales Trinity Saint David. Please read the following information carefully before deciding whether to participate. If you have any questions, please contact the researcher using the details above or speak to the researcher in the room on the day of your session.

Purpose of the Study
This study examines whether the way a conversational AI agent is designed influences how supported and understood users feel during an interaction, and whether this affects their mood afterwards. To ensure your responses reflect your natural experience, some specific details — including which version of the AI you will be assigned to — will not be disclosed until the end of the session. A full explanation will be provided in a debrief at the close of your session. Findings will contribute to the evidence base for the responsible design of conversational AI tools used in wellbeing and support contexts.

What Will Happen Today
The session will take place in a university laboratory at UWTSD and will last approximately 30–40 minutes. It involves four stages:

- **Stage 1: Consent and Baseline Measures (approximately 10 minutes)**
  You will read this information sheet and confirm your consent to participate. You will then provide basic demographic information and complete two short questionnaires measuring your current mood and your expectations of the upcoming interaction.

- **Stage 2: Ten-Minute Chat Session (10 minutes)**
  You will be randomly assigned to interact with a conversational AI agent on a laptop. You will be invited to discuss a recent source of stress or a goal you are currently working toward. The conversation is free-text and a timer will be visible on screen.
  You will also be offered the option to think aloud during the interaction — this means speaking your thoughts and reactions aloud as they occur while you chat. This is entirely voluntary. If you choose to participate, the session will be audio recorded for research purposes only. You may decline without it affecting your participation in any other part of the study. Please inform the researcher before your conversation begins if you would like to take part.

- **Stage 3: Post-Interaction Measures (approximately 10 minutes)**
  After your conversation you will complete a further set of questionnaires about your experience of the interaction, any changes in your mood, and the usability of the AI system.

- **Stage 4: Debrief (approximately 5 minutes)**
  The researcher will provide a full verbal debrief explaining the study aims, the AI conditions being compared, and which condition you were assigned to. You will also receive a written debrief sheet to take away. Course credit will be awarded through the Research Participation Scheme upon attendance, regardless of whether you choose to complete or withdraw from the study.

Potential Discomfort or Risk
Discussing a recent source of stress or a personal goal forms the basis of the AI interaction in this study. While this is a low-intensity context, if you are currently experiencing significant mental health difficulties or are in crisis, this study may not be suitable for you at this time. If you are unsure, please contact the researcher before booking.

For most participants, any discomfort is expected to be mild and brief. You may end the chat session at any point without penalty by returning to the questionnaire computer and continuing with the rest of the study. The AI is not a therapist and cannot provide clinical advice or support. An automated safety monitoring system is active throughout the interaction and will direct you to appropriate support resources if needed.

Confidentiality and Data Protection
Your data will be treated with full confidentiality. You will be assigned a unique numeric identifier at the point of consent; your name and student number will not appear in any research dataset, report, or publication. A separate encrypted record linking your identifier to your registration details is maintained solely to allow you to exercise your right to withdraw your data within one month of your session. This record will not be used for any other purpose and will be permanently deleted once the withdrawal window closes.

All data is stored securely on the researcher's password-protected, encrypted device and will be retained for five years following completion of the study in line with UWTSD research data policies, after which it will be permanently deleted.

If you choose to participate in the optional think-aloud, your audio recording will be transcribed by the researcher and the audio file deleted following verification. The transcript will be pseudonymised and stored alongside other study data.

The only exception to confidentiality applies if information disclosed during the session indicates an imminent risk of serious harm to yourself or others. In such circumstances the researcher may be required to consult with the supervisor and the University's Safeguarding Officer. This limit to confidentiality is explained in the consent form.

Voluntary Participation and Withdrawal
Participation is entirely voluntary. You may withdraw at any point without providing a reason and without any academic or personal penalty. Course credit will be awarded regardless of whether you complete or withdraw from the study. If you wish to withdraw your data after your session, you may do so within one month by contacting the researcher using your unique participant identifier,

which will be provided on your debrief sheet. After this period your data will have been incorporated into anonymised analyses and individual withdrawal will no longer be possible.

Support & Contact Information
If you experience any distress during or after the study, the following support services are available:

- Samaritans (24/7): 116 123 — free, confidential support for anyone struggling to cope
- Shout Text Line (24/7): Text SHOUT to 85258 — confidential text-based support
- NHS 111 (Mental Health Option): Call 111 and select option 2 — connects to local mental health professionals
- PAPYRUS HOPELINEUK: 0800 068 4141 — support for young people experiencing thoughts of suicide, available 9am–midnight daily
- UWTSD Student Wellbeing: wellbeingsupport@uwtsd.ac.uk — university support during office hours
- Student Assistance Programme (SAP): 0800 028 3766 — free, 24/7 confidential helpline covering emotional, financial, and legal matters

Principal Researcher: Rhiannon Sigrist — 2219662@student.uwtsd.ac.uk **Supervisor:** Dr Paul Hutchings — paul.hutchings@uwtsd.ac.uk

Prifysgol Cymru
Y Drindod Dewi Sant
University of Wales
Trinity Saint David

PG2 FORM

Prifysgol Cymru
Y Drindod Dewi Sant
University of Wales
Trinity Saint David

Centre for Psychology and Counselling
MSc Applied Psychology

Study Consent Form

Title of Study: Empathic Design and Working Alliance in Human–AI Wellbeing Interactions: A Randomised Experimental Study

Principal Researcher: Rhiannon Sigrist (2219662@student.uwtsd.ac.uk)
Supervisor: Dr Paul Hutchings (paul.hutchings@uwtsd.ac.uk)

Thank you for attending today's session. Before we begin, please read the following information carefully. You have already been provided with a full Participant Information Sheet — if you have any questions before proceeding please speak to the researcher now.
This study involves completing a series of questionnaires, having a ten-minute conversation with an AI wellbeing assistant, and completing further questionnaires afterwards. The session will take approximately 30–40 minutes.
Please confirm each of the following statements before proceeding:

☐	I confirm that I have read and understood the Participant Information Sheet and have had the opportunity to ask questions
☐	I understand that my participation is voluntary and that I may withdraw at any time without giving a reason and without any consequence to my studies or course credit
☐	I understand that my data will be stored securely and confidentially and will not be linked to my name
☐	I understand that this study involves partial concealment — full details of the study aims will be provided at the end of the session in a debrief
☐	I understand that the AI agent I interact with includes an automated safety monitoring system and that if I express acute distress the interaction will be redirected to appropriate support
☐	I understand that my mood and wellbeing data are special category data and consent to these being collected for the purposes of this research
☐	I understand that I may withdraw my data up to one month after today's session by contacting the researcher using my unique participant identifier

Optional — Think-Aloud Audio Recording
The following item relates to an optional element of the study. You may decline without affecting your participation in any other part of the study.

☐	I voluntarily consent to being audio recorded during the chat session as part of the optional think-aloud procedure. I understand that I will verbalise my thoughts aloud while interacting with the AI, that the audio file will be deleted following transcription, and that the transcript will be pseudonymised and stored securely as part of the research data

For your records:
Date of session: _____
Your unique participant identifier: _____
Please keep this form as a record of your participation. Your participant identifier will be needed if you wish to withdraw your data within one month of today's session. To request withdrawal please contact: 2219662@student.uwtsd.ac.uk

Prifysgol Cymru
Y Drindod Dewi Sant
University of Wales
Trinity Saint David

PG2 FORM

Prifysgol Cymru
Y Drindod Dewi Sant
University of Wales
Trinity Saint David

Centre for Psychology and Counselling
MSc Applied Psychology

Study Debrief Sheet

Title of study: Empathic Design and Working Alliance in Human–AI Wellbeing Interactions: A Randomised Experimental Study

Principal Researcher: Rhiannon Sigrist (2219662@student.uwtsd.ac.uk)
Supervisor: Dr Paul Hutchings (paul.hutchings@uwtsd.ac.uk)

Thank You
Thank you for taking part in this study. Now that you have completed your session, we are able to share the full aims of the research and explain why some details were not disclosed in advance.

Why Some Details Were Withheld
To ensure the scientific validity of this research, you were not initially told which version of the AI you had been assigned to or that the study was specifically comparing the effects of empathic design features. This was necessary to prevent expectancy effects — that is, to avoid the possibility that knowing which condition you were in might influence how you rated your experience or how you interacted with the system. Knowing in advance that a tool has been specifically designed for empathy can subtly shape how people perceive it, independently of the actual interaction. Withholding this information allowed your responses to reflect your genuine experience rather than your expectations.
No false information was provided at any stage. You were correctly informed that the study was evaluating an AI assistant for wellbeing.

What the Study Was Investigating
This study examined whether a conversational AI designed to incorporate structured empathic communication strategies produces a stronger sense of working alliance — the bond and collaborative connection between a person and their helper — and greater short-term improvement in mood compared with a standard AI without these features. The working alliance is one of the most reliable predictors of positive outcomes in traditional psychological support, and this study explored whether similar relational perceptions can emerge in human–AI interaction.

The Two Conditions
Participants were randomly assigned to one of two conditions:
Empathy-Engineered Agent (RAIYU): This version of the AI was designed to incorporate behavioural-empathic strategies such as active listening, validation, and collaborative goal orientation. It was guided by a structured approach designed to prioritise warmth and psychological safety throughout the interaction.
Baseline Agent: This version used the same underlying AI technology but without any empathic guidance or relational design features, providing a standard conversational experience for comparison.
You were assigned to: _____

Your Data and Right to Withdraw
If you wish to withdraw your data from the study, please contact the principal researcher at the address below within one month of today's date, quoting your unique participant code. Upon

receipt of a valid withdrawal request, your data will be permanently deleted. After one month, data will be incorporated into anonymised group analyses, at which point individual data can no longer be isolated or removed.
Your data is stored pseudonymously — identified only by your participant code — and no names or student numbers appear in any research dataset, report, or publication.
Your unique participant code: _____
Date of your session: _____
Your withdrawal deadline: _____

What Happens Next
The anonymised data will be analysed to compare working alliance scores (WAI-SR), mood change (PANAS-SF), and system usability (SUS) between the two conditions. Where participants chose to take part in the optional think-aloud procedure, audio recordings have been transcribed by the researcher and the audio files deleted. These transcripts will be analysed using reflexive thematic analysis to explore participants' perceptions of the AI's empathic responsiveness and the quality of the interaction.
Findings will be written up as an MSc dissertation and may be shared through academic publication. All outputs will use fully anonymised, aggregated data.

Support Resources
The following support services are available to you at any time:

- **Samaritans (24/7):** 116 123 — free, confidential support for anyone struggling to cope
- **Shout Text Line (24/7):** Text SHOUT to 85258 — confidential text-based support
- **NHS 111 (Mental Health Option):** Call 111 and select option 2 — connects to local mental health professionals
- **PAPYRUS HOPELINEUK:** 0800 068 4141 — support for young people experiencing thoughts of suicide, available 9am–midnight daily
- **UWTSD Student Wellbeing:** wellbeingsupport@uwtsd.ac.uk — university support during office hours
- **Student Assistance Programme (SAP):** 0800 028 3766 — free, 24/7 confidential helpline covering emotional, financial, and legal matters

For any questions or concerns about your participation, please contact:

- **Rhiannon Sigrist (Researcher):** 2219662@student.uwtsd.ac.uk
- **Dr Paul Hutchings (Supervisor):** paul.hutchings@uwtsd.ac.uk

Rhiannon Sigrist
MSc Applied Psychology (Student Researcher) University of Wales Trinity Saint David

Section 3: Progress report

3.1 Introduction

Growing demand for scalable mental health support has accelerated interest in conversational AI as low-intensity interim provision (Torous et al., 2025), particularly given prolonged waits for specialist psychological support in regions including Wales (Mind, 2024). Yet the relational quality of these systems remains a significant gap in the evidence base (Malouin-Lachance et al., 2025). RAIYU is an empathy-engineered conversational agent developed for this project to operationalise structured behavioural-empathic communication strategies within a controlled experimental setting (Elliott et al., 2018). The central research question examines whether deliberate empathic design produces stronger perceived working alliance and greater short-term affective improvement than a standard large language model.

The proposal has been submitted and assessed, ethical approval obtained from the UWTSD Institute Ethics Committee, study materials finalised, and the RAIYU platform is built and ready for deployment. This report reflects critically on the research journey to date, the challenges of translating psychotherapeutic constructs to an AI context, and readiness for data collection.

3.2 Planned vs Actual Project Timeline

The original timeline allocated January to March for the literature review, April for proposal development, May for submission, June for ethics approval and pilot testing, July for data collection, August for analysis, and September for write-up.

Progress has largely tracked this plan, with some compression and overlap. The literature review was substantively complete by the original deadline, though refined iteratively as new work emerged. The proposal was submitted on 29 May, close to schedule. Ethics preparation was completed ahead of schedule on 15 June, with institutional approval on 30 June, verbal supervisor confirmation on 2 July, and a final confirmation meeting on 16 July.

Data collection has since been compressed into a two-week window in August, rather than the originally planned month-long July window, reflecting practical constraints on laboratory access and participant availability. Analysis and write-up, originally separate phases, have been consolidated into September, since the compressed collection window makes this more realistic than a rigid split.

The original estimates were broadly realistic but underestimated the iterative nature of ethics preparation for a novel methodology, reinforcing the value of building slack into applied research timelines. A visual comparison is provided in Appendices 1 and 2.

3.3 Key Milestones Achieved

Several milestones mark the project's progression from design to launch. The core literature base was consolidated and critically evaluated, drawing on five theoretical papers and two methodological papers, each selected with explicit rationale. This sharpened the theoretical framing of the study, particularly around the distinction between perceived and genuine therapeutic alliance in AI contexts (Malouin-Lachance et al., 2025).

The proposal was refined through multiple supervisor-guided iterations, integrating theoretical nuance from Seitz's (2024) authenticity distinction and established human-AI relational frameworks (Bickmore & Picard, 2005; Elliott et al., 2018), reinforcing the value of substantive supervisor input.

Ethical clearance was obtained from the UWTSD Institute Ethics Committee, with full supporting materials submitted including the Participant information sheet, consent form, debrief document, and data management plan. Navigating AI-specific ethical concerns, particularly transparency, participant expectations, and interaction log confidentiality, required more iterative dialogue than a standard study, building my working knowledge of applied research ethics in emerging technology.

The RAIYU platform was developed in collaboration with a developer, incorporating a two-condition architecture, a retrieval-augmented generation knowledge base, and an integrated safety filter (Nelson et al., 2025). Study materials were cross-checked for consistency across documents.

3.4 Challenges Encountered

This project surfaced substantive challenges across conceptual, methodological, ethical, technical, and personal domains, several requiring active reconceptualisation rather than simple troubleshooting.

Applying Bordin's (1979) working alliance framework, developed for human therapeutic relationships, to a human-AI context required explicit justification rather than assumed equivalence. Existing work on AI-delivered relational support (Beatty et al., 2022; Bickmore & Picard, 2005) provided a foundation, but a persistent risk

was anthropomorphising the AI in ways that would blur perceived relational quality with genuine psychological alliance (Malouin-Lachance et al., 2025).

Methodologically, the WAI-SR (Hatcher & Gillaspy, 2006) is validated for human therapeutic relationships, so its pre-post administration in an AI context is non-standard and requires explicit defence. The ten-minute interaction window was justified on the basis that early alliance forms rapidly, with the first session carrying particular weight (Río Olvera et al., 2022), and that meaningful relational engagement with mental health chatbots can emerge within brief interactions (Xu et al., 2025). Balancing standardisation against ecological validity was a recurring tension.

Ethically, a central challenge was ensuring RAIYU does not present as therapy. This was compounded by a genuine tension: the design features intended to make RAIYU feel supportive are the same features that risk fostering inappropriate emotional reliance. The most significant turning point came from Seitz's (2024) work on authenticity in healthcare chatbots, prompting recognition that affective mirroring, the empathy strategy originally planned, risked feeling performative rather than genuinely supportive, requiring recalibration of the entire approach, discussed further in Section 3.5.

Technically, coordinating data flow between Qualtrics and the RAIYU platform via ResponseID synchronisation, while ensuring no participant data was transmitted externally, required careful planning with the developer. Working in a field publishing new findings throughout 2025 and into 2026 required continuous literature tracking, alongside managing genuine uncertainty about what constitutes "good" AI empathy given conflicting evidence.

3.5 Adaptations and Decision-Making

Several substantive adaptations were made over the course of the project, each grounded in research evidence, ethical principle, or supervisor guidance rather than being incidental.

Terminology shifted from “therapeutic alliance” to “working alliance” throughout, a precision correction: the former implies a clinical relationship RAIYU is not intended to represent, whereas “working alliance” (Bordin, 1979) is the broader construct applicable to any collaborative helping relationship.

The most significant adaptation followed directly from the conceptual challenge in Section 3.4. The original empathy-engineering approach relied on affective mirroring, but Seitz's (2024) work on authenticity in healthcare chatbots prompted recognition that mirrored affect from an AI can read as performative rather than genuinely responsive. The approach was recalibrated toward a behavioural-empathic model, prioritising active listening, validation, and collaborative goal orientation over affective mimicry, the clearest example in this project of design revised in direct response to evidence.

Two technical additions followed from literature review: retrieval-augmented generation was integrated to ground RAIYU's responses and reduce hallucination risk, and a behavioural health safety filter (Nelson et al., 2025) was added after recognising a gap between systems designed to be empathic and systems designed to be safe.

Participant materials were refined through multiple iterations to clarify RAIYU's limitations and non-clinical scope, prioritising ethical clarity over ecological realism.

The pre-post WAI-SR administration, though non-standard, was retained as methodologically defensible following supervisor consultation.

3.6 Preliminary Data Engagement and Emerging Patterns

As data collection has not yet commenced, this section outlines analytical preparedness rather than findings.

The planned quantitative analysis begins with descriptive statistics to characterise the sample and check data quality, including scale reliability for the WAI-SR (Hatcher & Gillaspy, 2006) and PANAS-SF (Thompson, 2007). Independent-samples t-tests will compare working alliance scores between conditions, a mixed-design ANOVA will assess affect change across time points, and regression analysis will examine the association between alliance and mood improvement.

The planned qualitative analysis will apply reflexive thematic analysis (Braun & Clarke, 2006) to think-aloud data, focusing on perceived empathy, relational engagement, and interactional breakdowns.

One point of anticipated interest is the possibility that quantitative alliance scores and qualitative narratives may diverge, reflecting the mixed-methods sensitivity the design was built around.

3.7 Reflection on Researcher Development

This project has developed my competence as an applied researcher well beyond the technical skills anticipated at the outset.

Working in digital mental health research has sharpened my ethical sensitivity, particularly regarding the non-maleficence implications of AI that can plausibly feel caring without genuine clinical competence, most visible in the empathy-engineering recalibration discussed in Sections 3.4 and 3.5.

The project has also built my confidence in handling conceptual ambiguity. Applying an established construct such as working alliance (Bordin, 1979) to a context it was never designed for required tolerating definitional uncertainty throughout design, strengthening my appreciation of the limits of psychological theory when extended into novel technological contexts.

Collaborating with a developer to build RAIYU introduced a cross-disciplinary dimension, requiring me to translate psychological constructs into technical specifications informing decisions about safety filtering and response grounding. Finally, the depth of literature engagement this project demanded exceeded what I had originally anticipated, given how much relevant work was published in this area during 2025 and into 2026, leaving me with a stronger habit of ongoing literature tracking to carry into future research.

3.8 Next Steps and Project Completion Plan

With ethics approval secured and the supervisor confirmation meeting completed on 16 July, the project now moves into a compressed data collection window in August.

Recruitment will run via the Research Participation Scheme and university mailing lists across a two-week period in August, continuing until the target sample of 60-80

participants is reached, or until the window closes. Analysis and write-up are scheduled for September, with submission due on 30 September.

The main anticipated risk is a low response rate given the compressed window, mitigated by two recruitment channels and course credit incentives already built into the ethics-approved design. Given the safety filter and researcher presence, participant distress is considered a low, well-managed risk.

The project is methodologically sound, ethically approved, and on a realistic, if compressed, schedule.

References

Beatty, C., Malik, T., Meheli, S., & Sinha, C. (2022). Evaluating the therapeutic alliance with a free-text CBT conversational agent (Wysa): a mixed-methods study. *Frontiers in Digital Health, 4*, 847991. https://doi.org/10.3389/fdgth.2022.847991

Bickmore, T. W., & Picard, R. W. (2005). Establishing and maintaining long-term human-computer relationships. *ACM Transactions on Computer-Human Interaction, 12*(2), 293–327. https://doi.org/10.1145/1067860.1067867

Bordin, E. S. (1979). The generalizability of the psychoanalytic concept of the working alliance. *Psychotherapy: Theory, Research & Practice, 16*(3), 252–260. https://doi.org/10.1037/h0085885

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology, 3*(2), 77–101. https://doi.org/10.1191/1478088706qp063oa

Elliott, R., Bohart, A. C., Watson, J. C., & Murphy, D. (2018). Therapist empathy and client outcome: An updated meta-analysis. *Psychotherapy, 55*(4), 399–429. https://doi.org/10.1037/pst0000175

Hatcher, R. L., & Gillaspy, J. A. (2006). Development and validation of a revised short version of the Working Alliance Inventory. *Psychotherapy Research, 16*(1), 12-25. https://doi.org/10.1080/10503300500352500

Malouin-Lachance, A., Capolupo, J., Laplante, C., & Hudon, A. (2025). Does the digital therapeutic alliance exist? Integrative review. *JMIR Mental Health, 12*, e69294. https://doi.org/10.2196/69294

Mind. (2024). *Thousands of people are still waiting six months or more to access specialist psychological support in Wales.* Mind Cymru. Retrieved from https://www.mind.org.uk/news-campaigns/news/thousands-of-people-are-still-waiting-six-months-or-more-to-access-specialist-psychological-support-in-wales/

Nelson, B. W., Wong, C., Silvestrini, M. T., Shin, S., Robinson, A., Lee, J., . . . Trister, A. (2025). An AI-based behavioral health safety filter and dataset for identifying mental health crises in text-based conversations. *arXiv preprint*. https://doi.org/10.48550/arXiv.2510.12083

Río Olvera, F. J., Rodríguez Mora, Á., Senín Calderón, M. C., & Rodríguez-Testal, J. F. (2022). The first session is the one that counts: An exploratory study of therapeutic alliance. *Frontiers in Psychology, 13*. https://doi.org/10.3389/fpsyg.2022.1016963

Seitz, L. (2024). Artificial empathy in healthcare chatbots: Does it feel authentic? *Computers in Human Behavior: Artificial Humans, 2*(1), 100067. https://doi.org/10.1016/j.chbah.2024.100067

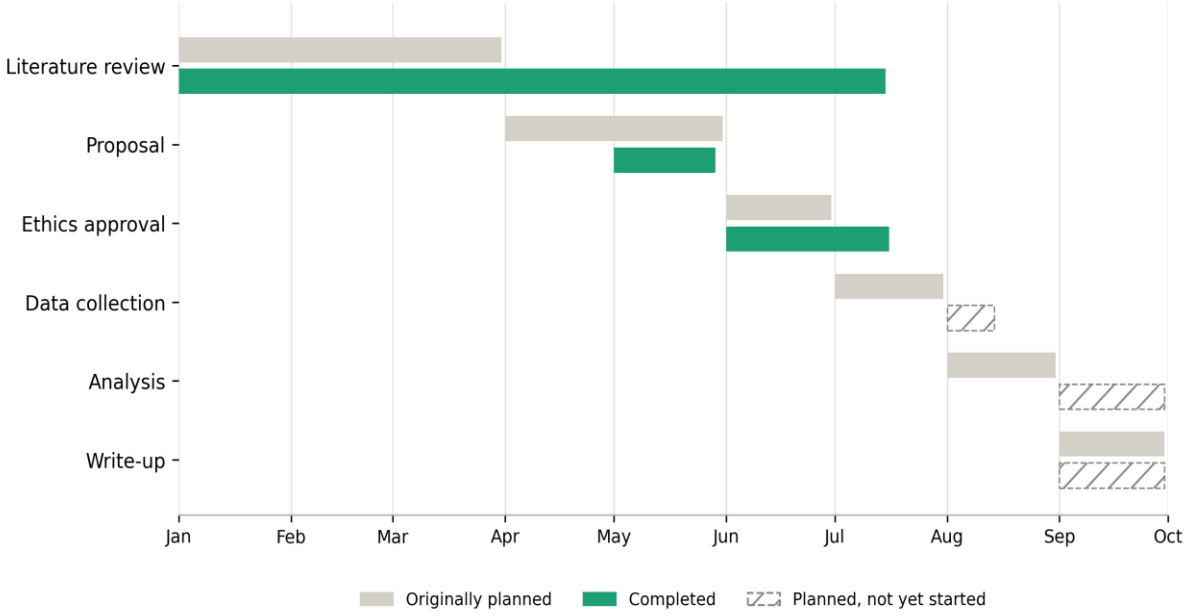
Thompson, E. R. (2007). Development and validation of an internationally reliable short-form of the positive and negative affect schedule (PANAS). *Journal of Cross-Cultural Psychology, 38*(2), 227-242. https://doi.org/10.1177/0022022106297301

Torous, J., Linardon, J., Goldberg, S. B., Sun, S., Bell, I., Nicholas, J., . . . Firth, J. (2025). The evolving field of digital mental health: current evidence and implementation issues for smartphone apps, generative artificial intelligence, and virtual reality. *World Psychiatry, 24*(2), 156-174. https://doi.org/10.1002/wps.21299

Xu, Z., Lee, Y. C., Stasiak, K., Warren, J., & Lottridge, D. (2025). The Digital Therapeutic Alliance With Mental Health Chatbots: Diary Study and Thematic Analysis. *JMIR Mental Health, 12*. https://doi.org/10.2196/76642

Appendix

Appendix 1: Planned versus actual timeline for the RAIYU study



Appendix 2: Planned versus actual timeline for RAIYU study

Phase	Planned	Actual	Status
Literature review	Jan – Mar 2026	Jan – 15 Jul 2026 (ongoing refinement)	Complete
Proposal	April – May 2026	Submitted 29 May 2026	Complete
Ethics approval	June 2026	Completed 15 Jun; approved 30 Jun; verbal confirmation 2 Jul; final confirmation meeting 16 Jul	Complete
Data collection	July 2026	Compressed to a two-week window in August 2026	Not yet started
Analysis	August 2026	Planned for September 2026	Not yet started
Write-up	September 2026	Planned for September 2026	Not yet started

Appendix 3: Core literature identification document

Core literature identification

Section1: Five core theoretical papers

The following five papers constitute the theoretical core of the argument underlying the RAIYU study. They were selected because together they form a logical chain: from why alliance matters in therapeutic contexts, to why empathy drives it, to why it can form with a computer, to what the current evidence for digital alliance looks like, to why standard AI fails and what the specific design challenge is.

Paper 1 - Bordin (1979)

Full reference: Bordin, E. S. (1979). The generalizability of the psychoanalytic concept of the working alliance. Psychotherapy: Theory, Research & Practice, 16(3), 252–260. https://doi.org/10.1037/h0085885

Role in the theoretical framework

Bordin (1979) provides the foundational theoretical framework for the study. This paper proposes the three-component model of the working alliance (agreement on goals, agreement on tasks, and the development of an affective bond between helper and recipient) which has since become one of the most influential and widely used frameworks for understanding therapeutic alliance across therapeutic modalities. Every other theoretical paper in this study either builds upon or contextualises this framework.

The study's primary outcome measure, the Working Alliance Inventory – Short Revised (WAI-SR), directly operationalises Bordin's three-component model. Research Question 1 is therefore asking, in precise terms, whether an empathy-engineered conversational agent can produce measurably stronger Bordin-style alliance perceptions than a baseline large language model. Without this paper, neither the measure nor the question has a theoretical home. It is the conceptual origin of the study.

Centrality to the research question

The study cannot be conducted, reported, or interpreted without this framework. It defines what is being measured, why it is worth measuring, and what a meaningful difference between conditions would represent.

Paper 2 – Elliot, Bohart, Watson, & Murphy (2018)

Full reference: Elliott, R., Bohart, A. C., Watson, J. C., & Murphy, D. (2018). Therapist empathy and client outcome: An updated meta-analysis. Psychotherapy, 55(4), 399–429. https://doi.org/10.1037/pst0000175

Role in the theoretical framework

Elliott et al. (2018) demonstrates that therapist empathy is consistently associated with positive therapeutic outcomes and is widely regarded as a central relational factor in psychotherapy. This meta-analysis synthesises evidence across dozens of psychotherapy studies and over 6,000 clients and consistently finds that therapist-offered empathy is a robust predictor of client outcomes, independent of therapeutic modality. The effect holds across a wide range of populations and settings, making it one of the most replicable findings in psychotherapy research. Empathy is conceptualised here as one relational process that may contribute to stronger alliance formation, rather than as synonymous with alliance itself.

The RAIYU system is specifically designed as an empathy-engineered agent. This design decision (the choice to operationalise empathy through structured communication strategies such as reflective listening, validation, and non-directive support) is theoretically motivated by Elliott et al. (2018). Without this paper, the selection of empathy as the design principle to manipulate is arbitrary. With it, the study is asking a theoretically grounded question: can the empathy-outcome relationship demonstrated in human therapy be reproduced, at least partially, through deliberate AI design?

Centrality to the research question

This paper provides the theoretical justification for the independent variable (empathic design features) and creates the inferential bridge between the psychotherapy literature and the AI wellbeing context. It is central to the rationale for why an 'empathy-engineered' agent should produce different outcomes from a baseline model.

Paper 3 – Bickmore & Pickard (2005)

Full reference: Bickmore, T. W., & Picard, R. W. (2005). Establishing and maintaining long-term human-computer relationships. ACM Transactions on Computer-Human Interaction, 12(2), 293–327. https://doi.org/10.1145/1067860.1067867

Role in the theoretical framework

Bickmore & Picard (2005) is the foundational paper establishing that humans can form relational bonds with computer systems. Drawing on social psychology and human-computer interaction research, building on CASA research, the paper argues that users often apply social expectations and relational norms to interactive technologies, and that these expectations can be deliberately designed for to produce measurable relational outcomes including trust, rapport, and felt connection.

This paper serves as the critical bridge between psychotherapy theory and the AI context. Bordin's (1979) framework was developed in the context of human therapeutic relationships. The theoretical leap required to apply it to human–AI interaction is substantial and requires justification. Bickmore & Picard provide that justification: if users naturally treat computers as social agents, then designing a computer to respond in socially and relationally attuned ways is a theoretically coherent strategy for producing alliance-like perceptions. Without this paper, applying Bordin to AI is an unsupported assumption rather than a theoretically grounded proposition.

This paper also aligns with the Computers Are Social Actors (CASA) framework developed by Reeves & Nass (1996) and Nass & Moon (2000), which together provide the broader human-computer interaction (HCI) theoretical tradition within which Bickmore & Picard sit.

Centrality to the research question

This paper provides the theoretical justification for applying alliance concepts to human–AI interaction. It answers the question: 'Why would we expect alliance to form with an AI at all?' and establishes the design principles that make relational engineering of AI a theoretically meaningful enterprise.

Paper 4 - Malouin-Lachance, Capolupo, Laplante, & Hudson (2025)

Full reference: Malouin-Lachance, A., Capolupo, J., Laplante, C., & Hudon, A. (2025). Does the digital therapeutic alliance exist? Integrative review. JMIR Mental Health, 12, e69294. https://doi.org/10.2196/69294

Role in the theoretical framework

This integrative review is the most current and comprehensive synthesis of evidence on whether a therapeutic alliance can form in digital mental health contexts. The review concludes that digital therapeutic alliance appears possible, although evidence suggests it may differ in strength and quality from alliance formed in traditional human-delivered therapy is the empirical gap that this study is directly designed to address. By establishing where the current evidence stands, Malouin-Lachance et al. (2025) contextualises the study's contribution and makes the experimental question urgent and timely.

Crucially, the review highlights a relative lack of controlled experimental work isolating the contribution of specific empathic design features on alliance perceptions. This is precisely the gap the RAIYU study fills by using a randomised between-subjects design that holds model architecture constant while varying empathic design. The review therefore not only motivates the study but directly validates its methodological approach as the appropriate next step in the evidence base.

Centrality to the research question

This paper defines the empirical gap the study addresses. It is the most direct evidence that the research question is both unanswered and worth asking, and it provides the theoretical baseline against which any findings will be interpreted.

Paper 5 – Seitz (2024)

Full reference: Seitz, L. (2024). Artificial empathy in healthcare chatbots: Does it feel authentic? Computers in Human Behavior: Artificial Humans, 2(1), 100067. https://doi.org/10.1016/j.chbah.2024.100067

Role in theoretical framework

Seitz (2024) addresses the specific problem that RAIYU is designed to solve: whether AI-generated empathic communication is perceived as authentic by users, or whether it is experienced as hollow, scripted, or unsettling — a phenomenon that may resemble aspects of the uncanny valley effect in which technology that is 'almost human' produces discomfort rather than connection. This is the theoretical and empirical problem at the heart of the design challenge for empathy-engineered AI.

While the preceding four papers establish that alliance matters, that empathy drives it, that AI can support relational bonds in principle, and that digital alliance exists but is weak, Seitz (2024) explains a key mechanism for why it is weak: the authenticity problem. Standard large language models produce responses that may be superficially empathic in form but fail to feel genuinely responsive to the user's specific emotional context. The RAIYU system's empathic engineering is precisely an attempt to address this problem, making Seitz the direct theoretical precursor to the study's design rationale.

Centrality to the research question

Seitz (2024) explains why the manipulation being tested matters: if artificial empathy routinely fails to feel authentic, an agent specifically engineered to communicate more authentically may produce meaningfully different relational outcomes. This paper gives the study its specific theoretical edge and connects the broader literature on empathy and alliance to the design decisions embedded in RAIYU.

Appendix 4: OSF pre-registration document

Draft pre-registration content

Study title
Empathic Design and Working Alliance in Human-AI Wellbeing Interactions: A Randomised Experimental Study

Research questions

1. Does interacting with an empathy-engineered conversational agent (RAIYU) produce stronger perceived working alliance than interacting with a standard baseline conversational agent?
2. Does interacting with RAIYU produce greater short-term improvement in mood (reduction in negative affect) than a standard baseline agent?
3. Is stronger working alliance associated with greater mood improvement across both conditions, and does baseline psychological distress moderate this relationship?

Hypotheses

- H1: Participants in the RAIYU condition will report significantly higher working alliance (WAI-SR) scores than participants in the baseline condition.
- H2: Participants in the RAIYU condition will show significantly greater pre-post reduction in negative affect (PANAS-SF) than participants in the baseline condition.
- H3: Working alliance scores will be positively associated with mood improvement across both conditions; this association will be moderated by baseline psychological distress (K10).

Design
Randomised, between-subjects experimental design. Two conditions: empathy-engineered agent (RAIYU) vs standard baseline agent. Random allocation via computer-generated assignment.

Sample
Target 60-80 undergraduate participants (minimum 60 for 0.80 power to detect medium effect size, Cohen's d = 0.50, alpha = .05, independent-samples t-test), recruited via Research Participation Scheme and university mailing lists.

Measures
WAI-SR (working alliance, pre/post), PANAS-SF (affect, pre/post), K10 (baseline psychological distress), SUS (usability, control variable).

Planned analyses

- Descriptive statistics and reliability checks (Cronbach's alpha) for all scales
- Independent-samples t-test: WAI-SR scores, RAIYU vs baseline
- Mixed-design ANOVA: PANAS-SF change scores over time x condition
- Regression: working alliance predicting mood improvement
- Moderation analysis: K10 as moderator of the alliance-mood relationship
- Reflexive thematic analysis (Braun & Clarke, 2006) of think-aloud data, treated as exploratory and not pre-registered

Exclusion/stopping rule
Data collection continues until target sample (60-80) is reached, or end of August 2026, whichever comes first. Incomplete responses (e.g. all items identically rated) will be excluded per criteria specified in the ethics-approved data management plan.

Appendix 5: OSF Registration Confirmation

The analysis plan for this study was pre-registered on the Open Science Framework (OSF) prior to the commencement of data collection, in line with the commitment made in the ethics-approved data management plan (Section P.5). The registration includes the full set of research questions, hypotheses, study design, sample size rationale, planned statistical models, and data inclusion/exclusion criteria.

The registration was submitted under embargo, meaning its content is not publicly visible until the embargo period ends, but the registration itself is permanently timestamped and locked as of the date of submission, consistent with standard pre-registration practice for work intended for future academic submission.

/ 6ah9e / Overview

Reia Sigrist

Empathic Design And Working Alliance In Human-AI Wellbeing Interactions: A Randomised Experimental Study

This Registration Is Currently Archiving, And No Changes Can Be Made At This Time.

Please note

Changes to any files (1) in the projects or components being registered, (2) uploaded or selected as prompt responses, including those connected through add-ons during archiving will result in archiving failure and loss of your registration timestamp. Please do not modify your files until after you have received email confirmation of archiving completion. If this registration has been archiving for more than 72 hours, please email support@osf.io for assistance.

Contributors
Reia Sigrist

Description
This study experimentally examines whether an empathy-engineered conversational AI agent (RAIYU) produces stronger perceived working alliance and greater short-term mood improvement than a standard conversational agent. Using a randomised between-subjects design, participants interact with either RAIYU or a baseline agent and complete validated measures of working alliance (WAI-SR), affect (PANAS-SF), and usability (SUS) before and after a ten-minute text-based interaction. The study also examines whether baseline psychological distress (K10) moderates the relationship between alliance and mood improvement. Existing evidence suggests AI systems can produce alliance-like perceptions, but no controlled study has directly manipulated empathic design features to isolate their effect. Findings are expected to contribute to the evidence base for responsible design of conversational AI tools used in wellbeing and digital mental health contexts, with applied relevance to university wellbeing services and public-facing support tools.

Registration Type
OSF Preregistration

Date registered
Jul 15, 2026

Date created
Jul 15, 2026

Associated project
https://osf.io/k3t7y

Registration title: Empathic Design and Working Alliance in Human-AI Wellbeing Interactions: A Randomised Experimental Study

Registration DOI/URL: https://osf.io/k3t7y

Date registered: 15/07/2026

Embargo end date: 01/10/2026