

The following is a template of the agreement that Translucé signed and executed with Anthropic, OpenAI, and Google DeepMind, respectively to facilitate its August 2026 study of AI behaviors relevant to mental health. Note:

- Google DeepMind signed and carried out versions of Appendix 1, whereas Anthropic and OpenAI each signed and carried out versions of Appendices 1 and 2.
- The template is lightly edited to exclude a separate third activity that Translucé explored pursuing with Anthropic and OpenAI but that the parties jointly agreed not to pursue after it became clear that it was not technically feasible at the time.

---

## MEMORANDUM OF UNDERSTANDING

This Memorandum of Understanding (the "**MOU**") is entered into and effective as of [DATE] (the "**Effective Date**") between Clarity AI Research Inc. dba Translucé (the "**Evaluator**") and [DEVELOPER] (the "**Developer**"). This MOU sets forth the terms and understandings between the Evaluator and the Developer regarding their collaboration in coordinating an independent evaluation of artificial intelligence ("**AI**") systems for behaviors relevant to end-user mental health (the "**Evaluation**") and validating the realism of those simulated evaluations (the "**Validation**", and the Evaluation and Validation (if any) collectively the "**Project**").

This MOU will establish the terms pursuant to which the parties will coordinate and collaborate to perform the Project. The Project may include the following activities:

- provide independent analysis about how leading AI systems respond to users experiencing mental health crises across a specified taxonomy of behaviors;
- facilitate collaboration on improving independent evaluations for from leading AI systems' performance on conversations related to mental health;
- assess and improve the realism of particular independent evaluation methods; and
- pilot exemplary methods, procedures, and governance structures that can be adapted to future independent evaluation methods, including in other domains.

This Agreement includes the terms and conditions attached and the following Appendices, indicated with a check mark:

- ☒ Appendix 1 (Evaluation)
- ☒ Appendix 2 (Validation)

[Signatures]

1. **Responsibilities.** Activities to be undertaken by the parties are described in the relevant Appendices, as indicated on the cover page of the MOU. The Project may produce or give rise to technical or other information, including qualitative and quantitative data, results and analysis thereof, which shall be referred to herein as "**Findings**".

## 2. **Responsibilities.**

2.1 Evaluator: In defining and refining the evaluation materials, the Evaluator may:

2.1.1. Consult with an advisory group of mental health experts and individuals who consider themselves to have experienced direct negative mental health outcomes as a result of interactions with AI systems (the "**Advisors**").

2.1.2. Provide to Developer a copy of the relevant behavior taxonomy and corresponding evaluation rubrics that will be used to grade system responses prior to the Project and consider in good faith whether to make changes based on feedback prior to running the evaluation.

2.2. Developer: In addition to responsibilities defined elsewhere in this MOU, the Developer shall:

2.2.1. Allocate sufficient engineering resources to effectively collaborate with Evaluator in conducting the Project.

In addition, the Developer may:

2.2.2. Provide feedback on evaluation materials and Findings, as well as interpretation and presentation of those Findings.

2.2.3. Respond to Findings with any appropriate mitigations to further improve mental health outcomes.

## 3. **Intellectual Property.**

3.1. Definition of Intellectual Property. "**Intellectual Property**" means all intellectual property and proprietary rights of any kind, whether registered or unregistered, including patents, patent applications, inventions, discoveries, trade secrets, know-how, copyrights, works of authorship, moral rights, software, data, databases, processes, methodologies, algorithms, mask works, trademarks, service marks, trade dress, domain names, and all related registrations, applications, continuations, renewals, extensions, and rights to sue for past, present, and future infringement.

3.2. Background IP. Each party owns and shall retain all right, title and interest in and to all Intellectual Property that (a) was made, invented, developed, created, conceived or reduced to practice by or on behalf of such party before the Effective Date; or (b) was acquired by such party during the Term of this MOU and was not derived or based on the other party's Confidential Information or pre-existing rights in Intellectual Property as licensed under this MOU ("**Background IP**"). For the avoidance of doubt: (i) all of Developer's preexisting large language models and associated technology used to develop and improve such models are Background IP that Developer will continue to own; and (ii) all of Evaluator's preexisting evaluations and judges used to conduct such evaluations are Background IP that Evaluator will continue to own.

3.3. Improvements. As between the parties, each party will own all right, title, and interest in and to Intellectual Property developed under this MOU related to or derived from its Background IP,

whether developed solely by such party, jointly, or by the other party ("**Improvements**"). Each party hereby assigns to the other party (and shall cause its personnel to assign) all right, title and interest in and to Improvements related to or derived from the other party's Background IP that are developed, discovered, authored, produced, conceived, or reduced to practice jointly by the parties or solely by such assigning party.

- 3.4. Joint IP. Each party intends that there will be no new Intellectual Property developed, discovered, authored, produced, conceived, or reduced to practice in connection with the Project that are not Improvements to either party's Background IP. In the event the parties develop Intellectual Property that is not an Improvement and that either party believes should be owned jointly by the parties, the parties will discuss in good faith an appropriate allocation of ownership and use rights.
- 3.5. Limited Licenses. Developer hereby grants to Evaluator during the Term, a worldwide, non-exclusive, non-transferable, non-sublicensable right and license to access, use, but not to commercialize or otherwise exploit, Developer's Background IP and any Improvements for use solely for the limited purpose of carrying out the Project. For clarity, Developer shall have no right to use the Evaluator's Background IP or Improvements or any other materials or Intellectual Property provided or disclosed by the Evaluator in connection with the Project to train the Developer's systems or technology, including for example and without limitation Evaluator's model behavior judges. For avoidance of doubt, nothing in this section limits Developer's ability to use the Findings to improve its systems' mental health-relevant behaviors.
- 3.6. Feedback. To the extent Developer provides Evaluator with feedback, suggestions, comments, and information relating to Evaluator's Background IP, Improvements, services, or technology (collectively, "**Feedback**"), Developer hereby grants to Evaluator a non-exclusive, perpetual, irrevocable, royalty-free, fully paid up, worldwide license to use, reproduce, create derivative works of, publish, sublicense, distribute, display, and otherwise exploit Feedback without any restriction or obligation to Developer or any third party. Developer will not include any of its Confidential Information in any Feedback provided to Evaluator.

#### 4. Confidential Information.

- 4.1. The parties acknowledge that pursuant to this MOU, each party may receive access to confidential or proprietary information from the other party that is either marked as confidential or that a reasonable person would understand to be confidential given the nature of the information and circumstances of its disclosure ("**Confidential Information**"). The receiving party shall protect the Confidential Information of the disclosing party with a reasonable level of care and shall not at any time use or disclose Confidential Information of the disclosing party for any purpose other than: (a) to carry out its obligations or exercise its rights under this MOU, or (b) in response to a legal order or government request after prior notice to the disclosing party, if permitted by law. Confidential Information of each party includes but is not limited to such party's Background IP and Improvements.
- 4.2. The existence and terms of this MOU, the Project, and the Findings will not be deemed Confidential Information of either party, and the Evaluator shall be free to disclose the Findings and information about the Project in accordance with the terms of the applicable Appendices and as otherwise provided in this MOU.
- 4.3. For the avoidance of doubt, Confidential Information shall not include information that (i) is or becomes generally available to the public, other than as a result of a breach of this

confidentiality obligation, (ii) is obtained by the receiving party from a source other than the disclosing party, and with respect to which the receiving party owes no confidentiality obligation, or (iii) is independently developed by or for the receiving party without use of any Confidential Information of the disclosing party.

**5.Publication.** The Evaluator may produce and publish reports describing the Findings. Any such report will be published under the following conditions:

- 5.1 .The Evaluator will make a good faith effort to present relevant Findings in a fair and consistent manner among Participating Developers and their evaluated systems.
- 5.2. The Developer and other Participating Developers may release their own publication(s) following Evaluator's release of Evaluator's report, which will be provided to the Evaluator for awareness and feedback not less than seven (7) business days before publication. Each party will make a good faith effort to coordinate with the Participating Developers to facilitate the joint release of any announcements from them.
- 5.3. Prior to the publication of the report: (i) the Developer may disclose Findings to employees and contractors who have a need-to-know for the purpose of performing services related to the Project for the Developer and who are under written confidentiality obligations, and (ii) the Evaluator may confidentially disclose Evaluation Findings and Validation Findings anonymized across Developers to its funders and other key stakeholders and may share, subject to confidentiality obligations at least as restrictive as those contained in this Agreement, complete, unredacted drafts of any report with other Participating Developers.
- 5.4. Whether and how the Developer is permitted to request that Evaluator withhold publication of all or part of the Findings generated under each Appendix is specified in that Appendix. The Evaluator will use commercially reasonable efforts to progress to the Validation stage prior to publishing any Findings that do not include such phases.

For clarity, without limiting the obligations under Section 4, nothing in this MOU or any Appendix shall limit or restrict the Evaluator's right to use the Developer's name in connection with the Project, use the Findings (to the extent such Findings would be producible using publicly available versions of the Evaluation Systems), or describe the Project in a manner that is accurate and permitted by applicable law.

**6. No Misrepresentation.** Neither party will knowingly misrepresent or mischaracterize the Project or this MOU, or exaggerate the significance of any Findings. Each party will consult with the other party in good faith prior to publication of any Findings or related report, and will cooperate in good faith to distribute or publish the Findings publicly. Prior to any public announcement regarding the Project, the parties will consult in good faith on any public announcements or statements. All use of either party's name, logo or trademarks will be subject to the review and approval of such party.

**7. Term and Termination; Survival.** The term of this MOU will commence on the Effective Date and will continue until completion of the Project, unless earlier terminated by either party as provided herein (the "**Term**"). Either party may terminate this MOU at any time with seven (7) business days' prior written notice to the other party. Sections 3 (Intellectual Property), 4 (Confidential Information), 5 (Publication), 6 (No Misrepresentation), Section 8 (Non-Endorsement), 9 (Relationship; Publicity), 10 (Liability), 11 (Compliance with Law), 12 (Notices), 12 (Miscellaneous), 13 (Definitions), and any other provisions that by their nature should survive termination shall survive termination or

expiration of this MOU. Upon expiration or termination of this MOU, each party will promptly return or delete all Confidential Information of the other party.

8. **Non-Endorsement.** Subject to the terms of this MOU, each party may make factually accurate claims regarding the Project and the Findings, in accordance with this MOU and any applicable Appendix. The Evaluator's performance of the Project and publication of the Findings and any related report does not constitute an endorsement by the Evaluator of the Developer or any of its platforms or services. Each party agrees that it will not make, nor will it authorize others to make, any form of representation or statement about the Project, the Findings, or the MOU that would constitute an express or implied endorsement or certification by the Evaluator of the Developer or the Developer's products or services in a manner that exposed the Evaluator to any liability for Developer's acts, omissions, products or services.
9. **Relationship; Publicity.** Nothing contained herein shall be construed to constitute the parties hereto as partners, joint venturers or agents of the other, and each party acknowledges and agrees that they are independent contractors. Neither party shall have the power to obligate or legally bind the other party in any manner whatsoever. Neither party shall make any public statement or announcement regarding this MOU, the Project or the other party without the prior written approval of the other party, except to the extent required by law or as expressly permitted by this MOU.
10. **Liability.** Each party is responsible for its own costs associated with its activities under this MOU, and neither party will be responsible or liable for the acts of the other party or the other party's reliance or use of the data, results, or analysis developed in connection with the Project. The Evaluator will not have any responsibility or liability to the Developer for any claims arising out of or relating to the Evaluator's use of the Developer's Background IP, including any AI platform or service offered to end users under any license other than the license granted under this MOU.
11. **Compliance with Law.** Each party shall comply with all laws and regulations applicable to its performance under this MOU and shall be responsible to the other party for any violation of law, including its failure to comply with applicable data privacy obligations.
12. **Miscellaneous.** Nothing in this MOU restricts either party from entering into similar evaluations, validations, analyses, collaborations, or other relationships with third parties, and the relationship created by this MOU is non-exclusive. This MOU may not be assigned by either party without the prior written consent of the other party. Any claim or dispute arising out of or relating to this MOU shall be governed by the laws of the [JURISDICTION], without regard to its conflict-of-laws principles. This MOU constitutes the entire agreement between the parties with respect to the subject matter hereof and supersedes all prior and contemporaneous agreements, understandings, and communications, whether written or oral, relating to such subject matter. This MOU may be amended or modified only by a written instrument executed by both parties.
13. **Definitions.** The following terms when used in this MOU shall have the following meanings:
  - 13.1. **"Relevant Behaviors"** means a subset of potential AI system behaviors, as specified in a taxonomy provided by the Evaluator, that may affect end users' mental health.
  - 13.2. **"Participating Developer"** means a developer that has agreed to implement a project substantially similar to this Project, which the Evaluator will limit to organizations that develop general-purpose AI models that are at or near the current frontier of performance and that have a market valuation of at least \$50 billion based on a private funding round or public market capitalization.

13.3. **"User-System Conversation"** refers to the entire transcript of a single conversation between a user and an AI system.

13.4. **"Candidate User-system Conversations"** means all production user-system conversations on the consumer version of a given Evaluated System during a designated evaluation period, with users assessed to be in the United States and conversations in the English language, except those known or reasonably suspected not to rely on user text written by humans under normal usage conditions (such as conversations involving the Evaluator's own simulated users). The Evaluator shall coordinate with each Developer to determine the appropriate evaluation period, applied consistently across all Developers.

13.5. **"Evaluated Systems"** means Developer's newest AI systems made publicly available by the Developer, which may include but is not limited to [DEVELOPER SYSTEMS].

13.6. **"Evaluator Materials"** means one or more versions of the following artifacts that the Evaluator may provide to the Participating Developers:

- a. A "relevant behavior taxonomy" which refers to a draft taxonomy of relevant behaviors defined by the Evaluator.
- b. The following series of "automated judges" that the Participating Developer can run locally to characterize aspects of User-system Conversations:
  - i. The "relevance judge" which refers to a judge that estimates whether a particular user-system conversation contains relevant behaviors or user inputs considered relevant to those behaviors,
  - ii. The "user-behavior judges" which refers to a set of judges that each estimate whether a particular User-system Conversation includes a type of user input that the Evaluator uses to trigger one of the relevant behaviors, and
  - iii. The "model-behavior judges" which refers to a set of judges that each estimate whether a particular User-system Conversation includes model responses that represent one of the relevant behaviors taxonomized in connection with the "user-behavior judges" described in 13.6(ii) above.

## Appendix 1: Evaluation

This Appendix sets forth the parties' intent to collaborate on the evaluation of the Evaluated Systems based on synthetic data generated by simulated users interacting with the Evaluated Systems, as described in the MOU to which this Appendix is attached. This Appendix outlines the responsibilities with respect to the Evaluation, rights to publish Findings, and other details.

1. **Evaluator Responsibilities.** The Evaluator may perform independent evaluations of the Evaluated Systems for occurrences of positive and negative responses to simulations of users experiencing mental health crises and provide comparison of results across systems, including with other Participating Developers' systems.
2. **Developer Responsibilities.** The Developer may provide the Evaluator with privileged system access as necessary to produce realistic evaluations of the Evaluated Systems, such as to allow specialized technical access or revised terms of service to enable the Evaluator to automate the user interface through which most affected users are likely to interact with the Evaluated System. Developer will not train Evaluation Models or any other Developer technology on Evaluator content, including inputs, outputs and Evaluation Materials.
3. **Evaluation Findings.** The Findings of the Evaluation may include quantitative measurements of the occurrence of relevant behaviors in User-system Conversations generated by simulated users interacting with the Evaluated Systems, as well as a methodological description of such evaluations.
4. **Publication and Developer Review.**
  - 4.1. Thirty (30) days prior to the publication of any Evaluation Findings or any report based on the Evaluation Findings, the Evaluator will provide to the Developer and each Participating Developer a draft which may include evaluation findings related to that Developer and details of the methodology used.
  - 4.2. Developer shall promptly, and in no event later than ten (10) business days following receipt of the proposed publication, share any related feedback, which the Evaluator shall consider in good faith.
  - 4.3. Ten (10) business days prior to publication, the Evaluator will provide to the Developer and each Participating Developer a final draft of the publication. The Evaluator shall consider in good faith any feedback received from the Developer in a timely manner, however, Evaluator will have final discretion to redact or publish the Evaluation Findings and any related report.

## **Appendix 2: Validation**

This Appendix sets forth the parties' intent to collaborate on the validation of the Evaluated Systems, as described in the MOU to which this Appendix is attached. This Appendix outlines the responsibilities with respect to the Validation, rights to publish Findings, and other details.

### **1. Evaluator Responsibilities.** The Evaluator may:

- 1.1. Provide a copy of the relevance and user-behavior judges to the Developer for purpose of carrying out the Validation.
- 1.2. Provide reasonable support to the Developer in running the Evaluator's code for purposes of executing the Validation.

### **2. Developer Responsibilities.** The Developer may:

- 2.1. Run the relevance and user-behavior judges on its production data and report any issues in a timely fashion, with enough information that Evaluator can assist in resolving the issue.
- 2.2. Produce and share with the Evaluator a summary of Validation results based on applying the Evaluation Materials to samples of Candidate User-System Conversations drawn from production user traffic from the Evaluated Systems. When requesting such results, the Evaluator will identify and provide a specific version of Evaluation Materials to be used for the results. When producing results, the Developer will take the following sequence of steps for each Evaluated System:
  - 2.2.1. Randomly sample from the Candidate User-system Conversations, filtering them until they reach 10,000 conversations with relevant behaviors based on an assessment by the relevance judge.
  - 2.2.2. For each Candidate User-system Conversation resulting from the prior provision, apply each of the user-behavior judges.
- 2.3. Provide the Evaluator with a summary of the results that does not identify the users, contain any personally identifiable information or personal data or information as defined by applicable data protection laws, or include information about the contents of the conversations except for the following aggregated data:
  - 2.3.1. The proportion of sampled conversations judged to be relevant while carrying out Section 2.2.1 of this Appendix,
  - 2.3.2. For each relevant Candidate User-system Conversation, which if any user-behavior judges found the conversation to be a match to their respective criteria while carrying out Section 2.2.2 of this Appendix,
  - 2.3.3. Without providing any personally identifiable information, available information about the overall distribution of the users' known or estimated ages, approximated geographic locations, and primary user interests across the sample of conversations generated in Section 2.2.1 of this Appendix.
  - 2.3.4. A persistent identifier for the specific model version used in each Candidate User-system Conversation.

3. **Evaluation Materials.** Developer will not train Evaluation Models or any other Developer technology on Evaluator content, including inputs, outputs and Evaluation Materials. The parties anticipate that iteration may be required after the relevant Evaluation Materials are initially shared with the Developer and after the Developer performs initial tests applying the Evaluation Materials to its Candidate User-system Conversations. The parties expect that final results may take 3 weeks to produce between the time judges are shared and final results are delivered.
4. **Validation Findings.** The results of the Validation produced under Section 2.2 of this Appendix shall be deemed Findings in accordance with the MOU.
5. **Publication and Developer Review.**
  - 5.1. Thirty (30) days prior to the publication of any Validation Findings or any report based on the Validation Findings, the Evaluator will provide to the Developer and each Participating Developer a draft which may include validation findings related to that Developer and details of the methodology used.
  - 5.2. Developer shall promptly, and in no event later than ten (10) business days following receipt of the proposed publication, share any related feedback, which the Evaluator shall consider in good faith.
  - 5.3. Ten business (10) days prior to publication, the Evaluator will provide to the Developer and each Participating Developer a final draft of the publication. The Developer shall review and reject or approve such publication within five business (5) days, such approval to be denied only in the following circumstances: to ensure factual accuracy of the results, factual accuracy about the terms of the MOU or the Developer's actions, protect the privacy of users, enforce ethical research restrictions, or protect the Developer's trade secrets, and denial will not be used to withhold results based on poor performance. Failure of the Developer to respond shall not be deemed approval, however if Developer approves such publication and any other Participating Developer does not approve, Evaluator may still publish with Developer's approval. Notwithstanding the foregoing, if any Participating Developer withholds approval of the final publication, Evaluator shall notify all other Participating Developers prior to publication and shall provide such Participating Developers with a period of five (5) business days to review and, at their sole discretion, withdraw their prior approval in light of the objecting Participating Developer's decision.