AI Transcript Coding Service: Data Handling, Method Documentation, and Research Deliverables

This document provides a brief overview of how the NTR AI Negotiation Coder transcript-coding service handles submitted materials and how researchers may describe and cite the method in academic work.

1. Overview of the service

The AI Negotiation Coder website provides several transcript-coding models for negotiation research. The website describes the coding schemes used by each model, gives examples of the codes, explains how the models were developed and validated, and provides instructions for transcript formatting and submission. The output returned to researchers is a spreadsheet-style file containing the coded text units. The site also lists model-specific papers to cite when reporting results.

2. Data handling for submitted transcripts

Security for submission and transfer

Researchers submit files through the official NTR AI Negotiation Coder website at the HTTPS lab domain. Uploads submitted through the official HTTPS site are encrypted in transit between the researcher's browser and the lab's server.

Submitted files are received by a lab-managed application hosted on Amazon Web Services (AWS) in the United States.

Processing workflow

After upload, the transcript text is extracted from the submitted file and sent through the Anthropic API for coding. According to Anthropic's current commercial/API documentation, API inputs and outputs are automatically deleted from Anthropic's backend within 30 days, except where a different agreement applies or where longer retention is required for policy enforcement or legal reasons. Anthropic also states that API data is not used for training unless the customer has a separate agreement stating otherwise. See this link and this link for Antrhopic's retention and privacy policies.

Storage on the lab server

The lab does **not** retain the original uploaded input file after processing. The model's coded output is temporarily stored on the lab's AWS server so the researcher can retrieve it.

That output file is stored for **24 hours** and then automatically deleted. The output is not backed up for long-term retention by the lab.

Retrieval and access

Researchers receive a **Job ID** and a **Passcode** to retrieve their output. These credentials are generated on the server using a random generator.

Access to the server is limited to the lab's authorized personnel. At present, administrative access is limited to the administrators' root account and the engineering team's shared IAM role. Authorized team members may inspect outputs within the 24-hour retrieval window if needed for debugging or user support.

Logging

Operational logs are retained for service administration and usage tracking. These logs record:

- number of lines processed,
- date/time processed,
- the email address, name, and institute (optionally) associated with the submission.

The logs do **not** retain transcript text.

Secondary use

Submitted transcripts are processed for the researcher's requested coding task. The lab does not retain submitted transcripts as a separate stored dataset for secondary research use. The original uploaded input files are not stored after processing, and temporary coded outputs are deleted after the 24-hour retrieval window.

3. Retention and deletion summary

In summary:

- Uploads through the official HTTPS website are encrypted in transit.
- The original uploaded input files are not retained after processing.
- Coded output files are stored temporarily on the lab's AWS server for up to 24 hours;
- Coded outputs are then deleted automatically.
- No long-term backup of those output files is maintained by the lab.
- Anthropic's API-side retention and deletion practices are governed by Anthropic's commercial/API policy.

4. Third-party LLM provider

The service currently uses **Anthropic Claude via API** as part of the coding workflow. Researchers should therefore be aware that transcript text submitted for coding is transmitted from the lab server to Anthropic for processing.

Anthropic's current commercial/API policy (see this link and this link) states that:

- API inputs and outputs are automatically deleted from Anthropic's backend within 30 days, unless an exception applies; and
- Anthropic does not use API data for model training unless there is a separate agreement stating otherwise.

Researchers with institution-specific privacy or ethics requirements should review Anthropic's current commercial data-handling documentation in addition to this summary.

5. Journal-compliant deliverables

The service provides the materials researchers typically need to describe the coding workflow and archive coded outputs for publication or replication purposes.

These materials include:

- the **code definitions** used by the selected model (see the Coding Scheme Used tab under each model on the website);
- a description of the **coding procedure** and unit of analysis used by that model (see the Coding Process tab under each model on the website);
- the **coded output file** returned by the system at the chosen unit of analysis, including the transcript name, speaker, text unit, assigned code, and a consistency score;
- and the relevant **method/validation citation** for the selected model (listed at the end of the introductory paragraph for each model).

Replication considerations

Researchers may describe and archive:

- the coding scheme used,
- the model selected,
- the LLM version used by that model,
- the unit of analysis,
- the coding procedure,
- and the coded output file produced for their project.

Note that, over time, LLMs occasionally shut down older versions as they create newer versions. As this happens, we update the LLM version used in our model. As a result, the exact version used on a given coding task at a certain point in time (e.g., Anthropic Sonnet 3.7) may not be available in the future.

Confidentiality/De-identification:

Any IRB requirements regarding confidentiality and/or de-identification in the transcripts are handled by the researcher before submitting to this service.

Prompt disclosure:
The full prompt used by the NTR AI Negotiation Coder is available on the website:

Model 1 Full Prompt – https://www.ainegotiation-ntr.com/model_aslani/

Model 2 Full Prompt – https://www.ainegotiation-ntr.com/model_brett/

Model 3 Full Prompt – https://www.ainegotiation-ntr.com/model_kern/

6. Recommended method description for academic papers

A reasonable, concise description for an academic paper would be:

> We coded negotiation transcripts using the NTR AI Negotiation Coder transcript-coding service. The selected model applies a predefined coding scheme to each submitted text unit using a large language model with in-context learning, guided by coding instructions and examples derived from human-coded transcripts. We used the model corresponding to [MODEL NAME], and archived the resulting coded output at the [sentence / thought-unit / speaking-turn] level.

This wording is consistent with the lab website's public description of the service and with the overview paper explaining that the lab's approach uses in-context learning with coding instructions and examples derived from prior human-coded transcripts.

This matches the site's public description of model development, validation, and output structure.

7. How to cite

General overview citation

For a general description of the lab's transcript-coding approach, the website points users to:

Friedman, R., Cho, J., Brett, J., Zhan, X., et al. (2024). *An application of large language models to coding negotiation transcripts*.

Model-specific citations

- **Model 1:** Friedman, R., Brett, J., Cho, J., Zhan, X., et al, 2024, Coding negotiations with AI: Instructions and validation for coding model 1.
- **Model 2:** Friedman, R., Brett, J., Cho, J., Zhan, X., et al, 2024, Coding negotiations with AI: Instructions and validation for coding model 2.
- **Model 3:** Friedmann, R., Zhan, X., Tyagi, S., Brett, J., Hooper, M., Babbit, K., Acharya, M., 2026, Coding Negotiations with AI: Instructions and Validation for Coding Model 3.

8. Suggested short ethics-application wording

Researchers may find the following summary useful for ethics or IRB-style applications:

> Transcripts are uploaded through the NTR AI Negotiation Coder's HTTPS web interface and processed on a lab-managed AWS server in the United States. Transcript text is extracted and transmitted via the Anthropic API for coding. The original uploaded input file is not retained after processing. The coded output is stored temporarily on the lab server for up to 24 hours to allow retrieval via a randomly generated Job ID and Passcode, after which it is deleted automatically. Operational logs retain only limited metadata (number of lines processed, processing date/time, and associated email address), not transcript text. Access is limited to authorized lab personnel for administration, debugging, and support. Anthropic's commercial/API data policy applies to API-side processing.