

---

# Stress-Testing Long-Context Language Models with Lifelong ICL and Task Haystack

---

## Supplementary Material - Dataset Documentation

### 1 Code Repository

Our proposed Task Haystack evaluation is based on the consolidation of 64 existing datasets. We include all of our code, including datasets used, model configurations, and evaluation frameworks, in the supplementary material. The uploaded supplementary materials include a `/data` directory for storing the datasets and a `/preprocessing` directory containing scripts to download and pre-process the datasets from HuggingFace Datasets. The model configurations can be found in the `/model` directory. The main evaluation scripts are located in `run_*.py` files. We also provide a detailed `README.md` to assist with reproduction and future evaluations.

### 2 Access and Long Term Preservation of the Dataset

All datasets used in our research are stably and permanently hosted on HuggingFace Datasets. The corresponding HuggingFace Datasets identifiers are listed in Appendix A.2. Additionally, we provide a preprocessed version of the data, along with the preprocessing code in our code repository. We plan to release the code publicly on Github upon acceptance of our work. The authors will answer any questions related to the dataset and code via Github issues.

### 3 Guide to Reading the Dataset

Each task is organized within its respective directory under `/data`, which includes 5 `train_i.jsonl` files containing sampled entries for in-context learning demonstrations, `test.jsonl` for testing, and `config.json` for setting names, options, instructions, and the inference prompt format. During evaluations, the demonstrations and instructions are prepended to the `inference_prompt`, and the input and label fields in the prompt are replaced by those specified in either `train_i.jsonl` or `test.jsonl`.

### 4 License for Dataset and Code

We direct readers to adhere to the licenses provided by the original datasets, as detailed in Appendix A.2. Our code is released under a Apache-2.0 license.