STAI-X 2026 Reproducibility Checklist

Accepted STAI-X papers will undergo a reproducibility check. Authors should provide code, data, and instructions that are as complete as possible so that as many main text figure or table result units as possible can be reproduced. For this checklist, authors only need to consider result units that appear as figures or tables in the main text; appendix and supplementary material results do not need to be included for now.

For each checklist item below:

- For Item 1, replace `\answerTODO{}` with exactly one of `\answerYes{}`, `\answerPartial{}`, `\answerNo{}`, or `\answerNA{}`.
- For Items 2 and 4, replace `\answerTODO{}` with exactly one of `\answerYes{}`, `\answerNo{}`, or `\answerNA{}`.
- Use `\answerPartial{}` for Item 1 when coverage is only partial, and explain the partial coverage in the justification.
- Replace `\justificationTODO{}` with a short 1–3 sentence explanation.
- If code is hosted on GitHub or another repository, provide the URL and the exact branch, tag, release, or commit used for this checklist.

After the numbered checklist, please complete the execution elements table.

Paper title: [FILL: paper title]
Primary supported platform(s): [FILL: e.g., Ubuntu 22.04, Python 3.11]
GitHub / code repository URL (optional): [FILL: repository URL, if applicable]
Repository version / location (optional): [FILL: branch / tag / release / commit SHA; subdir if needed]

1. **Main text figure or table reproducibility**

   Question: Are the main text figures and tables reproducible from your submitted code on at least one supported platform?

   Answer: **[TODO]**

   Justification: **[TODO]**

   Guidelines:

   - Main text outputs should include the paper's central figures and tables from the main text.
   - If only a subset is reproducible, answer [Partial] and specify which outputs are omitted or blocked.
   - If no main text outputs are reproducible from the release materials, answer [No] and identify the blocking issue.

2. **Dependency specification**

   Question: Is a dependency specification included?

   Answer: **[TODO]**

   Justification: **[TODO]**

   Guidelines:

   - Examples include `requirements.txt`, `environment.yml`, `renv.lock`, `Project.toml`, Dockerfiles, or explicit system package lists.

3. **Data availability**

   Question: What is the availability status of the data required for the main text results?

   Answer: **[TODO: Included / Public / Partly available / Restricted / Not available]**

   Justification: **[TODO]**

   Guidelines:

- Use **Included** if the required data are bundled directly with the release materials.
- Use **Public** if the required data are available through public links or repositories with clear access instructions.
- Use **Partly available** if only a subset of the required data are included or publicly available.
- Use **Restricted** if the main results depend on data that require controlled, licensed, or case-by-case access.
- Use **Not available** if the required data cannot be made available to STAI-X organizers or readers.
- The justification should state which parts are available, which parts are restricted or missing, and whether a reduced or demo substitute is provided.

4. **External services, credentials, or hosted infrastructure**

   Question: Are any external services, credentials, hosted models, public APIs, paid APIs, rate limited APIs, external database endpoints, or similar hosted dependencies required?

   Answer: **[TODO]**

   Justification: **[TODO]**

   Guidelines:

   - Answer [Yes] if reproduction requires any external hosted dependency (for example, hosted models, public or paid APIs, rate limited APIs, external database endpoints, private infrastructure, gated services, non-public credentials, or mutable public endpoints); otherwise answer [No].

Execution Elements Table Requirements

- For this checklist, include only main text figure or table result units. Do not include appendix or supplementary material results for now.
- Complete one row per main text figure or table result unit that requires running experiments, processing data, or producing released outputs.
- Each row must describe exactly one main text figure or one main text table result unit. Do not combine multiple result units in one row, even if they are produced by the same script, command, or workflow.
- Do not include purely conceptual workflow diagrams, architecture diagrams, or schematic figures unless they are generated from experiment data or release artifacts.
- Use exact relative paths and commands whenever possible.
- In the result description column, state any expected tolerance, randomness, or approximate-match criteria when exact agreement with the manuscript value is not expected.
- In the reproducibility column, enter exactly one of [Yes], [Partial], or [No] for the corresponding result item.
- If reproducibility is [Yes] or [Partial], fill in the expected output and runtime column. If reproducibility is [No], leave the expected output and runtime column blank; this result item will not be included in the STAI-X reproducibility check.
- If reproducibility is [Partial] or [No], explain the omitted or blocked parts in the notes on computation and stability column.
- In the notes on computation and stability column, state any large compute requirement (for example, LLM fine tuning or large model training) or reproducibility stability issue (for example, live LLM API calls, hosted judges, mutable external APIs, or nondeterministic services).

Paper locator	Result description and tolerance	Entrypoint	Canonical command	Data availability	Repro-ducibility	Expected output and runtime	Notes on computation and stability
[FILL: Figure / Table / section]	[FILL: what this result shows; randomness and tolerance statement]	[FILL: script or notebook path]	[FILL: exact command]	[FILL: Included / Public / Partly available / Restricted / Not available]	[FILL: Yes / Partial / No]	[FILL: output file or directory; expected runtime]	[FILL: hardware, large compute, API or stability caveats]
[FILL: Figure / Table / section]	[FILL: what this result shows; randomness and tolerance statement]	[FILL: script or notebook path]	[FILL: exact command]	[FILL: Included / Public / Partly available / Restricted / Not available]	[FILL: Yes / Partial / No]	[FILL: output file or directory; expected runtime]	[FILL: hardware, large compute, API or stability caveats]
[FILL: Figure / Table / section]	[FILL: what this result shows; randomness and tolerance statement]	[FILL: script or notebook path]	[FILL: exact command]	[FILL: Included / Public / Partly available / Restricted / Not available]	[FILL: Yes / Partial / No]	[FILL: output file or directory; expected runtime]	[FILL: hardware, large compute, API or stability caveats]
[FILL: Figure / Table / section]	[FILL: what this result shows; randomness and tolerance statement]	[FILL: script or notebook path]	[FILL: exact command]	[FILL: Included / Public / Partly available / Restricted / Not available]	[FILL: Yes / Partial / No]	[FILL: output file or directory; expected runtime]	[FILL: hardware, large compute, API or stability caveats]

Example row Authors may delete this example from the completed checklist.

- **Paper locator:** Table 3
- **Result description and tolerance:** Main predictive performance metrics on the full clinical dataset; manuscript values are means over seeds 1–5 and should match within 0.5 percentage points.
- **Entrypoint:** `scripts/run_main_experiment.py`
- **Canonical command:** `python scripts/run_main_experiment.py -config configs/full.yaml -seeds 1 2 3 4 5 -outdir outputs/table3`
- **Data availability:** Restricted
- **Reproducibility:** [Partial]
- **Expected output and runtime:** `outputs/table3/summary_metrics.csv`; about 3.5 h.
- **Notes on computation and stability:** 1 GPU with 24 GB VRAM, 32 GB RAM; exact manuscript values require licensed data access, so public release reproduces only the synthetic-data version.