# Online Resource 6: Use of AI tools

**Article:** When Does Supervision Beat Reconstruction? Magnification Shift and the Robustness of AI Quality Control in Digital Pathology

**Journal:** Journal of Medical Systems

**Authors:** Konstantinos Moutselos, Ilias Maglogiannis

**Corresponding author:** Konstantinos Moutselos, Department of Digital Systems, University of Piraeus, Piraeus, Greece; kmouts@unipi.gr

Reference numbers in square brackets refer to the reference list of the article. Sections, tables and figures without the prefix S are those of the article.

**Tool.** Claude, a large language model developed by Anthropic.

- **Versions and dates (code sessions):** taken from the co-author trailers of the commits on our computing server (867 commits from 19 August to 18 October 2026; 370 name a model, the other 497 record none):
  - Claude Opus 5: 27 August–1 September 2026 (15 commits, 8 of them with the 1M-token context variant);
  - Claude Opus 5.5: 30 September–18 October 2026 (339 commits);
  - Claude Sonnet 5.5: 14–15 October 2026 (15 commits);
  - Claude Fable 5: 8 October 2026 (1 commit).
- **Versions and dates (application sessions):** the Claude models available in the application at the time (August–October 2026); the model was not recorded per session. The current session runs Claude Opus 5.5.
- **Access:** two user interfaces; no API calls, and no sampling parameters (temperature, top-p) set by us. The interfaces' default settings were used throughout.
  - The claude.ai web and desktop application: analysis planning, review of code and results, drafting of text.
  - claude.ai/code sessions running on our computing server through a terminal: writing and running analysis code.
- **How the two were used together.** The first author relayed messages between the two. The application session drafted the instructions for the code session and reviewed its proposed scripts and reports; the code session wrote the scripts. No script ran before the first author approved it.

**Tasks.**

1. **Analysis code.** The model drafted and revised the Python scripts for evaluation, statistics, figures and tables. It also proposed analyses and decision rules; the authors decided which to adopt.

2. **Laboratory record.** The model drafted the entries of an internal, dated laboratory log: analysis plans, results and corrections. The log is not released with the paper.
3. **Manuscript.** Under the authors' direction, the model drafted the manuscript text, figure captions, tables and supplementary material; the authors revised every section, and the model revised the text in response.
4. **Literature.** The model summarised papers that the authors supplied and drafted the reference entries. Every reference was retrieved by the authors from the publisher or an indexing database and checked against the statement it supports.

**Human verification and quality assurance.**

- **Before a run:** every script was shown to the first author as a complete diff and approved before it ran. Each evaluation ran once, from a clean, version-controlled tree. Its commit, start and end times and output checksum were recorded.
- **Reproduction gate:** before any new measurement, each script had to reproduce previously stored results exactly; otherwise the run stopped.
- **Numbers:** every number in the manuscript was traced to a stored output file. Tables were generated by scripts from stored outputs, with a check that every value matched its source; no value was typed by hand.
- **Figures:** figures were regenerated from stored outputs. Each new version was checked to leave the plotted values unchanged; only display properties changed.
- **Images:** no histology image was generated or altered by an AI tool. Overlays were drawn from stored model outputs, without colour or contrast adjustment.
- **Responsibility:** the authors take full responsibility for the content, including all citations.

**Representative prompts.** Quoted text is verbatim; text without quotation marks is a description.

| Task | Prompt | Output | Human check |
|---|---|---|---|
| Analysis code (application session → code session) | Instruction drafted by the application session and relayed by the first author: "Label-only v2/v3 publication copies. Every value cell must be identical to the current _pub CSV (assert cell by cell)." and, for one table, "Propose the label map to me first, as a two-column list (old → new), before writing the script." | The code session proposed a script that rebuilds the existing tables byte for byte before relabelling them and stops on any changed value, and a label map for review. | The application session reviewed the script and the map; the first author approved them before the single run. Value identity was asserted by the script. |
| Code review (application session → code session) | "Show the revised diff, and its sha256, before running." (after a requested change to a figure-rendering script) | A revised script and its checksum. | Run only after the first author's approval; the run record holds the checksum. |
| Manuscript (first author → application session) | The first author asked for the open points that the model had listed on a reading copy of the manuscript to be fixed (missing table bodies, a figure cited but not produced, the word limit, internal labels in two tables, figure fonts and three small checks), and for new versions (description). | Revised reading copies of the manuscript and supplement. Table bodies were inserted by script from the stored table files; the uncited figure was removed and its citations redirected; the AI-use subsection was shortened, with the details moved to this Online Resource; the items | The authors read the copies line by line. Table values were checked against their source files by the build script. |

| Task | Prompt | Output | Human check |
|---|---|---|---|
| | | that needed the computing server were listed. | |
| Laboratory record | The first author asked for a correction entry about the training-monitor acceptance check (description). | A draft entry. | The first author read the text before it was inserted. It was inserted verbatim, compared byte for byte, and committed alone to the append-only log. |
| Literature | The first author supplied the PDF of a newly published paper and asked for the model's summary of it (description). | A summary and a suggested place for the citation. | The authors decided to cite it and checked the reference in Zotero [pending]. |