

Language services

Before you trust an AI reviewer's score

From the field note "AI translation quality assurance: cut false positives, keep the real errors".

The short version

- ◇ In a book-length test, only 4 of 18 AI flags (about 22%) were correct as coded, so AI translation quality assurance must be built for precision first.
- ◇ A few general review principles and an evidence-checking validator replace a growing pile of span-specific prompt exceptions.
- ◇ Computable checks (do-not-translate terms, glossary renderings, approved translation memory) run in code, the same way on every run.
- ◇ Book-level consistency needs its own pass, because per-segment reviewers never see other segments.
- ◇ A small gold-label set with a recall guard, and a review chain where only fresh disagreement escalates, keep precision measurable and arbitration small.

The checklist

- ☐ Have a senior linguist label 15 to 20 AI flags from one real file, keep the real errors as a recall guard, and add two or three more files before quoting precision.
- ☐ Replace span-specific prompt exceptions with general principles and a validator that keeps only evidence-backed flags.
- ☐ Check DNT and glossary terms in code with boundary-aware matching, and never duplicate a model's finding.
- ☐ Normalize quotes, dashes and spaces before TM matching, and soft-flag instead of deleting.
- ☐ Run a whole-book pass for register and term drift in small chunks, and send every fix to a linguist.
- ☐ Escalate only fresh disagreement, and use one filter for every count a linguist sees.

Want a second pair of eyes on your answers?

Book a free 30-minute discovery call at www.seriustech.com/contact or email contact@seriustech.com. We reply within 1 business day.