Using AI to Type Transcripts: What Works, What Breaks, and How to Check It

Veri Eval Practitioner Brief · November 2026

Temitope Petrus-Nihi, Senior Director, Credential Evaluation, Veri Eval

Every foreign transcript has to be typed into a course-by-course record (course code, title, credits and grade) before an evaluator can convert grades or work out a GPA. It is slow work and easy to get wrong. AI tools now say they can do this typing in seconds, and many offices are starting to use them. We wanted to answer a simple question: if you hand a transcript to an AI tool, how much of what comes back can you trust, and how would you know when it is wrong?

This brief is a practitioner's summary of the white paper *Typist, Not Interpreter: Measuring and Improving AI Accuracy in Transcript Data Extraction for Credential Evaluation* (verieval.com/research/typist-not-interpreter), which sets out the full method, results and limits. It is written for evaluators and office managers, and no technical background in AI is needed to apply what it describes.

What we did

We gave AI tools a set of transcripts and asked them to type out every course line. We then compared each line with the correct answer and counted how many came back exactly right, meaning the code, title, credits and grade all matched.

- **The transcripts.** Five real applicant transcripts from Liberia, Kenya, Pakistan, Bangladesh and Afghanistan, with personal details removed, and ten made-up transcripts that we built ourselves for invented schools. The made-up transcripts cover common layouts and six languages, including Spanish, French, Chinese, Arabic and German, with 561 course lines in all. Because they are fictional, we can share them freely, and we know the correct answer for every line.
- **The tools.** Microsoft AI Builder, the AI feature in Microsoft's Power Platform that our own office uses, tested with three different AI models; and the free versions of three chat apps that many people already use: ChatGPT, Claude and Gemini.
- **The instructions.** Every AI tool works from a written instruction, called a prompt. Ours, the "typist prompt," tells the tool to copy only what is printed, exactly as printed, never to convert or calculate, and to flag anything it cannot read. For comparison, we also tried no rules at all, and extra rules on top.

- **The time.** We timed a member of our staff typing a transcript by hand, and then checking records that an AI tool had already filled in.

What we found

The best results were very good, but not perfect. With the typist prompt, the three chat apps copied between 98.2% and 99.6% of the course lines on our made-up transcripts exactly right, and Microsoft AI Builder, with its best model, copied 96.1%. On the real transcripts, which were scanned and had more complicated layouts, that tool got about three lines in four exactly right.

Bigger AI models were not better. Inside Microsoft AI Builder, the two larger models got only about 61% of lines right, well behind the smaller one.

No tool made up a course, but the mistakes look right. A B+ became a B, a grade was taken from the line next to it, or a whole term was left out. Nothing about these errors stands out on the screen.

The tools almost never warn you. On the real transcripts, 442 lines came back wrong or missing, yet only 2 of them were marked as uncertain.

Too few rules and too many rules both cause problems. When we gave the tools no rules at all, they read the transcript correctly but then changed what they wrote, for example combining a mark and a letter into "C (51)." When we added extra rules, they began making their own judgments; one left a whole credit column blank because it decided that "Hours" were not credits. The short typist prompt, which sits between the two, was the best or within two lines of the best in every tool we tried, and it was the only prompt that never caused a large drop. Gemini barely reacted to the prompt at all.

Checking a filled-in record is much faster than typing one from scratch. By "checking" we mean taking the record the tool has already filled in, holding it beside the transcript, reading every line to confirm the code, title, credits and grade, and correcting anything that is wrong. Typing one 52-line transcript by hand took our staff member 26 minutes 42 seconds. Checking four records of 64 and 84 lines that a tool had filled in, and fixing the errors in them, took her 9 to 13 minutes each, and she found every wrong course line.

For an office, then, the real question is not "Can AI type transcripts?" but "How do we check what it types?"

Eight mistakes the tools made

Comparing the tools' records with the correct answers showed eight types of error. Each comes with a tip on how to spot it when you check a record.

1. The missing plus or minus. An A− comes back as an A, or a B+ as a B. This was the most common error and the hardest to see, because the grade looks normal. *How to spot it:* compare every grade letter by letter with the transcript, or use the GPA check described below.

2. The grade from the next column. Where a transcript prints two terms or years side by side, the tool sometimes takes a grade from the wrong column, so the course and credits are right but the grade belongs to another course. *How to spot it:* on side-by-side layouts, read across each row and make sure the grade sits on the same line as its course.

3. The missing term. A whole term is left out, often one printed in small type or at the bottom of a column. On one real transcript, 8 courses disappeared this way. *How to spot it:* count the terms on the transcript and in the record, and compare the credit totals.

4. Points instead of the grade. Some transcripts print a mark (say 75) and, beside it, points (the credits multiplied by the mark, say 225). The tool may copy the points as if they were the grade. *How to spot it:* a "grade" far above the top of the grading scale, such as 225 on a 100-point scale, is almost always points.

5. The wrong total. Transcripts often print totals for each page, each term and the whole program, and the tool may copy an earlier total instead of the final one. People make this mistake too: in one of our timed checks, the checker replaced a correct final average with a year average printed higher up the page. *How to spot it:* make sure the total in the record is the last one printed on the transcript.

6. Grades that are too good. When a tool misread a letter grade, the wrong grade was higher than the real one about four times in five (193 times out of 236), far more often than chance would explain, so a C+ could become an A. Most of these errors came from Microsoft AI Builder, but they appeared on every letter-grade transcript we tested. Because this kind of error favors the applicant, it is easy to miss, and it quietly raises the GPA. *How to spot it:* if the GPA worked out from the record is higher than the one printed on the transcript, look for grades that have been read too high. When you read the record line by line, give the best grades a second look, because a wrong grade is usually a better one.

7. Two grades swapped. One tool gave two neighboring courses each other's grades. Because both grades were still there, every total and the GPA still added up. *How to spot it:* only by reading each line against the transcript; no arithmetic check can catch this one.

8. Sums added without being asked. Some tools added their own arithmetic to their notes, such as "courses add to 181; printed total is 179." ChatGPT, Gemini and Microsoft AI Builder got most of these sums wrong, even when the lines they had copied added up correctly, although Claude's sums were right. Since you cannot tell in advance whether your tool is a reliable calculator, it is safest to treat it as a typist only. *How to spot it:* treat any number in the record that is not printed on the transcript as suspect, and do your own adding up in a spreadsheet.

Three simple checks

The good news is that every transcript already contains a way to check itself: its own printed totals. If you add up what was copied and compare it with what the transcript prints, most of these errors show up.

Check	What you do	What a mismatch tells you
1. Credit check	Add up the credits in the record, for each term and for the whole transcript, and compare each sum with the credit total printed on the transcript.	A course or a whole term is missing, or some credits were copied wrong.
2. Points check (only where the transcript prints points)	Multiply each course's credits by its mark, add them up, and compare the result with the points total printed on the transcript.	A grade was copied wrong, or the points were copied in place of the grade.
3. GPA check (only where the transcript prints its own grading scale)	Using that scale, work out each term's GPA from the record and compare it with the GPA printed on the transcript.	A grade lost its plus or minus, or was read too high, which the credit check cannot catch.

When a check fails, it tells you which term to look at again. When all three pass, the record is very likely right, but a grade swapped between two courses can still slip through, so every line still needs to be read. Our free spreadsheet runs all three checks for you: you paste in the copied lines and type in the printed totals, and each term shows either OK or CHECK.

The single most important rule is that **AI should only copy, never calculate.** Leave all the adding up to the spreadsheet.

Together, these practices make up what we call the **TYPE method**: a **T**ypist prompt, **Y**our own test first, a **P**rinted-totals check, and an **E**valuator who reads every line and decides. The diagram on the next page shows the whole method in five steps, from the tool's first draft to the evaluator's decision, and who does each step.

The TYPE method: five steps

T Typist prompt Y Your own test first P Printed-totals check E Evaluator decides

Y Before you start: test your tool on the free made-up transcripts, and retest after any change.

T AI 1. Extract

The tool copies every course line and every printed total, exactly as printed, into fixed fields. Typist, not interpreter: no converting, no guessing.

P Arithmetic 2. Check

Add up the extracted lines and compare them with the printed figures:

- total credits
- points totals (credits × mark)
- GPAs, only where the transcript prints its own grade points

P AI or evaluator 3. Re-read once

For each block that does not match: one narrow re-read of that block only. Accept it only if it now matches the printed figures. Done by hand, the evaluator re-reads the block against the transcript.

Step 3 happens at most once per block. No loops.

P Arithmetic 4. Flag

Blocks that still do not match are flagged, with both numbers, so the evaluator sees them first.

E Evaluator 5. Decide

The evaluator checks every line against the transcript, clears or confirms each flag, and does the evaluation. The tool's output is a first draft, never an evaluation.

The one rule that never bends: never change a value to make the numbers agree. Re-read, then flag. A mismatch the tool did not cause (a school's own error, or an altered document) is noted or queried with the issuer, never corrected.

Who does what: AI = copying only · Arithmetic = a spreadsheet, a flow or a script, never the AI tool · Evaluator = every judgment. Keep real applicant data in tools covered by a data-protection agreement.

Seven things to do

1. **Use AI as a typist, not an evaluator.** Let it draft the course record, but never let it judge equivalency, convert grades or make any other decision.
2. **Give it a short typist prompt.** Ours is free to copy.
3. **Never ask AI to add up.** Do the arithmetic in a spreadsheet instead.
4. **Test your AI tool before you rely on it.** Run our free made-up transcripts through it, compare its answers with the answer keys, and test again whenever the tool or the prompt changes. A bigger or better-known AI is not necessarily more accurate.
5. **Check every record against the transcript's own totals.**
6. **Then read every line.** Some errors, such as two swapped grades, pass every total.
7. **Keep applicant records safe.** Use only business AI accounts whose contract protects your data, and where possible let a system send the files rather than copying them between windows by hand.

A word on writing your own rules

If you write your own prompt, keep it short, and write rules as plain instructions rather than warnings. Of four extra rules we added, two helped and two made things worse; a rule warning against copying the points column made the tool do exactly that, on 12 lines. Change one rule at a time, and test each change on the made-up transcripts first.

If you use ChatGPT, Claude or Gemini

The method works with the chat apps many offices already have: start a new chat for each transcript, turn memory off, paste in the prompt, attach the transcript and ask for nothing else. With our prompt, ChatGPT copied 99.5% of lines exactly, Claude 99.6% and Gemini 98.2%, but these apps change often, so test the version you use.

Above all, keep real transcripts out of personal or free accounts, whose providers may keep chats or use them to train their AI. Use a business version whose contract protects your data, and check it against FERPA, GDPR or your own rules.

The evaluator's role

These results change the evaluator's job rather than remove it. The tool can do the typing, but it will make mistakes, and the evaluator's check is what finds them, so the tool does not need to be perfect as long as every record is checked. That check cannot be rushed: even a record with no errors took about 9 to 10 minutes to check, because the time goes into reading every line, and records with errors took a few minutes more. The evaluator stays at the center of the process, as the person who checks the record and makes every decision about it.

Try it yourself

Everything is free to download under CC BY 4.0 at verieval.com/research/test-kit:

- Ten made-up transcripts in six languages, each with an answer key
- The typist prompt
- The spreadsheet checker
- Scripts to score any AI tool's replies

In an afternoon, you can find out whether your own tool is ready.

About the evidence

We first tested Microsoft AI Builder on the five real transcripts, then ran about 125 further tests on the made-up transcripts with all four tools. We wrote down what we expected to find before testing began, so the results could not be adjusted afterward. The full method, results and limits are in the white paper (verieval.com/research/typist-not-interpreter).

The author works for Veri Eval, which uses Microsoft AI Builder in its evaluation system. Veri Eval designed, ran and scored the tests, and so that anyone can check the work, the transcripts, answer keys, prompts and scoring rules are free to download.