The AI Collection Experiment

Experiment Worksheet

Use this worksheet as you test AI transcription on a small sample from one of your collections.

1. Choose something difficult

Collection:

_____

What material are you testing?

_____

_____

Why did you choose it? What makes it difficult?

_____

_____

Number of pages in your test: _____

2. Decide what success would look like

What are you hoping AI could help you accomplish?

☐ Make handwritten material searchable
☐ Improve accessibility
☐ Reduce staff transcription time
☐ Tackle a transcription backlog
☐ Create drafts for staff, students, or volunteers to review
☐ Extract names, dates, amounts, or other structured data
☐ Make material more discoverable to researchers
☐ Other: _____

What matters most for this collection?

☐ Overall accuracy ☐ Names ☐ Places ☐ Dates ☐ Numbers / amounts
☐ Specialized terminology ☐ Layout / structure ☐ Other: _____

Before you run the AI: What would make this experiment a success?

_____

_____

3–4. Run the AI and correct it

As you review the AI drafts, keep track of the human work involved.

Page	Review & correction time	What did AI struggle with?	Anything surprising?
1			
2			
3			
4			
5			
6			
7			
8			
9			
10			

Optional comparison

Time to transcribe one similar page without AI: _____

Time to review and correct a similar AI Draft: _____

Did AI eliminate work—or move the work somewhere else?

_____

_____

Results

5. Measure what happened

Best-performing page: _____

Character Error Rate: _____

What did AI do particularly well?

_____

_____

Worst-performing page: _____

Character Error Rate: _____

What caused problems?

_____

_____

Look beyond the error rate

Did the AI reliably get the information that matters most for this collection?

☐ Yes ☐ Mostly ☐ No ☐ It depends

What kinds of important errors did you notice?

_____

Did you encounter any plausible-looking errors that could easily have been missed?

Any hallucinations?

6. Look under the hood

Model: _____

Total token usage: _____

Anything about the model, prompt, or AI output you'd want to document if this became a real project?

7. Make something useful

You created an accessible PDF using:

☐ Human-reviewed transcription only

☐ Human-reviewed transcription + AI Drafts for unreviewed pages

Would this reading copy be useful to your researchers or users?

☐ Yes ☐ Maybe ☐ No

What would you be comfortable publishing without additional human review?

Where would you require human review?

8. Decide whether this scales

Quality

Is the AI output good enough for your intended use?

☐ Yes ☐ With human review ☐ Not yet

Human work

Who could realistically review the output?

☐ Staff ☐ Students ☐ Volunteers ☐ Contractors

☐ No one—we don't have the capacity ☐ Other: _____

Based on this experiment, would AI meaningfully reduce the work required?

☐ Yes ☐ Maybe ☐ No

What happens next?

What would you ultimately need from this collection?

☐ Searchable transcription ☐ Accessible PDF ☐ Structured data

☐ Online publication ☐ Export to another system ☐ Other: _____

What did you learn?

☐ Stop here — AI isn't good enough for this collection or use case.
☐ Experiment further — the results are promising, but we need to test more material or another approach.
☐ Run a larger pilot — AI looks useful, but we need to understand the workflow and human review required at a larger scale.
☐ Scale it — the results are strong enough that we're ready to think about processing the collection.

The most important thing I learned:

_____

_____

_____

Questions I still have:

_____

_____

Want a second opinion?

You've got real results now. If you'd like, share your experiment with FromThePage and we'll talk through what we'd try next.
https://calendly.com/fromthepage/30-minute-meeting