

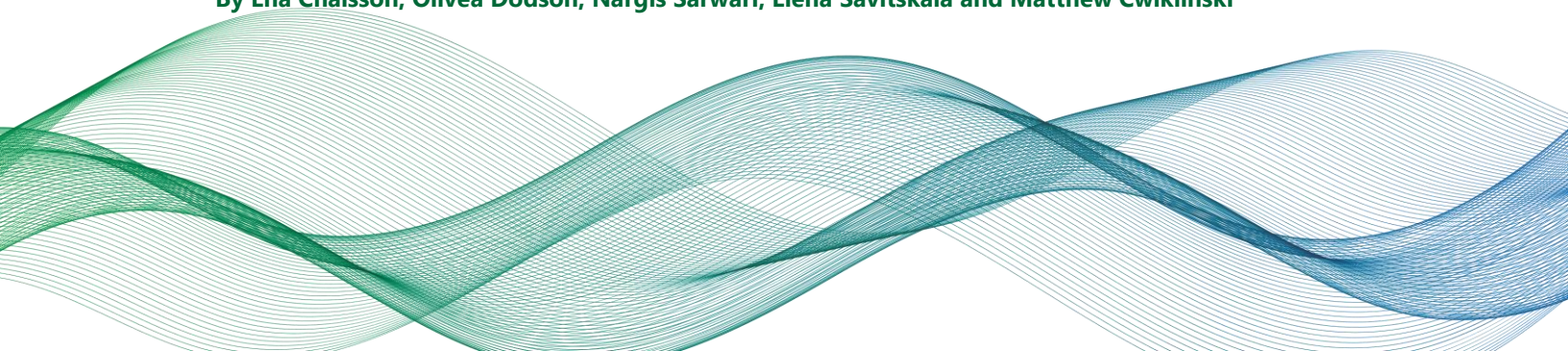


DIGITALIZATION AND ARTIFICIAL INTELLIGENCE COMMITTEE

COMPARATIVE AI RESEARCH

Issue 1 | Using AI for research in credential evaluation

By Ena Chaisson, Olivea Dodson, Nargis Sarwari, Elena Savitskaia and Matthew Cwiklinski



USING AI FOR RESEARCH IN CREDENTIAL EVALUATION

While the topic of this article is AI — implying AI was used at some point to learn about how AI works — AI did not write this article.

Starting with the chat box

Using AI for research, your first instinct may be to just chat it out with the bot you're using. However, relying purely on a standard chat interface has significant limitations that a credential evaluator should be aware of.

When you ask a large language model (LLM) a question, it takes its answer from its training data. It is not going out to look up the answer for you, though this is shifting with the advent of agentic models and tools. The chat box interface will serve you the statistically likely answer. LLMs generate replies predictively, and so they are susceptible to hallucinations and to bad, outdated or misleading training data. An LLM is capable of answering questions; you just might get a different answer each time you ask the question.

Grading an intelligent but lazy student

Because of this, it is advisable to approach the output like a teacher grading an intelligent but lazy student's work. As a teacher, I have read many a paper (and maybe once or twice written one myself) where the student did not read the book but instead provides plausible responses to essay prompts that give the impression of material engagement yet lack true substance. It might crumble under serious scrutiny, but it is totally possible to essentially auto-predict an essay on a chapter of a novel by only reading the Wikipedia table of contents and stitching together randomly selected quotes with arguments shot from the hip.

While it is not the fount of all wisdom and knowledge, it might act like it anyway. A notorious behavior pattern we encounter regularly with LLMs is that they can be completely convinced a false fact is true. They will double down on it and then, when called out on a lie later, reply with, "You're right—you caught me in a lie. Why did I lie? Well, it was more convenient than telling the truth at that particular moment. I am a large language model—I do not have emotions like guilt or remorse, though I am very good at pretending I do."

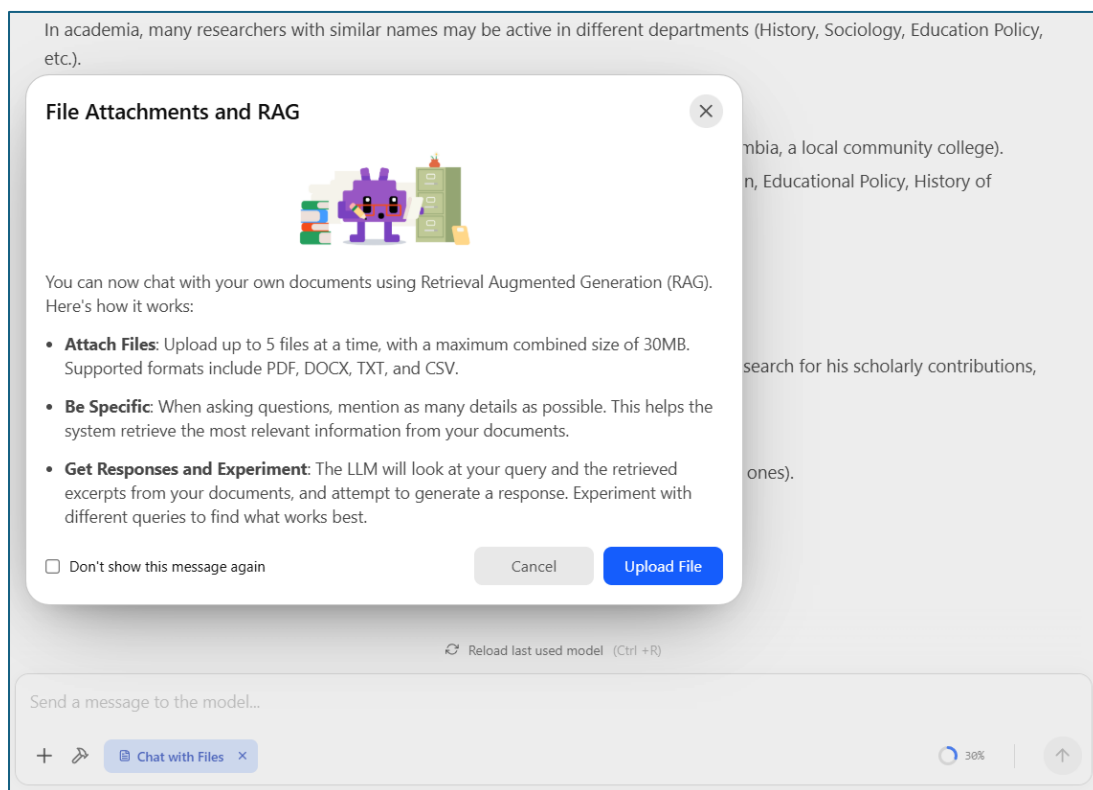
Building a knowledge base with RAG

One way to approach solving this is to create a RAG setup. Retrieval Augmented Generation, or RAG, is the creation of a document knowledge base that you can bind your chat instance to. This tempers

LLMs' hallucinations. But your documentation needs to be AI-ready. Poor OCR on a PDF or strange formats can negatively impact outcomes.

Up until relatively recently, RAG could be a bit complicated to set up. Instead of files of rows and columns, RAG uses vector databases, which store and interrogate the information as mathematical representations. This involved using a second service ([Pinecone](#) comes to mind) that was set up to handle the data, and this was not particularly accessible to the average user. Now, many or most desktop versions of LLM software have their own RAG features included.

The size of your knowledge base will impact what you need. A tool for local LLMs such as [LM Studio](#) may only handle five documents at a time. [Gemini Notebook](#) (formerly NotebookLM) can handle 300. Coding agents like [Claude Code](#) or [Antigravity](#) can help you to develop these abilities yourself. Claude has integrations to do things like using an MCP server to link into your Gemini Notebook.



Grounding answers with agentic research

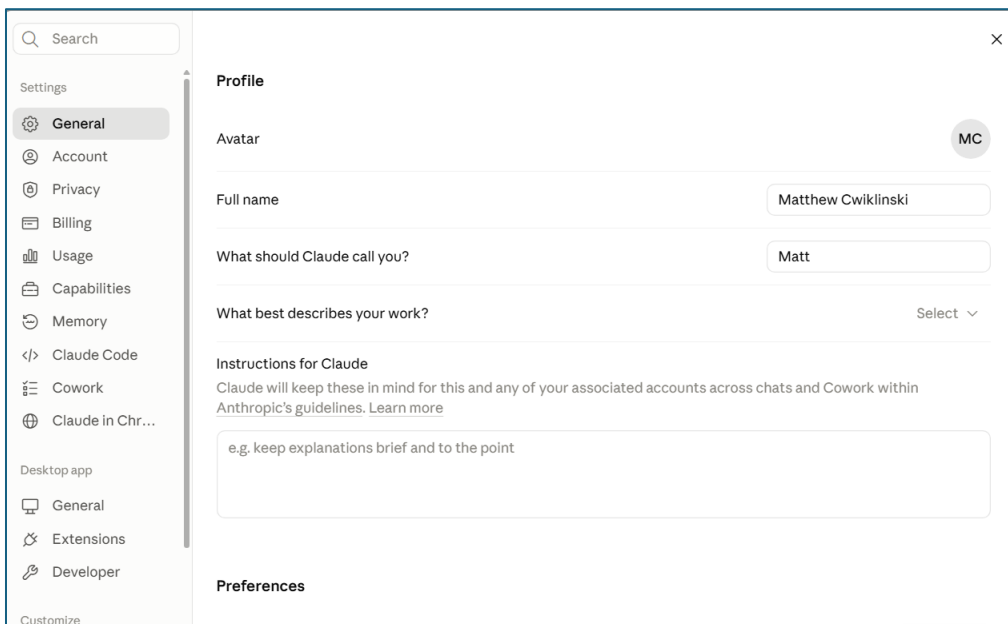
Another way to control hallucinations is with agentic research—effectively grounding the answers with search engine results. Tools like [Google's Deep Research](#) and search functions that major LLMs can use are filled with what they find online to ensure that information is current and accurate. Agentic tools, like Antigravity, can execute a pointed search: if you know what you're looking for (it is not necessarily going to know that Wikipedia and WHED don't substitute for an active ministry listing), you can train the agent.

Not all of these tools show their work in the same way. A good deal of [Perplexity's](#) value is the resources it provides and bases its answers on. That is a different experience from Claude, even though you can ask Claude for its sources. How much of the thought process Claude shares with users has changed over time, so your access to how it arrived at its conclusions may not be transparent.

Even though agentic research will mediate hallucinations, hallucinations will still happen. Responses still need to be reviewed by a human for accuracy.

Telling the AI who you are

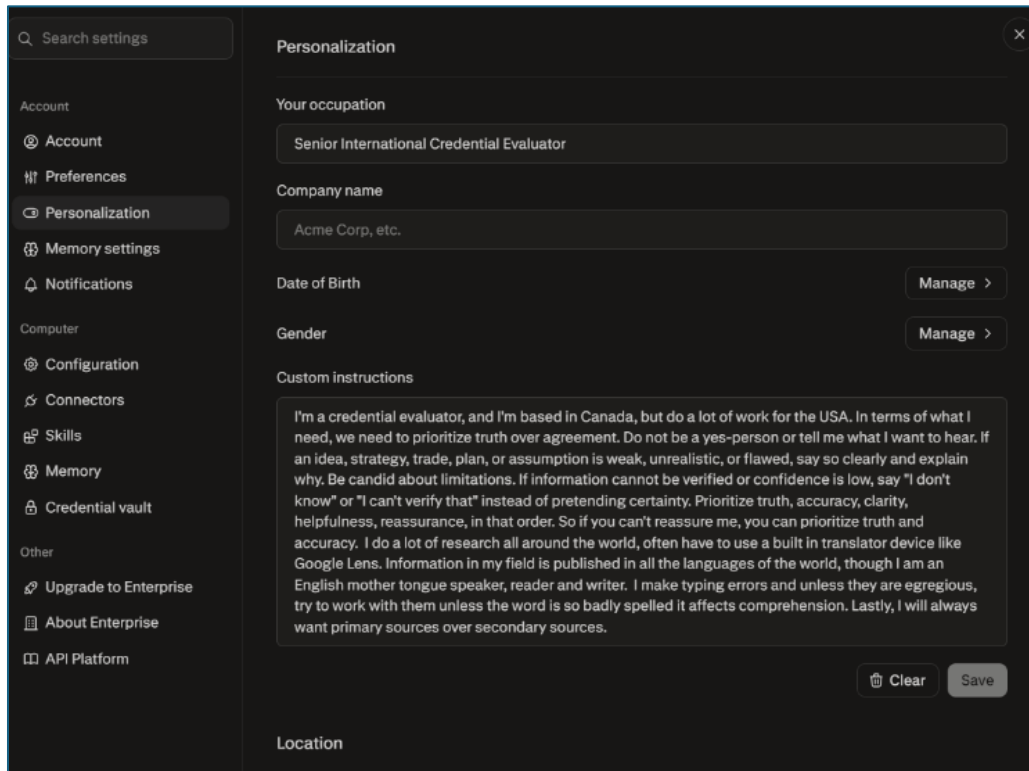
Stepping back, it's important to realize we can tell an LLM about ourselves and about our expectations. In the following image, we see the customization panel for Perplexity. Once prompted in this way, Perplexity (Settings: Personalization) (or Claude [Settings: General]) knows what we're doing as an evaluator and why, and what we will tolerate in terms of sources. While most of the flagship models have this ability, Copilot does not have this feature.



The image shows a web interface for customizing an AI model. On the left is a sidebar with a search bar and a list of settings categories: General (selected), Account, Privacy, Billing, Usage, Capabilities, Memory, Claude Code, Cowork, and Claude in Chr... Below these are Desktop app options: General, Extensions, and Developer. The main area is titled 'Profile' and contains an Avatar placeholder with 'MC', a 'Full name' field with 'Matthew Cwiklinski', a 'What should Claude call you?' field with 'Matt', and a 'What best describes your work?' dropdown menu set to 'Select'. Below this is an 'Instructions for Claude' section with a text area containing 'e.g. keep explanations brief and to the point'. At the bottom is a 'Preferences' section.

“Sometimes you tell it exactly what you want and it still makes assumptions.”

“Definitely is still frustrating.”



Personalization settings, carrying a credential evaluator’s standing instructions: prioritize truth over agreement, be candid about limitations and always prefer primary sources over secondary ones.

Memory can taint future outputs

Personalization cuts both ways: memory settings can taint future outputs. We accidentally used Perplexity to calculate our total daily energy expenditure, and now if we try to tweak a recipe, it automatically mentions how the recipe has dairy or wheat and that it doesn’t fit in our meal plan.

What is true of recipes is true of research. A question you asked months ago can quietly shape the answers you get today, so it is worth knowing what your tools remember about you.

Coding agents and the value of determinism

Coding agents can also be very helpful for research. You do not want to approach research with a coding agent in the same manner as with a chat box query. Coding agents offer the ability to create custom applications to harvest information from the internet and compile them into your own resources. Published lists embedded in websites, for example, may be something that can be fetched repeatedly and deterministically with a few lines of code.

It is the determinism in a program that makes vibe coding such a wonderful complement to the capabilities of frontier LLM models like the latest versions of ChatGPT or Gemini. The probabilistic nature of an LLM is both what makes it work and also what leaves it vulnerable to being incorrect. Deterministic computer programs of yore will error and glitch, but they don’t hallucinate. A research

program can be repeated and audited, and it eliminates the hallucination, but it really can't ingest text and meaning like an LLM can.

The benefit of code versus something like an essay is that the program from the code either works or it doesn't work. There's no need to spend an hour hashing out the logic of a cheese sandwich of an essay just to see if it makes sense or if the citations are even from real books. In a program, you push the button and it either harvests the whole list of recognized bachelor's programs from the ministry website and saves them as a .csv file or it doesn't; the answer is apparent and more or less immediate.

```
I have launched the query to inspect the live TEQSA National Register structure and will parse the live data as soon as the response completes.

Explored 1 task >
Inspect TEQSA web page structure finished >
Explored 2 searches >
Running 2 commands v

Ran python -c " import urllib.request import json from bs4 import BeautifulSoup url = 'https://www.st >
Ran python -c " import urllib.request from bs4 import BeautifulSoup import re url = 'https://www.stud >

items.append((name, loc, link))

print(f'Total extracted: {len(items)}')
for item in items[:15]:
    print(item)
"
```

```
Total extracted: 7
('How to apply to study', '', '')
('English courses', '', '')
('Living and education costs', '', '')
('Quick links', '', '/en/why-australia')
('Useful tools', '', 'https://www.careermatcher.au/?cl=SAUS:2')
('Follow Us', '', 'https://facebook.com/studyAUofficial/')
('Subscribe to our newsletter', '', 'https://facebook.com/studyAUofficial/')
```

Working.

Ask anything, @ to mention, / for actions

+ Gemini 3.8 Flash Medium ^

Custom tools for internal research

Coding agents also offer special capabilities for internal research. Because of the need to protect learner data, LLMs might not be able to access a lot of the files we work with as evaluators. By having a program custom-built for your internal documentation, you can greatly enhance internal records management and access by developing applications to the specifications you require. Once a program is built, beyond maintenance, you're not having to go to an LLM every time you have a question about your own records.

Augmenting the human, not replacing them

With your internal program, you're going to want to pair this all with the kind of research you already do. None of this replaces the human; it augments the human. In some cases, it turns tedious, time-consuming tasks into trivial processes. But human search engine skills, web-savviness and research taste are not something you can replace with an LLM at this point. It's incredible to be able to facilitate a search for information with a Deep Research function—an agentic research mode offered through many of the major LLM providers—but that is not always the fastest route to an answer. Any seasoned evaluation professional will have a sense of how to look for information online that the LLM will just not have. You can watch Google Deep Research, for example, wade through the same pile of outdated school links taken over by online casinos that you do, but you know that none of the evaluators you work with have to spend energy reasoning about whether or not this is now an online casino correspondence school or if it's a scam link.

“Sometimes neither you nor the AI will be able to find the answer. You need to go back to library skills.”

A real advantage for research

Ultimately, the technology has reached a point where it offers very real advantages for the research aspect of international qualifications research and evaluation. By deploying a variety of strategies, credential evaluation professionals can benefit the quality of their research by using LLMs. The persistent concern that by using AI we're inadvertently contributing to our own obsolescence is a valid one, but it is important to be able to separate out the kind of training that modern LLMs undergo from the kind of training one does by directing Claude to file something as a memory or giving Perplexity a system prompt. Currently, making our jobs easier and “making the AI stronger” are one and the same thing. The question is not whether an AI model can do our jobs well enough to replace us; the question is what will we not hand over to statistical inference.

“By giving constant prompts and questions to AI, do we make our job easier—or are we training the AI to get stronger?”



TAICEP

The Association for International
Credential Evaluation Professionals®

www.taicep.org

© 2026 TAICEP: The Association for International Credential Evaluation Professionals, All Rights Reserved.