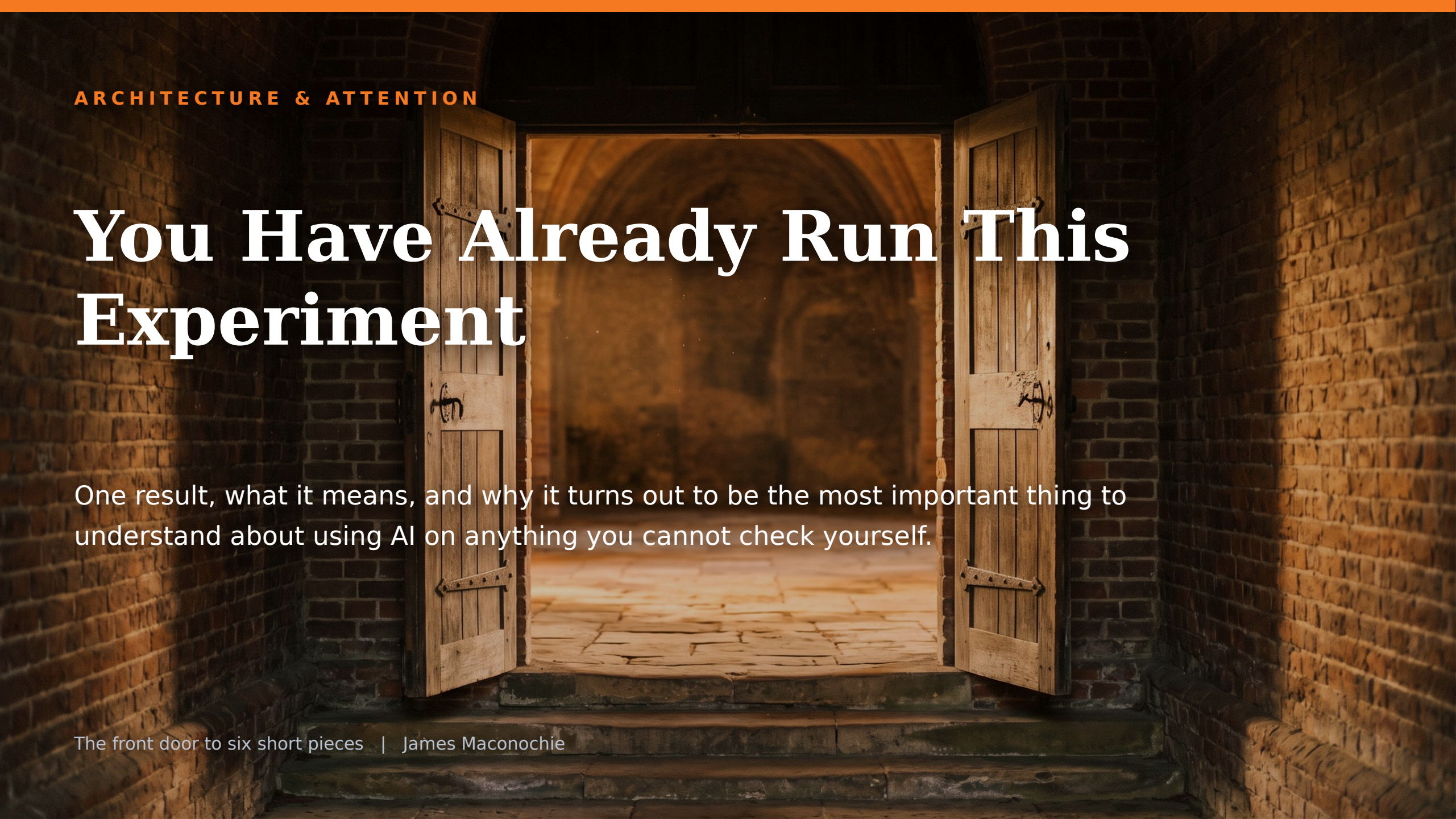


A photograph of an open wooden door in a brick wall, leading to a brightly lit interior space. The door is made of light-colored wood and is flanked by dark brick walls. The interior space is visible through the doorway, showing a stone floor and a large, arched opening in the distance. The lighting is warm and golden, creating a sense of depth and mystery.

ARCHITECTURE & ATTENTION

You Have Already Run This Experiment

One result, what it means, and why it turns out to be the most important thing to understand about using AI on anything you cannot check yourself.

The front door to six short pieces | James Maconochie

THE RESULT

Think back to the last time you caught one out

You asked an AI something. It answered immediately, fluently, and with complete confidence. And on one occasion, at least, it told you something you happened to know was wrong.

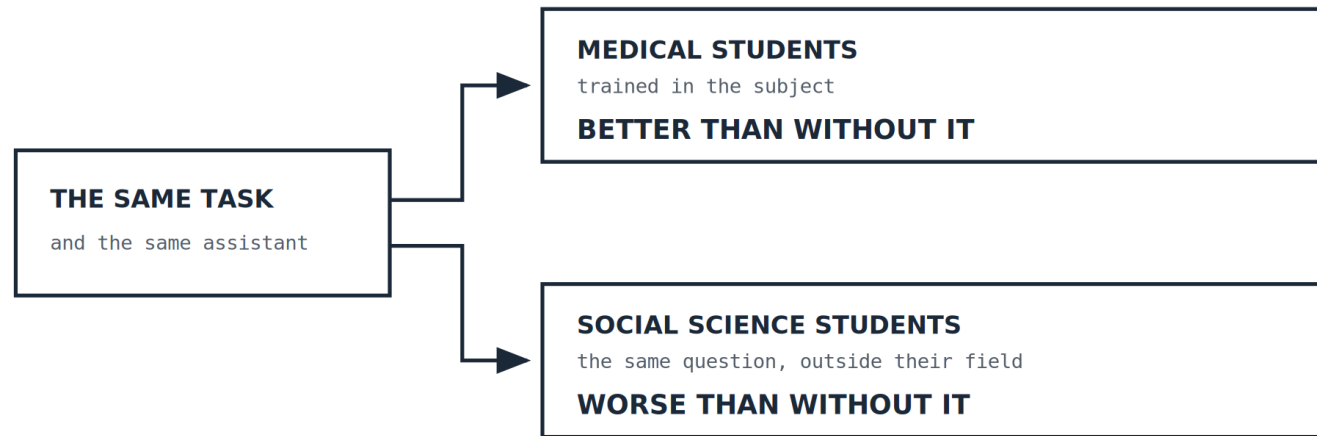
Not obviously wrong. Wrong in a way that would have sailed past almost anybody. Wrong about a detail in your own field, or your own town, or the thing you have been doing for twenty years.

You caught it. Most people file that away as a funny story about the technology. It is a result, and the interesting part is not about the machine. It is about you.

THE RESULT

Two groups, one tool

Researchers gave students an AI assistant to research a scientific question, and compared what they produced against what they produced without it. Same task, same tool. Only one group had been trained in the subject.



the only thing that differed was what each group already knew

Stadler, Bannert and Sailer, 2024, with a 2026 preprint adding domain expertise as the moderating variable. The follow-up is a preprint and has not been peer reviewed. The study reports which way each group moved rather than the size of the effect, so neither does this.

THE RESULT

And it felt easier

The earlier study in the same line of work found something more uncomfortable than a drop in quality. The students using the assistant reported a lower mental workload while producing worse reasoning.



It felt easier at exactly the moment it was going wrong. Which is why nobody catches this from the inside.

This is the part that makes it a structural problem rather than a training problem. The signal you would use to notice is the signal that has been removed.

So what was the one group carrying?

Not more intelligence. Not more effort. They already knew the subject well enough to argue with what came back: which claims to be suspicious of, which numbers smelled wrong, what had quietly been left out.

That capacity has a name. It is judgement.

It is built the only way it can be built: by making calls with something riding on them, and then finding out. The wrong ones show you where you personally tend to go wrong. The right ones tell you what your judgement is actually worth. Neither lesson arrives if nothing was at stake, which is why it takes a career and never quite finishes.

The debugging at eleven at night. The hundredth presentation of the same condition. The case that collapsed on the one witness you had not prepared.

WHAT IT MEANS

Nothing a model does costs it anything

A model has read every account of being right and every account of being wrong that anyone has written down. That is not a small thing. It is why the answers are as good as they are.

What it has not done is been through any of it. When it is wrong, nothing happens to it. No consequence arrives, so nothing in it updates.

That is not a complaint about the technology. It is a description of what it is, and it has one straightforward implication.

The judging has to stay with people.

Which would be fine, except for two things

NOBODY CAN SEE WHETHER IT HAPPENED

Two people review the same document and reach the same conclusion. One thought about it. The other agreed with it. Every record of those two decisions is identical, and nothing any organisation currently inspects can tell them apart. Not because anyone hid anything. Because there is no field for it.

A GREAT DEAL OF MONEY ASSUMES THE PROBLEM SOLVES ITSELF

Hundreds of billions of dollars are committed to an argument whose first step is that this gets handled by scaling. If it does not, the cost does not stay with the people who made the bet. It arrives in retirement accounts nobody chose the contents of, and in jobs that are quietly not refilled.

One is an inspection problem. The other is a balance sheet. They are the same problem seen from two distances.

So what would you build instead?

Here is the distinction that changes the question, and once you see it you cannot unsee it.

A TOOL THAT REPLACES YOU

It has to do the whole job. Every part of it, at least as well as you did, or what you end up with is worse than what you had.

A TOOL THAT AMPLIFIES YOU

It only has to do one part: the part where you were actually stuck. The rest stays with you, because you are still there doing it.

Replacement is five problems, and all five are unsolved. Amplification is one, and it happens to be the one that was slow anyway.

WHAT FOLLOWS

The same five functions, two designs

The left column has to be built from nothing. The right column is mostly already there.

	TO REPLACE YOU	TO AMPLIFY YOU
gather what is known	BUILD IT	BUILD IT
spot what is missing	BUILD IT	THE PERSON
weigh it against the case	BUILD IT	THE PERSON
decide	BUILD IT	THE PERSON
answer for the decision	BUILD IT	THE PERSON

replacement is five problems. amplification is one, and it was the slow one anyway

WHAT FOLLOWS

One question to take away

Whatever you do next, this is the one worth asking about any tool you are using seriously.

After a year of this, what can I do that I could not do before?

If the honest answer is “produce more of what I used to produce”, the tool widened your throughput. If it is “notice things I used to miss”, it widened you.

Those are different products. Nearly everything on the market is sold on the first. That is a choice somebody made, not a fact about the technology.

Six questions

Everything here is asserted. None of it is made out here. Each claim gets made properly in one of six short pieces, with the evidence, the sources, and the objections that have not been answered.

01 Why Does It All Feel Like Too Much?

Language got cheap to make. Attention did not.

02 Why Can't You Just Be Told?

The study above, and why judgement cannot be transferred in text.

03 What Holds a Claim Up?

Four checks for whether a claim can bear weight.

04 How Would You Know?

The two reviewers, and why no inspection anybody runs can separate them.

05 Who Pays If It's Wrong?

The argument carrying the money, and where the cost lands if it fails.

06 What Would You Build Instead?

The constructive half: five principles and a working prototype.

WHERE THIS COMES FROM

Six short pieces, six working papers

These are plain language versions of six working papers. The papers carry the sources, the objections and the limitations, including the ones the author cannot answer.

They are at jamesmaconochie.com. Nothing in them requires a technical background, and you can start anywhere.

If something here is wrong, that is worth more than if it is right.

Working papers are revised when they are shown to be wrong, and what changed is kept on the record.

James Maconochie | Architecture & Attention

HUMILITAS, HUMANITAS, ET CURIOSITAS · SEMPER

Humility, humanity, and curiosity. Always.

