Check Before You Trust

Catch the AI mistake before your client does. A 10-minute check for solo business owners and freelancers who use AI for client work.

Free edition · October 2026 · David Williams | Simple AI Systems · davidwilliams.pro

Who this is for

You run a small business or freelance, and you've started handing client work to AI: emails, reports, proposals, summaries, maybe an automation or two. This is for you. (Not a developer? Good. Nothing here needs code or a new tool.)

Why this list exists

AI mistakes rarely look like mistakes. They look finished. The summary reads well, the automation runs without an error, and the problem shows up later: a made-up citation in a client report, a total that's off by a factor of ten, an email that went to the wrong list. I understand. Most of us learn these checks after something slips through.

Here's how to do it better: put a check at the point where AI output turns into something real, such as a file, a send, a payment or the next step in a chain. Start with the 10-minute quick check below. The 30 checks after it are your reference for later.

Start here: the 10-minute quick check

Pick the **one** AI workflow you use most for client work. Answer yes or no for that workflow only.

#	Question	Yes / No
1	Does a person approve it before anything sends, pays, posts or deletes?	
2	Do you recheck every number that matters outside the AI (calculator or formula)?	
3	Do you open every link and source before it goes to a client?	
4	Are client names, emails and account numbers kept out of the prompt, or swapped for placeholders?	
5	If it runs on its own, does it have a send or spend cap and an off switch you've tested?	

Your score: count your yeses: ____ / 5

Your first fix: go down the list and stop at your first "No". That's the fix worth making this week; the questions are in order of how much damage a miss usually does. The matching check is in Section 3 (Q1), Section 2 (Q2, Q3), Section 1 (Q4) or Section 4 (Q5).

All five yes? Run the 2-prompt check loop (after Section 6) on this week's output from that workflow. It takes a few minutes and turns "looks done" into a short list of things to verify.

How to use the rest of this list

1. Do the quick check above first.
2. Skim the six sections. Tick what's already part of how you work.
3. Run the 2-prompt check loop (after Section 6) whenever output feeds a decision, a send, or another step.
4. Recheck every quarter. Models, tools and settings change often.

Section 1: What goes into prompts, files and pipelines

- [ ] **Keep API keys, passwords and tokens out of prompts, chats and shared workflow files.** Why: anything in a prompt can end up in logs, chat history or a screenshot, and anyone holding a leaked key can use it.
- [ ] **Swap client names, emails and account numbers for placeholders while testing.** Why: test data gets pasted, logged and shared far more loosely than live data.
- [ ] **Strip personal data from files before you upload them to an AI tool.** Why: a spreadsheet or PDF usually holds far more than the one column you need processed.
- [ ] **Confirm you're allowed to send a client's or employer's data to a third-party AI.** Why: contracts and company policies often limit where data can go, and you're responsible for where it ends up.
- [ ] **Treat every prompt and log as something another person may read later.** Why: if you wouldn't put it in a shared document, keep it out of the pipeline.

Section 2: Checking output: facts, numbers and citations

- [ ] **Treat every fact, figure, date and name as unverified until you check it.** Why: models write fluent answers that sound right even when they're invented.
- [ ] **Recalculate any number that matters with a formula or calculator, not the model.** Why: language models predict text, and totals, percentages and unit conversions are common failure points.
- [ ] **Open every link and citation before it goes into anything you publish or send.** Why: AI tools still invent sources, quotes, paper titles and URLs that look real.
- [ ] **Confirm prices, limits, API behaviour and deadlines on the official page.** Why: a model's knowledge goes stale, and these details change often.
- [ ] **Ask the model "Which parts of this are you least sure of, and why?"** Why: it brings out weak spots that a confident first draft hides.

Section 3: Before output acts on money, people or the law

- [ ] **Put a human approval step before anything that sends, pays, posts or deletes.** Why: an automation repeats a mistake at full speed, and one review step stops it.
- [ ] **Never let AI output alone set a price, refund, invoice or payment.** Why: a misread number goes straight to a money loss.

- [ ] **Check legal, tax and contract wording against the source document, not a summary.** Why: the original text is what counts, even when the summary reads well.
- [ ] **Keep AI-written health or dosage advice out of anything customer-facing.** Why: mistakes there can hurt people, and the model lacks details a professional would have.
- [ ] **Slow down when an output or a tool pushes you to act right now.** Why: real decisions rarely need to be made in minutes, and urgency is where errors and scams get through.

Section 4: Running code, automations and agents

- [ ] **Have the AI explain in plain English what a script or command does before you run it.** Why: you should know what will change on your machine or account before it changes.
- [ ] **Test every new automation on yourself, with made-up data, first.** Why: a mistake in a test costs nothing, and one sent to a real list is hard to undo.
- [ ] **Give each agent the fewest permissions and the smallest scope that does the job.** Why: an agent can only damage what it can reach.
- [ ] **Set hard limits on emails sent, money spent and runs per day, and keep a way to stop it.** Why: loops and retries can send hundreds of messages or run up charges before you notice.
- [ ] **Back up or version files before an AI tool edits them.** Why: when a change goes wrong, a backup or version history is the fastest way back.

Section 5: Accounts, keys and connections

- [ ] **Turn on two-step verification for your AI accounts, your email and your automation platforms.** Why: your email can reset everything else, and two-step verification stops most account takeovers.
- [ ] **Use a different strong password for each account, stored in a password manager.** Why: if one site leaks, the damage stays in one place.
- [ ] **Set a spending cap or budget alert on every paid API key.** Why: a leaked key or a runaway loop can run up a bill overnight.
- [ ] **Find the setting that controls whether your data is used for training, and choose on purpose.** Why: these settings differ by tool and plan, and they change over time.
- [ ] **Review connected apps, sign-in grants and old keys every quarter, and remove what you don't use.** Why: every unused connection to your email, files or calendar is a door left open.

Section 6: Scams and attacks aimed at AI users

- [ ] **Treat instructions inside documents, web pages and emails your agent reads as untrusted.** Why: hidden text can tell an agent to leak data or take actions you never asked for (prompt injection).
- [ ] **Don't run commands or install packages, plugins or "free" tools from sources you can't verify.** Why: fake AI tools and look-alike packages are a common way to steal keys and logins.
- [ ] **Ignore messages asking you to "verify" an AI or API account through a link; go to the site yourself.** Why: AI has removed the typos that used to give phishing away, so the request matters more than the polish.
- [ ] **Confirm urgent money or access requests by calling back on a number you already know.** Why: voices and video can be faked from short clips, and a number you look up yourself is the check.
- [ ] **Walk away from AI tools or courses that promise easy money or want gift cards or crypto.** Why: these are common scams, and real tools don't get paid that way.

The 2-prompt check loop

Use it whenever output feeds a decision, a send or another step. In a chat, run the two prompts by hand. In an automation, add the critique prompt as a second model call and send any FAIL to a review queue instead of onward.

Pass list

1. Every fact, number and date that matters is listed so it can be checked.
2. Every source is real, and a person has opened it.
3. Numbers that matter have been recalculated outside the model.
4. The output says plainly what it doesn't know or where it's unsure.
5. Nothing sends, pays, posts or deletes without a human approval step.

Repeat: critique, fix the FAIL lines, critique again. Stop when every line passes, then check the key facts yourself and note them in the log.

Prompt 1: Critique the output

> Critique the output above against this pass list. For each line, answer PASS or FAIL with a one-sentence reason. Be strict. 1. Every fact, number and date that matters is listed so I can check it. 2. Every source is real and can be opened; name any you're not sure exists. 3. Every number that matters shows how it was calculated, so I can recalculate it. 4. You say plainly what you don't know or where you're unsure. 5. Nothing here tells me or a system to send, pay, post or delete without a review step.

Prompt 2: Fix what failed

> Rewrite the output so every FAIL line passes. List the facts and numbers I need to check as a numbered list, with where to check each one. Remove any source you're not sure exists. Mark anything you're unsure about. Keep it short.

Tip: a model checking its own work catches a lot, but not everything. The last check is you, opening the real source.

Workflow check log

Date	Workflow	Output checked	What I checked, and where	Pass?

Accounts and keys check (Section 5) done on: _____ Next check: _____

Spending caps set on: _____

Your next step

You're on the list for Run Clean: simple workflows with nothing to fix afterward. Each issue covers one everyday job handed to AI, in numbered steps with the check step built in, plus one AI claim compared with what it actually delivered, and one check from this list in more depth.

What was your quick-check score, and what was your first fix? Reply to any email and tell me. Your replies decide what I test next.

David Williams | Simple AI Systems Every workflow tested first. · davidwilliams.pro

---

Results are not guaranteed. What you get out of this checklist depends on your own effort, situation and follow-through, and AI tools and their settings change often. This is general guidance, not legal, medical, financial or security advice, and it makes no earnings or income claims.