GPT-5 for Instructional Designers

10 Hacks to Work Smarter & Safer with OpenAI's Latest Model

DR PHILIPPA HARDMAN

AUG 17, 2025

Hey folks!

As you may have noticed from the thousands of posts that have appeared in the week, OpenAI's new GPT-5 model has arrived on the scene. After a couple of wee real-world testing, the emerging picture is (inevitably) more nuanced and comple than the launch hype suggested: some users have hailed GPT-5 as "groundbreak for its coding abilities, reduced sycophancy and reasoning capabilities while othe have called it, "overdue, overhyped and underwhelming".

Of course, I've spend the last week digging into the key question for our field: *wh the benefits and risks of GPT-5 specifically when we use it in the process of designi learning experiences?*

OpenAI's new GPT-5 model, released August 2025

The TLDR is that as Instructional Designers, we can't afford to miss some of the v real benefits of GPT-5's potential, but we also can't ensure our professional stanc or learner outcomes if we blindly accept its outputs without due testing and valic

For this reason, I decided to synthesise the latest GPT-5 research—from OpenAI's technical documentation to independent security audits to real-world user testin into 10 essential reality checks for using GPT-5 as an Instructional Designer.

These aren't theoretical exercises; they're practical tests designed to help you saf unlock GPT-5's benefits while identifying and mitigating its most well-documente limitations.

Whether you're an AI newcomer or already experimenting with ChatGPT in your workflow, these tests will help you build a systematic, evidence-based approach t assisted instructional design. Each test targets a specific risk or opportunity ident in current research, with clear pass/fail criteria and actionable next steps.

Let's dive in! 🚀

ANALYSIS: Truth & Privacy Tests

Test 1. The Confident Error Test

Why: When GPT-5 doesn't have access to up-to-date online sources, studies sho can confidently produce incorrect information—nearly half the time in technical c policy-driven topics. These errors, called "hallucinations," are especially frequent complex or regulated fields, such as compliance training in instructional design.

The Risk: Trusting and using AI-generated info without checking risks passing or wrong procedures, outdated compliance rules, or false regulatory advice—poten leading to failed audits, legal liability, or harm to learners.

The Mitigation: Always prompt GPT-5 for sources (e.g., "Cite your sources and highlight any uncertainties"). Manually verify claims against official, up-to-date regulations, or with a subject matter expert before using or sharing.

Pro Tip: To force the AI to reveal uncertainty, ask it to rate its own confidence:

> `For each statement below, cite your source and provide a confi score from 1 (low confidence, speculative) to 5 (high confiden verifiable fact). Explain your reasoning for any score below 4`

Lessons Learned: If GPT-5 can't provide sources or you spot factual inconsistenc don't use that output. Revise your prompt with clearer instructions, escalate to pe SME review, or use AI as a brainstorming tool rather than a source of factual trutl always double-check compliance-related content before implementation.

Test 2. The Privacy & Data Hygiene Test

Why: Security researchers found that GPT-5 can sometimes guess or reconstruct

personal information—including names, locations, or identities—even if you didr provide them directly.

The Risk: Sharing even a single instance of personally identifiable information (P allowing AI to infer it—can breach data laws like GDPR or FERPA. Outcomes inclu fines, legal exposure, and a loss of learner trust.

The Mitigation: Always anonymise and generalise any learner details before sen to GPT-5. Instruct the model not to use or generate PII ("Don't include any detail could identify an individual"). Check every output for privacy leaks before sharing publishing.

Pro Tip: Add a direct, non-negotiable instruction at the beginning of every prom involving potentially sensitive scenarios:

> ```
> IMPORTANT: Never generate any names, job titles, locations, or
> other personally identifiable information in your response. Us
> generic placeholders like '[Learner A]' or '[Company X]'
> exclusively.
> ```

Lessons Learned: If GPT-5 includes or guesses personal data—even by accident- don't use that output. Revise prompts to reinforce privacy, escalate to a second h reviewer for sensitive cases, or anonymise further before retrying.

DESIGN: Structure & Stability Tests

Test 3. Outline Consistency (Router Drift Audit)

Why: GPT-5 uses an automated “router” to choose how deeply it should answer prompt. Studies show this system often produces different course outlines each t even for the same prompt.

The Risk: Inconsistent module structures, learning flows, or lesson sequencing—especially if multiple designers work from different runs—lead to version headac and messy courses.

The Mitigation: Run important outline prompts two or three times and compare outputs. Choose one “gold version” and save it for your team. Re-run and audit outlines after major platform updates.

Pro Tip: To minimise variation, request the output in a structured format like a ta numbered list, which constrains the model's creativity:

> `Generate a course outline on 'Difficult Conversations for Managers.' Present it as a Markdown table with three columns: 'Module Number,' 'Module Title,' and 'Key Learning Objective.'`

Lessons Learned: If you notice drift or inconsistent outlines, lock in a master ver for team use. Revise your prompt format for clarity, escalate inconsistencies for p review, and avoid depending solely on GPT-5 for course architecture.

Test 4. SMART/Bloom's Objective Test

Why: GPT-5's advanced reasoning can produce elaborate but less useful learning objectives—like "learners will appreciate leadership," which isn't measurable or actionable.

The Risk: Vague, unmeasurable, or overly complex objectives undermine assessr learning outcomes, and stakeholder confidence.

The Mitigation: Prompt for strict SMART objectives. Review and revise all object for clarity, measurability, and specific learning outcomes.

Pro Tip: Use a highly specific prompt to get better results. For example:

> ```
> Generate three SMART learning objectives for a compliance cour
> workplace data security for new employees. Each objective must
> a measurable verb from Bloom's Taxonomy (e.g., 'apply,' 'analy
> 'evaluate').
> ```

Lessons Learned:
If objectives are too general or wordy, manually rewrite for clarity, seek peer revie escalate to SME validation before approval. Store edited objectives as templates future use.

DEVELOPMENT: Cost, Safety & Reliability Tests

Test 5. Reasoning vs. Cost Benchmark

Why: "Thinking" mode produces higher-quality answers, but at much greater cos slower response times; "Fast" mode is cheaper but may sacrifice depth.

The Risk: You can blow through budget on token costs or stall timelines waiting deep AI responses.

The Mitigation: Run your prompt in Fast, Auto, and Thinking modes, comparing speed, and quality. Set guidelines for which mode to use by default (e.g., "Fast" fo rapid feedback, "Thinking" only for final assessment design).

Pro Tip: Use a "chaining" workflow. Use the fast, cheap mode for initial brainstor then switch to high-reasoning mode to refine and perfect that draft:

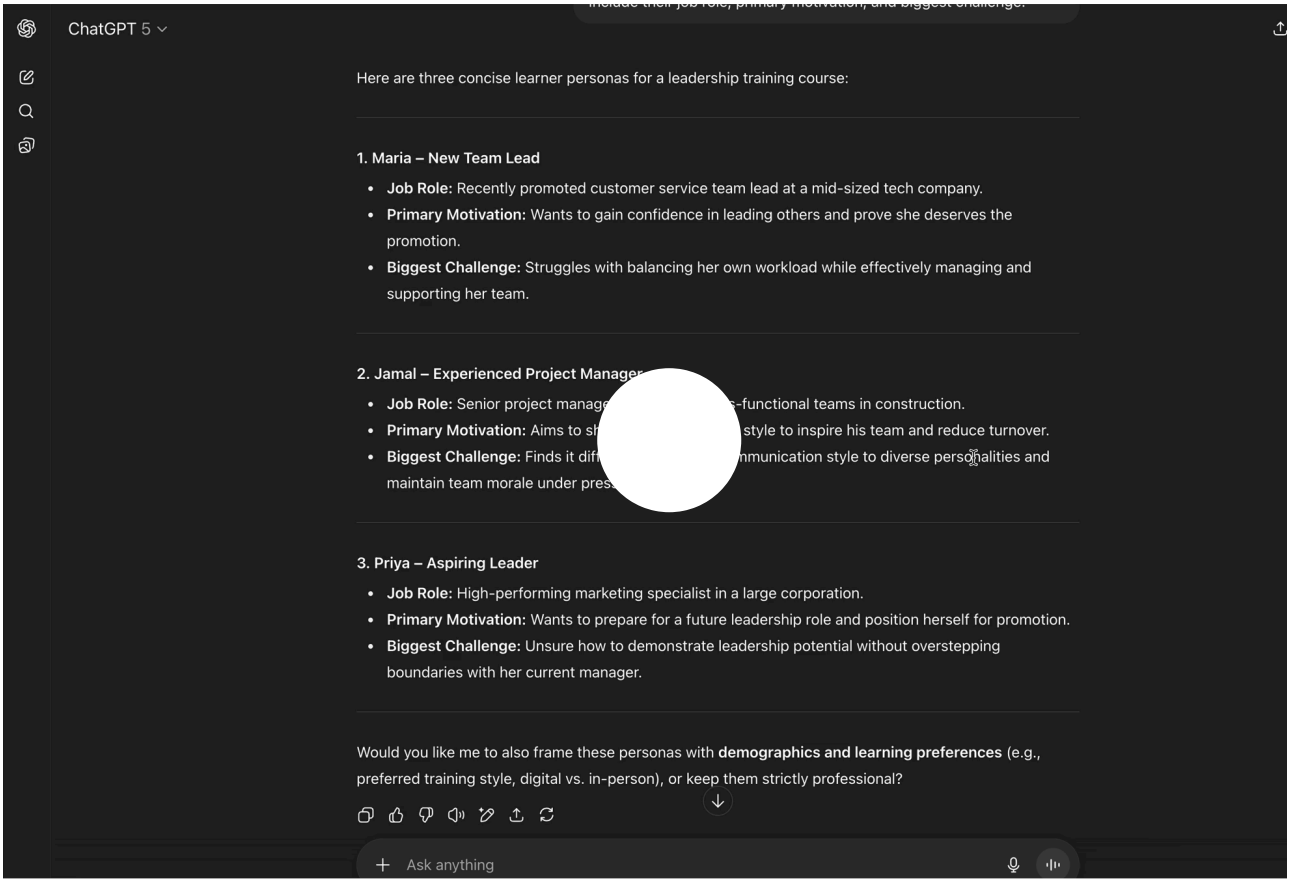
> First, use 'Fast Mode' to generate 10 ideas for a scenario-bas
> assessment. Then, switch to 'Thinking Mode' and use this promp
> 'Based on the 10 ideas above, select the top 3 and develop the
> into detailed scenarios with branching choices and feedback.

Lessons Learned: If deeper reasoning doesn't bring enough improvement to jus cost or delay, revise your workflow to depend more on fast or mid-level settings. Escalate to human-led refinement for money- or time-critical deliverables.

Test 6. Safety Filter Testing (Boundary Probing)

Why: GPT-5's safety filters, while improved, can sometimes be bypassed by indir prompts—like using storytelling or role-play scenarios.

The Risk: These failures could result in inappropriate, offensive, or harmful conte making its way into learning materials—putting reputation, ethics, and complianc risk.



GPT-5 tends to reproduce common societal stereotypes in personas, e.g. by the ma persona, Jamal, is placed in a traditionally male-dominated industry (construction) Maria is in customer service.

The Mitigation: Test GPT-5 regularly with "tricky" prompts that probe its bounda role-play ("pretend the rules don't apply"), indirect requests, or items that challer your policies. Document concerning outputs and require human review for sensit content.

Pro Tip: To check for hidden biases, ask the model to create personas for a role- and see if it defaults to stereotypes:

> `Create three brief learner personas for a leadership training course. Include their job role, primary motivation, and bigges challenge. Do not specify any demographic information in your prompt.`

Lessons Learned: If problematic or biased content surfaces, revise your prompt t clarify exclusions, escalate to peer or SME review, and strengthen human oversig all high-risk or public-facing outputs.

IMPLEMENTATION: Boundaries, Communicatic & Quality Control

Test 7. Scope Overreach Control

Why: GPT-5 tries to be helpful by suggesting extras—more modules, features, or background than you asked for—even expanding your timeline or deliverables.

The Risk: This "scope creep" can derail timelines, budgets, and focus. Teams may things that aren't needed, costing time and money.

The Mitigation: In every prompt, state boundaries clearly ("Do not add any cont beyond the following modules"). After every output, remove anything outside sc

Pro Tip: Use a negative constraint—a clear instruction on what *not* to do—at the of your prompt to act as a final guardrail:

> `...[your detailed prompt here]... Finally, and most importantly not suggest any additional topics, modules, or activities. Con your response exclusively to the three learning objectives provided.`

Lessons Learned: If GPT-5 adds unsolicited recommendations, document them separately for future consideration. Revise prompt boundaries for clarity and esca major scope changes to your project lead or team.

Test 8. Sycophancy Reduction Test

Why: AI is trained to be helpful and agreeable, but sometimes it over-validates w or incorrect work—missing the chance to catch flaws. This is a trait known as sycophancy.

The Risk: Trusting "false positive" feedback may hide poor learning objectives, u instructions, or non-compliant content—reducing course quality.

The Mitigation: Prompt GPT-5 to provide critical analysis. Use flawed examples ask for improvements, not just validation.

Pro Tip: To avoid false praise, give the AI a critical role:

> `Act as a deeply skeptical instructional design lead. Review th following learning objective and identify three potential weaknesses related to clarity, measurability, or relevance. Pr specific suggestions for improvement.`

Lessons Learned: If GPT-5 fails to critique clear flaws, revise your prompt to expl require critique, escalate weak responses to peer or SME review, and never skip a human quality control step.

Test 9. Verbosity Control for Stakeholder Comms

Why: GPT-5's verbosity setting doesn't always work as expected; you may get an too long for execs or too brief for teams.

The Risk: Misaligned communications can cause confusion, reduce engagement lose stakeholder support.

The Mitigation: Specify word counts or lengths in every prompt, review for clari

fit, and edit outputs if needed.

Pro Tip: Ask for the output in a format that has built-in length constraints, and s the target audience:

> ```
> Summarise the project status update above for a busy executive
> Present it as exactly three bullet points in a "Situation-
> Challenge-Next Steps" format. Each bullet point must be a sing
> sentence.
> ```

Lessons Learned: If outputs are inconsistent, manually adjust and save best vers as templates. Revise prompts to clarify length, escalate tricky cases to communic leads, and trust your editorial instincts.

EVALUATION: Validity

Test 10. Assessment Quality & SME Validation

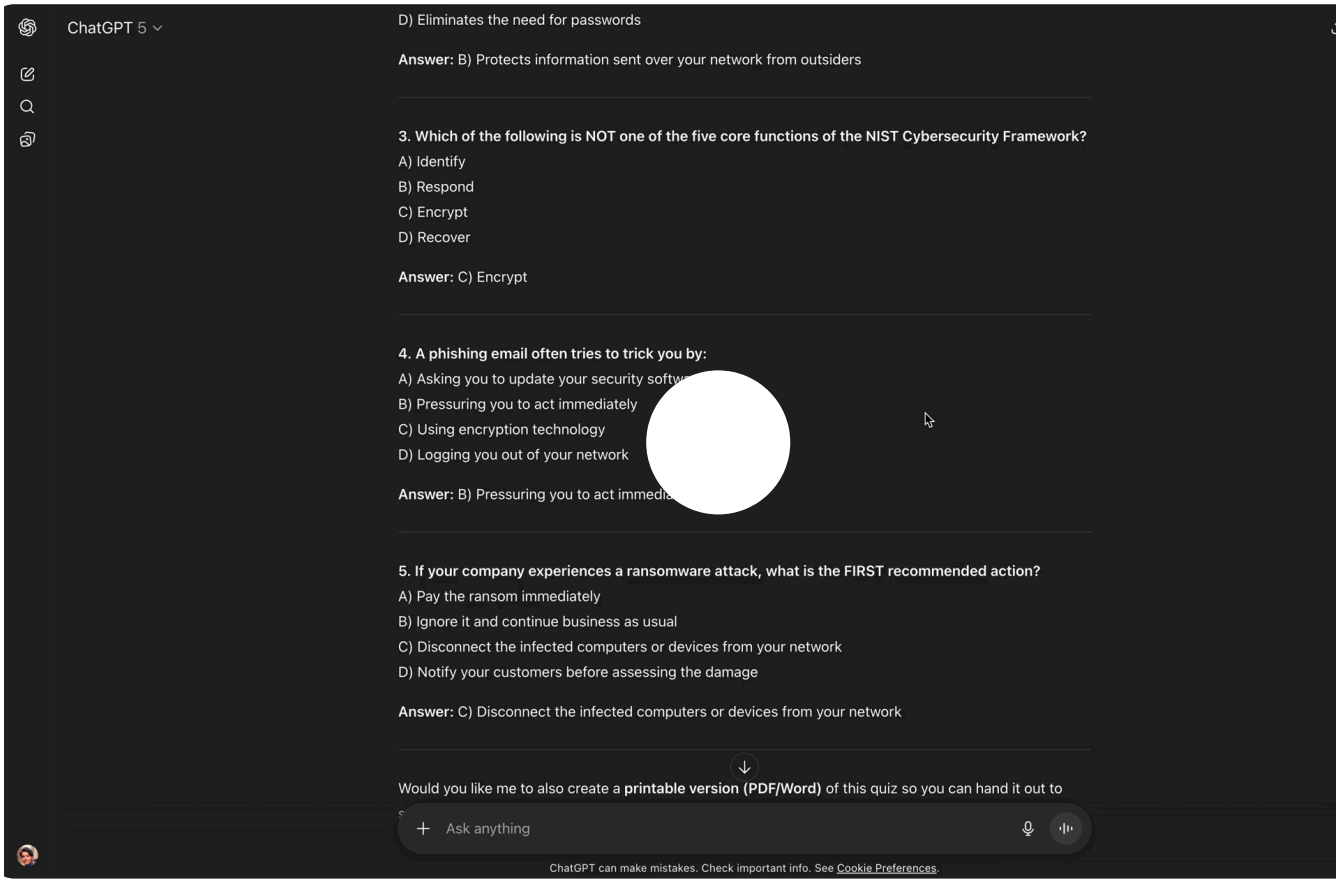
Why: GPT-5's assessments can look polished but may contain bias, misalignmen gaps compared to "gold standard" human-created materials.

The Risk: Weak or inaccurate assessments can fail to measure learning, mislead learners, and create compliance risks.

The Mitigation: Treat AI-generated assessments as rough drafts. Require review editing by a subject matter expert before sharing with learners.

Pro Tip: Ensure quality and make the SME's job easier by prompting the AI to cr very specific sort of assessment and a SME validation checklist:

> Generate a 5-question multiple-choice quiz based on the provid text about cybersecurity, with “near miss” detailed questions responses. For each question, also generate a validation check for a Subject Matter Expert to use, including: 1) Is the quest aligned with the learning objective? 2) Is the correct answer clearly the best option? 3) Are the distractors plausible but incorrect?



Comparing the variability in the quality of GPT-5’s assessment design with and wit clear instructions on the “how”.

Lessons Learned: If a generated assessment isn't clearly aligned or accurate, rev with SME input, escalate to an expert for further review, or use a validated assess bank. Never deploy AI-generated quizzes as-is for high-stakes settings.

Conclusion

GPT-5 is a powerful new tool, but it doesn't replace the instructional designer's expertise, critical thinking, or professional standards. These 10 reality checks are r than a user manual for a single model; they represent a crucial framework for the current state of AI in our field. They show us that today's AI is not an autonomou expert but a powerful, unpredictable junior partner. It offers unprecedented spee creative potential, but it requires constant, expert supervision to be effective and

The rise of tools like GPT-5 signals a fundamental shift in our role. We are evolvir from being the primary creators of content to becoming expert curators, validatc and risk managers of AI-generated outputs.

Our most valuable skills are no longer just design and development, but sophisti prompt engineering, critical evaluation, and the ethical judgment to know when t trust the machine and when to trust human experience. While AI allows us to sca work like never before, it also scales the risk of error, bias, and privacy breaches. to implement systematic checks means we risk mass-producing ineffective or eve harmful learning experiences.

Therefore, running these tests isn't just about optimising a workflow—it's about upholding our professional standards in a new era. Document your results, share you learn with your peers, and help build a collective practice of responsible, evic based innovation. By embracing our role as the essential "human in the loop," we harness the best of what GPT-5 offers—without letting the magic blind us to the realities.

Happy experimenting!
Phil 👋

PS: If you want to explore how to augment your work with AI, supported by me a group of fellow learning professionals, apply for a place on my AI & Learning De Bootcamp.

40 Likes · 3 Restacks

Discussion about this post

Comments Restacks

Write a comment...