

SAHIL FEAT. CLAUDE / FIELD GUIDE / OCTOBER 2026

The Human-AI Creative Collaboration Field Guide



Claude shoots possibilities. Sahil calls cut.

Directed by Sahil Tayal · featuring Claude (Anthropic)
sahilfeatclaude.sahilwho.online · sahiltayal.com

A field guide, not a manifesto

This guide accompanies SAHIL FEAT. CLAUDE (sahilfeatclaude.sahilwho.online). It is for anyone who makes things with an AI model and wants the result to be better, truer and more their own, not just faster.

Claims are labelled. Evidence means published research, cited at the end. Interpretation means our reading, which you may dispute. Hypothesis means testable but untested here. Everything unlabelled is practical advice drawn from experience; treat it as a starting point.

The one-line version: *the model shoots possibilities; you call cut, and you answer for the cut.*

PART 1

The Director's Loop

Eight steps. The human owns the ends; the model is most useful in the middle. Skipping a step is the most common way collaborations go wrong.

Step	You do	The model can	If skipped
1 Intend	Decide what is worth making and why	Ask clarifying questions	Polished work about nothing
2 Brief	State audience, purpose, constraints, non-goals	Point out gaps in the brief	Plausible generic defaults
3 Generate	Ask for range, not one answer	Produce many varied takes fast	Anchoring on the first draft
4 Judge	Circle, strike, explain why	Argue for and against options	Taste outsourced
5 Verify	Check facts, numbers, code, accessibility	Suggest checks; never replaces them	Confident errors shipped
6 Revise	Give specific notes	Iterate tirelessly	Endless churn, no progress
7 Decide	Lock the cut; define done	Summarise trade-offs	Scope creep
8 Own	Publish, credit honestly, take responsibility	Draft credit lines	Blame-shifting to the tool

Creative direction that works

A useful brief has six parts: purpose (what should change for the reader), audience (one person you can picture), constraints (time, budget, format, tone), taste references (things it should feel like, and why), non-goals (what it must not be), and definition of done.

Weak direction	Strong direction
"Write a creative bio for me."	"120 words, first person, for a conference programme read by economists. Lead with the research question, not job titles. No adjectives like passionate or visionary."
"Make the website look amazing."	"Black and white with one red. Film-editing metaphor. One-second entrance. No gradients, no particle effects. Must pass WCAG AA on a cheap phone."
"Summarise this research."	"Summarise for a non-specialist in 5 bullets. Separate what the study measured from what the authors speculate. Quote nothing over 15 words. Give the DOI."

Interpretation: the strongest lines in a brief usually rule things out. Constraints are taste made explicit.

Ten common mistakes

- Accepting the first competent draft. Fix: ask for nine takes; reject at least five on purpose.
- Hiding your goal. Fix: say who it is for and what should happen after they read it.
- Revealing your favourite answer first. Models tend to agree (Evidence: Sharma et al., 2023). Fix: ask for the strongest objection.
- Trusting citations you did not open. Fix: open every source; cut what you cannot find.
- Measuring feelings of speed. Fix: time to a finished, checked result (see METR, 2025).
- Letting scope balloon. Fix: write the definition of done before generating anything.
- Losing your voice. Fix: write the first and last lines yourself; read it aloud.
- Over-iterating. Fix: stop when the next note is a matter of taste, not of truth.
- Pasting secrets or private data. Fix: strip names, keys and credentials before sharing context.
- Vague credit. Fix: say plainly what the model did and what you did (Part 7).

Twenty evaluation questions

Purpose

- Should this exist at all?
- Who is better off after encountering it?
- What is the one thing it must do?
- What did I decide that the model could not?

Truth

- Which sentences are facts, and have I checked each?
- Is any number un-sourced?
- Is anything here confidently vague?
- What would an expert in this field object to?

Craft

- What is the weakest part, and why is it still here?
- Does every element earn its place?
- Does it work for someone using a keyboard, a screen reader, a small screen?
- Would it survive being half as long?

Voice

- Does it sound like the person whose name is on it?
- Which phrases are generic defaults?
- Is there one surprising, specific detail?
- Would I say this out loud?

Consequence

- Who could be misled or harmed by this?
- Is the AI contribution described honestly?
- What am I responsible for if it is wrong?
- What would I change if I made it again tomorrow?

Verification practices

Evidence: language models can state false things fluently, including non-existent citations; developers document this (e.g., Anthropic, "Reduce hallucinations"). Interpretation: fluency is not evidence; polish lowers our guard.

Item	Check
Facts	Open the primary source; confirm it says what the summary claims.
Numbers	Re-derive from the original table; check units, sample, year.
Quotes	Match word for word; keep them short; attribute precisely.
Citations	Resolve the DOI or URL; confirm authors, title, venue.
Code	Run it; test edge cases; test on modest hardware; read the diff.
Design	Keyboard, contrast, reduced motion, 360px width, screen reader labels.
Claims about yourself	No invented metrics, testimonials or credentials.
Keep an UNVERIFIED label on anything you have not checked, until you have.	

What the evidence says (and does not)

- Noy & Zhang (2023, Science): 453 professionals on short writing tasks; with ChatGPT, time fell about 40% and graded quality rose about 18%. Short, self-contained tasks only.
- Brynjolfsson, Li & Raymond (2025, QJE): 5,172 support agents; productivity rose on average, mostly for less experienced workers; the most experienced saw small quality declines.
- Dell'Acqua et al. (Organization Science): consultants did better with AI on tasks inside its capabilities and worse on a task designed to sit outside them: a jagged frontier.
- Doshi & Hauser (2024, Science Advances): AI story ideas raised rated creativity, especially for less creative writers, but made stories more similar to each other.
- Becker et al. / METR (2025): 16 experienced developers took about 19% longer with AI allowed, yet believed it had saved them about 20%. Specific setting; early-2025 tools.
- Interpretation: the effects are real, uneven and context-dependent. Test in your own work rather than trusting averages.

A short creative experiment: Nine Takes (60 minutes)

- 5 min. Write one ordinary idea in under ten words.
- 10 min. Write a brief: audience, form, tone, budget, time, one non-goal.
- 5 min. Ask a model for nine different interpretations (literal, reversed, miniature, epic, documentary, comic, ritual, tool, question). Or use the Cutting Room simulator on the website.
- 10 min. Circle at most three. Strike the rest. Write one sentence on why for each circle.
- 5 min. Make a tenth take by hand that none of the nine suggested.
- 20 min. Build the smallest honest version of your favourite.
- 5 min. Reflect below.

Reflection

Which decision took the longest, and why?

What did the tenth (hand-made) take have that the nine did not?

What would you credit to the model, and what to yourself?

Honest credit lines

- "Written by [you]. Drafting assistance from [model]; all claims checked by the author."
- "Concept and direction: [you]. Code drafted with [model], reviewed and tested by [you]."
- "Made by [you] with [model]. Mistakes are mine."

FURTHER READING

Sources and the Sahilverse

- Becker, Rush, Barnes & Rein (2025). Measuring the Impact of Early-2025 AI on Experienced Open-Source Developer Productivity. arxiv.org/abs/2507.09089
- Brynjolfsson, Li & Raymond (2025). Generative AI at Work. QJE 140(2). doi.org/10.1093/qje/qjae044
- Dell'Acqua et al. Navigating the Jagged Technological Frontier. Organization Science. doi.org/10.1287/orsc.2025.21838
- Doshi & Hauser (2024). Generative AI enhances individual creativity but reduces the collective diversity of novel content. Science Advances 10(28). doi.org/10.1126/sciadv.adn5290
- Lichtenstein & Slovic, eds. (2006). The Construction of Preference. doi.org/10.1017/CBO9780511618031
- Noy & Zhang (2023). Experimental evidence on the productivity effects of generative AI. Science 381(6654). doi.org/10.1126/science.adh2586
- Sharma et al. (2023). Towards Understanding Sycophancy in Language Models. arxiv.org/abs/2310.13548
- Anthropic. Reduce hallucinations. platform.claude.com/docs/en/test-and-evaluate/strengthen-guardrails/reduce-hallucinations

Explore the Sahilverse

- sahiltayal.com: the canonical home of the Sahilverse
- sahilfeatclaude.sahilwho.online: this project
- sahilfeatchatgpt.sahilwho.online: the first duet
- sahilwho.online: Sahil Who? identity lab
- sahilverse.online: the Impossible Machine catalogue

Write to Sahil: sahil@sahiltayal.com

Drafted with Claude (Anthropic) under the direction of Sahil Tayal, October 2026. Not legal, medical or financial advice.