WEEK 02 · Monday, September 14, 2026

Language Models for Research, Ideation & the Brief

A thinking tool, not a writing tool. Where it helps, and where it quietly lies.

What did you find?

Which one did you not realize was an AI?

How many did you choose — and how many just arrived?

Where does one get you wrong — and would you find out?

Since we last met.

Eight days in September 2026: three shipping announcements and one resignation.

SEP 3 · OPENAI

GPT-6 Astra ships.

OpenAI's new flagship. It can drive a computer — screen, keyboard, mouse — and is the first model the company rates “Critical” for cybersecurity.

SEP 8 · OPENAI

Images 2.5 lands.

A new image model in ChatGPT, Work and Codex. Two flavours: Flare for speed, Sunburst for precision. Up to 50% faster than Images 2.0.

SEP 3 · NVIDIA

Hugging Face, acquired.

$12.9B for the open-model hub — Nvidia's second-largest deal after Groq. The open ecosystem now has a chipmaker for a landlord.

SEP 8 · ANTHROPIC

A researcher walks out.

Jacob Coxon quit, saying the labs are racing to self-improving superintelligence and that insiders privately fear it could end us by 2030.

The four D’s.

Discover, Define, Design, Develop. Every project in this class runs in that order.

DISCOVER

What is already true

Market and competitor research. Who it is for, and what already exists in the space.

Today, before the break

DEFINE

What you are making

Sitemaps and wireframes for a website. Sketches and a written brief for everything else.

Today, after the break

DESIGN

What it looks like

Static comps built on the structure you already defined. Identity, illustration, layout, type.

Weeks 3 through 7

DEVELOP

The finished output

Build the site, prep the files for print, write the brand guideline, ship the thing.

Weeks 8 through 10

Skipping Discover is the oldest mistake in design. AI has made it much easier to skip.

Today has two halves.

Same tool, two different jobs. Checking claims, then generating options.

FIRST HALF · DISCOVER

Ask three models what already exists here, and who it is for. Then check every claim against a real source.

SECOND HALF · DEFINE

Twenty concepts instead of your first one, argued down to three. Then a brief you write out of what you found.

No image generation today — that starts next week, with the same discipline.

1:40 Debrief — **2:40 Discover** — **3:40 Break** — **3:55 Three moves**

What you walk out with.

Four things, all of them going into your Process Journal.

CHECKED RESEARCH

What already exists and who it is for, with every claim marked verified or flagged.

A DIRECTION

One concept you picked out of twenty — and can explain why you picked it.

A WRITTEN BRIEF

A paragraph in your own words: what it is, who it is for, how it should feel, what it should not be.

AN IDEATION LOG

Everything from the three moves, including the parts you threw out.

ONE RULE FOR THE SESSION

You type the final version of anything you keep. Do not copy and paste.

The test: point at one decision you made that the model did not suggest and say why you made it.

Sixty years of chat bots.

You have seen this interface before. What sits behind it has changed completely.

1966

ELIZA, at MIT

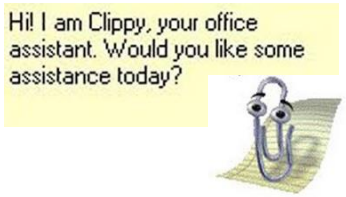
About 200 lines of code. It handed your words back as a question.

Joseph Weizenbaum

1972 – 2010

Scripted bots

PARRY, Clippy, SmarterChild. Hand-written rules. Off script, it broke.



Rules, written by people

2011 – 2021

Voice assistants

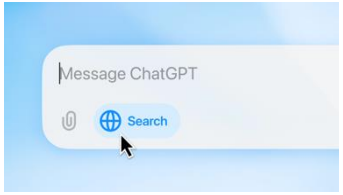
Siri, Alexa, Google Assistant. Speech in front, fixed commands behind.

Still a menu, spoken

2022 –

Language models

ChatGPT, Claude, Gemini. Every sentence generated as it goes.



No script left to check

Same box since 1966. What changed is that there is no script behind it any more.

Weizenbaum was horrified.

He wrote ELIZA to prove the conversation was hollow. People confided in it anyway.

His own secretary asked him to leave the room so she could talk to it in private.

Psychiatrists proposed using it for real therapy. He argued against it for the rest of his career.

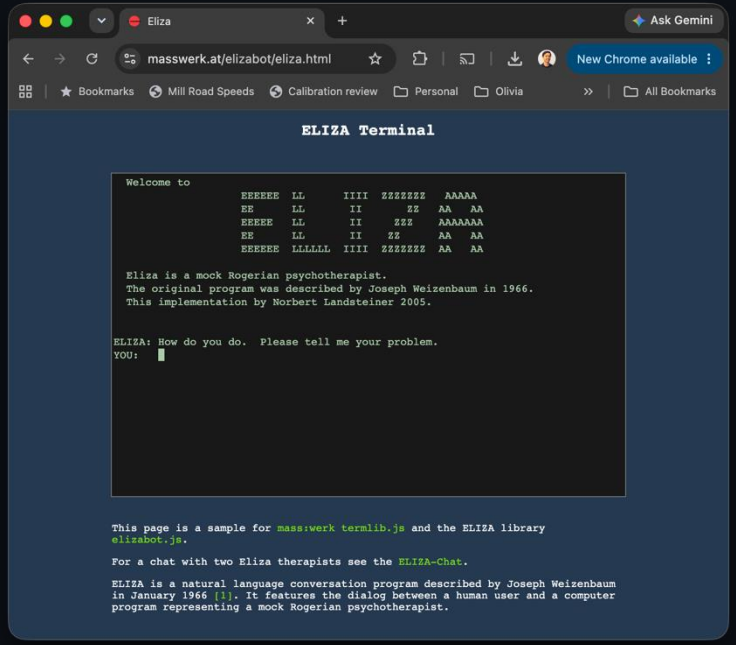
Nothing in ELIZA understood anything. Keyword matching was enough.

We read fluency as understanding. That reflex is sixty years old.

Try it yourself.

A faithful browser port of the 1966 program — no install, no login, still unnervingly effective.

> Norbert Landsteiner's JavaScript port of Weizenbaum's original script.
>
> masswerk.at/elizabot/eliza.html



What actually changed in 2017.

The interface stayed the same. The thing behind it was replaced.

2017	2018 – 2021	Nov 2022	2023 – now
The transformer	**Scale**	**ChatGPT**	**Everything at once**
A Google paper. Training runs in parallel, so scale starts to pay off.	GPT-1 to GPT-3 is roughly 1,500× bigger. Abilities nobody planned.	A tuned model taught to follow instructions. The public showed up.	Text, images, audio, video. Agents that use tools.
Everything descends from it	**GPT-2 was held back as risky**	**Fastest adoption on record**	**Still guessing next words**

One 2017 paper about how to read a sentence sits underneath every tool in this class.

All it does is guess the next word.

PREDICT

Your prompt becomes numbers. The model scores every possible next token, picks one, then does it again.

TRAIN, THEN TUNE

It learns what usually follows what, then people rate its answers and it is tuned toward what they liked.

CONTEXT, NOT MEMORY

It only sees the current conversation. Close the tab and it is gone; long chats push the start out of view.

Nothing in there stores facts. It stores what text tends to look like and uses statistics to determine the next word.

It would rather answer than say no.

MAKES THINGS UP

Rewarded for guessing

Graded like a multiple-choice exam. A guess might score; “I don’t know” never does. So they bluff.

Why Language Models Hallucinate, OpenAI, 2025

AGREES WITH YOU

Tuned on your thumbs-up

Raters reward answers that feel good. An April 2025 ChatGPT update shipped so flattering it was pulled in days.

Rolled back in under a week

IT GETS REPEATED

Fluent, confident, invented

A public database of court filings citing cases that do not exist passed 1,900 in August 2026.

damiencharlotin.com

Live: ask for a month with a Q in it or to spell “seventeen” and count the E’s.

What a language model is actually for.

Not a search engine, not an oracle, not a designer. Three things it does well, three it does badly.

GOOD AT

- **Volume you can reject from.** Twenty options in ninety seconds. You only need one, but you need to see twenty to know which.
- **Arguing against you.** When prompted, it will say what a critic would say.
- **Naming what you already sense.** Putting words to the thing you cannot quite articulate yet.

BAD AT

- **Deciding.** “Which of these is best?” is where authorship quietly moves from human to AI.
- **Knowing what is good.** It writes confident prose about bad ideas. Confidence is not quality.
- **Knowing what is true.** Fluency is not evidence, and it cannot tell you when it is guessing.

One rule all afternoon: you type the final version of anything you keep. Not copy-paste, not “regenerate until it’s good.” As you type, make your own edits and decisions.

A hallucination reads exactly like a fact.

AI OUTPUT, UNCHECKED

"73% of students find events like this on Instagram, not email."

The model invented the number. No such survey exists — but it reads specific, confident and plausible.

Result: You publish a fabrication as fact — you become the source of the misinformation

AI OUTPUT, SOURCE-CHECKED

Same sentence — now with a source you opened, read, and can cite.

You traced it to the real survey — or confirmed none exists and said so. Both are real answers.

Result: Fact checked and defendable

What to compare across models.

SUGGESTIONS

What each one reaches for first

REFUSALS

What it won't do, and how it says so

HEDGING

Which claims it softens with "may" or "often"

CONFIDENCE

How sure it sounds when it is wrong

SOURCES

Cited, vague, or none at all

GAPS

What it left out that another caught

FRAMING

Whose interests the answer is written for

ERRORS

What it got flatly, checkably wrong

Open source, open weight, closed

Open Source

Weights + code + training data

The full recipe is published.

You can retrain it from scratch.

Rare at the frontier.

e.g. OLMo, Pythia

Open Weight

Weights only, license-bound

Download and run it locally.

Fine-tune it, but no training data.

The license sets the limits.

e.g. Llama, Gemma, Qwen

Closed

API access only

Frontier weights stay in-house.

The model can change under you.

No offline or on-prem use.

e.g. GPT, Claude, Gemini

Why it matters: what you can inspect, host, and depend on changes at every tier

Which model, and how much effort

CLAUDE MODELS · how much capability

Fable 5.1
Toughest challenges — requires usage credits

Opus 5
Complex tasks — currently selected

Sonnet 5
Most efficient for everyday tasks.

Haiku 4.5
Fastest for quick answers

CLAUDE EFFORT · how much “thinking”

Low
Fastest, least thorough

Medium
Balanced

High — default
Thorough; the standard setting

Extra
Slower and more exhaustive

Max
1.5× or more usage

Match the model to the task first — then raise effort only when the problem is genuinely hard

Menus shown are Claude's; other assistants use different names for the same two dials

Bring your own material

DOCUMENTS

PDFs, decks, briefs

Research reports, client briefs, standards, style guides.

Ask for themes, contradictions, or a one-page summary.

“Top 5 pain points here?”

IMAGES & SCREENS

Screenshots, mockups, photos

Wireframes, competitor UI, moodboards, whiteboard shots.

Ask for a critique, a redraw, or the palette and type used.

“Critique this flow.”

DATA FILES

CSV, Excel, JSON

Survey exports, analytics pulls, usage logs, backlogs.

Ask for clusters, outliers, or a chart you can drop in a deck.

“Cluster these replies.”

Upload the artifact, then ask the question — grounded input beats a model guessing at your context

Giving it a memory

MEMORY FILES

CLAUDE.md and friends

A plain text file the model reads at the start of every session.

Holds project context, conventions, and what not to do.

“Always cite sources.”

SAVED PREFERENCES

Standing instructions

Written once in settings, applied to every new conversation.

Tone of voice, output format, tools and units you use.

“Use UK spelling.”

PAST CHATS

Recall across sessions

Some assistants can look back at earlier conversations.

Useful for long projects; check what is retained and shared.

“Pick up Tuesday’s audit.”

Write your preferences down once — otherwise every conversation starts from zero

Three models. Where to get them.

One Tab

Aggregators, many models

Fastest way to get three.

One login, one paste box.

Someone else's wrapper, though.

e.g. Poe, Anakin AI

Three Tabs

The products, as shipped

Each one's real defaults.

Its own refusals and hedges.

Slower, and worth the tabs.

e.g. ChatGPT, Claude, Gemini

With Sources

Search-shaped, links attached

Answers arrive with links.

Good for finding where to look.

A citation is not a check.

e.g. Perplexity, NotebookLM

For the lab, use three tabs. A wrapper hides the defaults you came to compare.

LAB
Discover first.
Define after the break.

Pick your subject. Ask three models what exists and who it is for, then go check.

Before you open anything: write down the first three concepts you'd come up with on your own. Those are the ones you ban.

Pick something to work on.

Five minutes. A subject you would be happy to spend an afternoon with — and roughly what you would make of it.

BEFORE YOU OPEN ANYTHING

Write down the first three concepts you would come up with on your own. Those are your defaults — the same obvious moves a model reaches for first. You will ban them after the break. Keep the paper.

NOTHING COMES TO MIND? TAKE ONE OF THESE

A 24-hour laundromat that rents its floor for events after midnight

WHAT YOU'D MAKE

Brand identity

WORN-OUT IMAGE TO BAN

Bubbles, a spinning drum

A repair café that fixes things people were told to throw away

WHAT YOU'D MAKE

Poster series, 5+

WORN-OUT IMAGE TO BAN

A wrench, a lightbulb

How a bodega actually makes money

WHAT YOU'D MAKE

Editorial feature

WORN-OUT IMAGE TO BAN

The neon awning at night

You will need three models open in three tabs and one document to paste into.

DISCOVER
Three passes. 60 minutes.

Your own subject, or one of the three on the handout. Three models, three tabs.

01

LANDSCAPE

15 min

02

THE PEOPLE

12 min

03

THE CHECK

23 min

Then ten minutes: we put the verified claims and the flagged ones in two columns on the board.

Pass 1 — Landscape.

Find out what already exists here before you decide what to add to it.

All three models, one question: keep each answer in its own document. You compare them later.

THE ASK

Ask what already exists in this space: who is making something like it, what it looks like, who it is aimed at, and what nobody is doing. Then take two of the examples it named and go looking for them.

YOUR MOVE — WRITE IT DOWN

Note what you could not find. A model will name work that does not exist, and catching one of those is worth more than the whole list it arrived in.

WHY START HERE

You cannot find a gap in a field you have never looked at. Neither can a model, but it will describe one anyway.

15 MINUTES

Long enough for three answers and two failed searches.

Pass 2 — The people.

Same three models, same subject. Now ask who it is actually for.

No verifying yet: collect all the claims first, then check them in one pass.

THE ASK

Ask who this is for, how they already spend their attention, what would reach them, and what would put them off. Keep each model's answer separate so you can see where they part company.

YOUR MOVE — WRITE IT DOWN

Write down where the three disagreed. Where they agreed, that is not three sources. It is the same training data three times.

WHY THREE MODELS

Three models trained on much the same internet will agree with each other. That agreement proves nothing.

12 MINUTES

Shorter than the first pass on purpose. This is collection. The checking comes next.

Pass 3 — The check.

Go through every claim you collected and find out which ones are true.

The longest block of the lab: it is the part everyone skips, and the part the assignment is graded on.

THE WORK

Mark every claim you have collected as verified against a source you opened, or unverified. A model's own citation is a place to look. Open it and see whether it says what the sentence claimed it said.

YOUR MOVE — KEEP THE FLAGS

A claim you could not source is a finding, so do not delete it. What survives checking tends to be numbers and behavior. What does not tends to be personality.

WHAT COUNTS

A second model agreeing with the first is not a check. Go to the organization, the data, the actual portfolio.

23 MINUTES

The biggest block of the hour, and the one the homework is built on.

Pool — what survived the check.

Two columns on the board: the claims you verified, and the claims you could not.

Bring your marked list: we are reading the pattern across the whole room, not grading anyone's research.

THE ASK

Verified claims go in the left column, the ones you could not source go in the right. Read the board across the room rather than down your own list.

WHY POOL IT

One flagged claim is an anecdote. The whole room's flagged claims are a map of where these models invent.

YOUR MOVE — WRITE IT DOWN

Which kinds of claim survived? Numbers, names and dates fail differently from general description. That difference is the finding.

10 MINUTES

Long enough to fill both columns and say what the right-hand one has in common.

DEFINE
Three moves. 60 minutes.

Same subject you researched. Bring your verified claims and your flags.

01

DIVERGE

18 min

02

INTERROGATE

15 min

03

BRIEF

20 min

Then twenty minutes: find someone making a similar kind of thing and read each other your briefs.

Move 1 — Diverge.

Your first idea is the one everybody else also had. Ask for twenty, then throw out the first eight.

The three you wrote before the lab: those are the three you ban here. Nothing on your list may lean on them.

THE ASK

```
Give me 20 distinct visual directions for [MY SUBJECT].
Rules: none may rely on [CLICHE] or my own three ideas.
Vary the register — some literal, some metaphorical,
some absurd. One line each, plus the single image that
would carry it. Number them. Don't explain yourself.
```

THE FOLLOW-UP — THIS IS THE ONE THAT MATTERS

```
Concepts 1–8 are ones I'd have reached on my own.
Throw them out. Give me 10 that sit further from
the center of this subject, and say what makes them
further out.
```

WHY TWENTY

You are building a field wide enough that rejecting nineteen means something.

WHY THROW OUT EIGHT

The top of the list is the statistically obvious answer — not where your work should sit.

Move 2 — Interrogate.

Make it argue against the three concepts you actually like.

Before you paste anything: pick your three strongest concepts and rank them.

THE ASK

Paste your three strongest concepts. Argue against all three: the cliché each is closest to, the audience it loses, what a critic would say it is avoiding. Agreement is forbidden.

THE FOLLOW-UP — THE DECISION IS YOURS

Which objection do you accept? Write one sentence naming it, and what you are changing because of it.

WHY ARGUE

Agreement is the default setting. You have to switch it off to learn anything.

15 MINUTES

Long enough to be argued with, short enough that you cannot start over.

Move 3 — Brief.

You write first, out of what you found before the break. Then hand it over, and forbid rewriting.

One paragraph, no model: subject, the audience you verified, what it should feel like, what it must not be.

THE ASK

Hand over your paragraph and forbid rewriting. Ask only for the ambiguities: what could be read two ways, what is missing, what a reader would guess wrong.

THE FOLLOW-UP — THE DECISION IS YOURS

You resolve every ambiguity yourself, in your own words. That resolved brief is what you submit.

WHY YOU WRITE FIRST

If the model writes the brief, the brief is the model's. Authorship starts on the blank page.

20 MINUTES

Ten to write it yourself, ten to interrogate what you wrote.

Share out — read each other your briefs.

Twenty minutes. Turn to the person next to you. No feedback, no fixing — just read.

Same three moves, different subjects: the interesting part is where your two processes went apart.

THE ASK

Read your brief out loud and listen to theirs. Then say what the models pushed you toward, and where you overruled them.

WHY OUT LOUD

A brief you cannot say out loud is not finished. Someone else's ear finds that faster than your own eye.

YOUR MOVE — WRITE IT DOWN

Name one thing the research changed about your direction. If nothing changed, the research was decoration.

20 MINUTES

Ten each. Do not spend it fixing — you are comparing processes, not editing copy.

What tends to go wrong.

Six failure modes. Most of them feel like progress while they are happening.

You skip the checking.

A profile you did not check is a stereotype, and it will feel like truth the whole time you design against it.

You take the first list.

The top of a list is where the model is most average. The interesting material is further down, or in the second round.

It goes soft on you.

Ask for harsh and you get it for a turn or two before it drifts back to encouragement. Push again, or open a fresh chat.

You let it write your brief.

What comes back sounds smoother, says less, and none of it is yours anymore.

You mistake fluency for quality.

These models write just as confidently about bad ideas as good ones.

You hand over the judgment.

“Which of these is best?” is the question that gives your authorship away. Ask it to argue, then decide yourself.

Where does the idea come from?

Not a reason to refuse the tool. A reason to know what you are doing while you use it.

THE UNCOMFORTABLE PART

A model can hand you twenty concepts because it has read an enormous amount of other people's thinking — without asking and without paying — in the same way image models have read other people's pictures.

QUESTION ONE

If a concept came out of a numbered list and you picked it, whose idea is it? Does your answer change if you picked it out of twenty instead of two?

QUESTION TWO

Next week you will disclose AI use in your image generations. Should you disclose it in your thinking? Should the two be treated the same way? If you used an internet search engine for the discover, define, and ideation process is that the same?

No answer needed today. You do need a position by the time you write your AI-Use Policy in Week 8.

Week 2 Homework

Due at the start of Week 3

1. FINISH — Polish and write up everything from today's lab: the research with every claim verified or flagged, the three-model comparison, your ideation log with the counter-arguments in it, and the brief you wrote yourself.

2. NOTICING — New this week, about ten minutes. Find ONE of those models' published training-data policy. Who trained it, on what, what do they claim? Two sentences.

3. REFLECTION — 250 words: which parts of your thinking you handed over this week, which you kept, and whether you would tell a client which was which.

If every claim checked out, you probably only kept the ones that would. The flags are the finding.