

• I — OPEN

DESIGN + AI SUMMIT

Recall versus inference

Building AI that actually knows you

Stuart Frisby • mrfrisby.com

Slide 1 of 63

MONDAY

Slide 2 of 63

On Monday morning, I sat down at my desk, and there was a briefing waiting for me telling me what my priorities that day were, and detailed meeting prep on the two people I was meeting that afternoon — including what we'd talked about last time and what I'd promised them. There was a two minute podcast of that same briefing, in my own voice, and a summary of what my team had been discussing across the previous week.

None of those things happened by magic. They happened because something remembered what I would otherwise have had to reconstruct from scratch, again, the way I do most mornings I don't have that document.

Booking.com → Deliveroo → Preply

Slide 3 of 63

So, quickly — who am I, and why you should listen to any of this. I'm Stuart Frisby. I've been doing design leadership jobs for about twenty years now. I ran the whole design discipline at Booking.com for a good while — 250-odd designers, absolute chaos — then went and did something similar at Deliveroo, with an eighty-person team. These days I'm Senior Director of Design at Preply — a global language learning platform, and I do a bit of consulting, I write a newsletter about design leadership, and very occasionally I speak at conferences like this one — but this is my first talk in years and years so if it goes badly it's actually due to a lack of due diligence on the part of the organisers, and not at all due to the middle-age-induced cognitive decline I have experienced in the years since I last spoke at an event like this one.

COSTA

Slide 4 of 63

I have a terrible memory. I think it's getting worse. Although I might've just forgotten how bad it always was.

Seven months ago as I started a new job and started forgetting new things I decided I needed to build myself a carer — a piece of software called COSTA that runs my working day, prunes my task list, preps my one-to-ones, drafts things in my voice, tells me every morning what actually matters. I've used it daily since. What I want to share is a report compiled from something I've been living inside of for seven months, and which has shaped my views on where design is going, and who we are going there with.

Recall, not inference.

Slide 5 of 63

This talk isn't about what AI can do for you. Everyone's already having that conversation, loudly, and mostly incoherently. It's about something much less exciting to say out loud, and much more useful to actually have: Why the real power of AI is not inference, but recall.

Interactive prototyping

Slide 6 of 63

Two things get talked about whenever “AI and design” are the subject of conversation. The first is interactive prototyping — tools that turn a rough idea into something clickable in a couple of minutes. That's not really a design conversation, whatever the marketing says. It's an engineering problem wearing a design costume. Useful, genuinely, but it has almost nothing to do with what makes a designer good at their job.

AI drawing interfaces

Slide 7 of 63

The second is getting AI to draw the interface itself. And AI is shockingly bad at that. Properly bad. You've all seen the evidence — generic layouts, decorative nonsense, components that don't correspond to any real design system, presented with total confidence as though somebody made a decision about them.

Figma Make

Slide 8 of 63

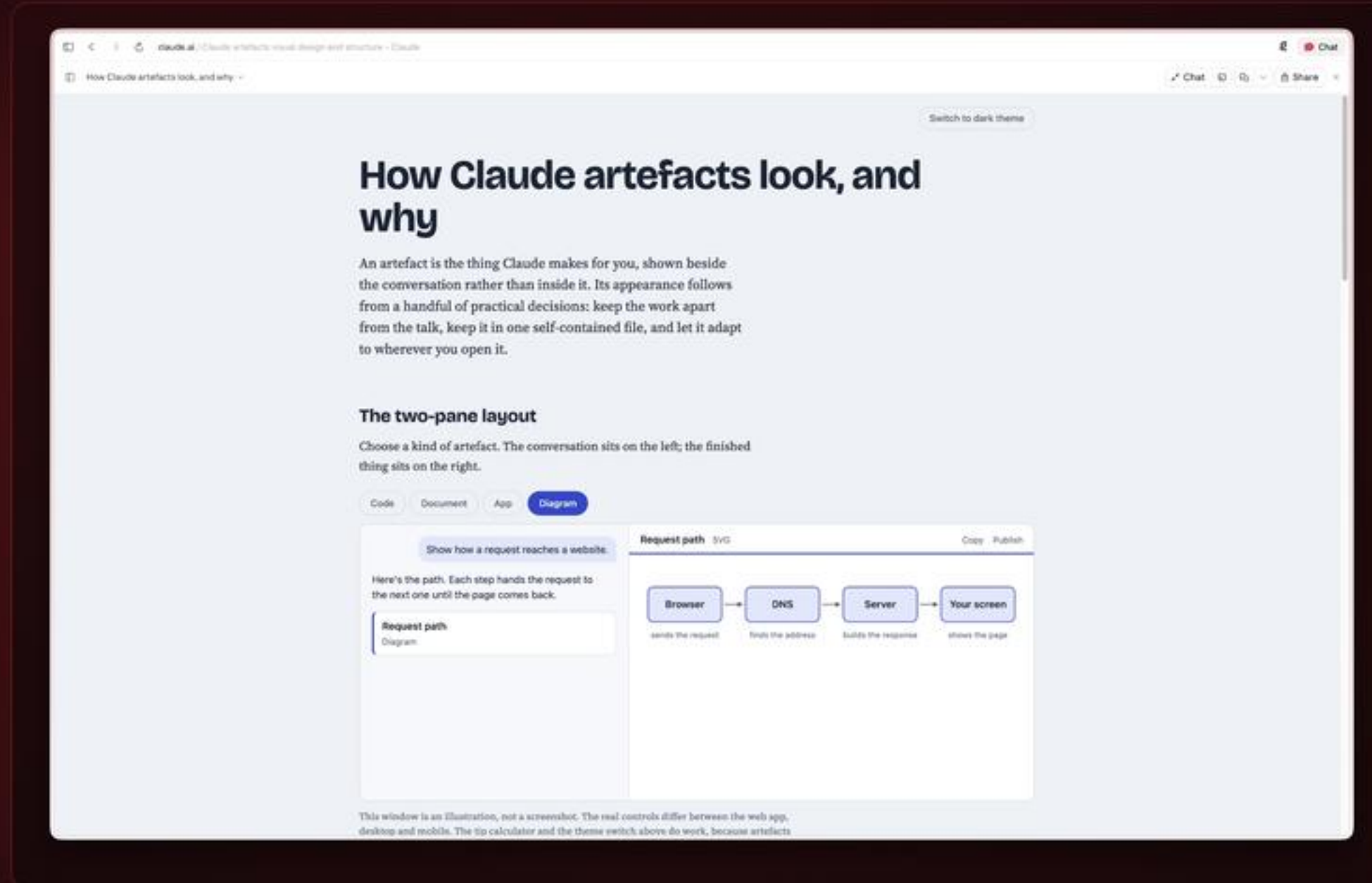
If the AI industry was being judged primarily on how bad Figma Make is at making things in Figma, the world would be a very different place right now.

Both of these design/AI tropes keep the conversation pinned to the same question: can these tools draw and code interfaces. That's the wrong question, and I also think it's really not that interesting a question in the first place.



Slide 9 of 63

Let's think about a different kind of design. A shitty poster for the local craft market — you know the one. Confident colour, garbled type, that one font they all use trying to do four jobs at once.



Slide 10 of 63

And its enterprise cousin: an endless stream of Claude artefacts — every single one of them somehow sharing the same typography, the same layout instincts, the same faintly over-confident voice. Different tool, different price point, identical failure.

An asset generator. Generating assets.

Slide 11 of 63

Neither of those is design. They are the output of an asset generator, generating assets. That's a fine thing for the world to have — it probably does need a fast way to knock out a poster for a craft market, although I'd love someone at OpenAI to demonstrate some form of taste. What design is, and what AI can do for design, is something else, and I think there really is something remarkable to be had here, if only we could see the beautiful wood for the terrible looking trees.

**Design is what happens
between your ears.**

Slide 12 of 63

Design is what happens between your ears. Not between calls to someone else's endpoint.

Abductive → Deductive

Slide 13 of 63

To design is to think. AI can't think for you. What it can do is to make thinking more efficient, and more effective.

There are, roughly, two modes available to you when you're solving a design problem you don't have complete information for — which is most of them. The first is abductive reasoning: the leap to the most plausible solution, made because you don't have enough information to justify anything more rigorous than a good guess. Most of design happens here, and there's nothing shameful about that. The second is deductive reasoning: working forward from what you actually know to be true.

The claim I want to make is a narrow one, which sounds really boring but is actually really exciting if you're a total nerd for design and you're here so...

**Capability without context
is just a very fast amnesiac.**

Slide 14 of 63

AI's real contribution to design isn't better guessing. It's shifting a significant portion of your guessing into knowing. Less abduction. More deduction. Not because AI got smarter, but because for the first time, it's actually viable for something to go and collect what you'd need to know, rather than you having to guess in its absence.

Capability without context is just a very fast amnesiac.

Not memory. Synthesis.

Slide 15 of 63

This is recall — but it's a bigger kind of recall than the word “memory” usually gets used for. Memory implies something was once known, and lost. A lot of what's actually useful here was never known by anyone in the first place — not lost, never held. One kind of that recall is synthesis: information that already existed, scattered, uncollected, sitting in research nobody reread and data nobody cross-referenced, turned for the first time into something you can actually reason from. The poster and the artefacts fail because there's no thinking behind them. The thing I'm describing works because, for the first time, there's something to think with.

The Force Multiplier Principle

Slide 16 of 63

And to be clear about what this doesn't do: it doesn't replace judgment. There's a framework I've used in design leadership for years, long before any of this — the Force Multiplier Principle. AI amplifies strong practitioners. It doesn't substitute for them.

**A weak designer with good
AI habits is still a weak designer.**

Slide 17 of 63

A weak designer with good AI habits is still a weak designer. Recall doesn't change who's doing the thinking. It changes what they've got to think with.

(Note for later: this only holds if what it's recalling is actually bounded — an unbounded guess dressed up as recall is worse than an honest one. That's the distinction the honest-limits section comes back to properly.)

Notebooks. Slack threads. Meeting notes.

Slide 18 of 63

Here's the other half of recall — the part that actually is memory, in the ordinary sense of the word. Something once known, properly known, by an actual person, and then lost. This one's universal. Every person in this room already lives it: who said what, what got decided, what you believe and why. It lives in a notebook filled up and never reopened. In meeting notes nobody rereads. In a Slack thread that had the right answer eleven months ago. In your own head, at eleven at night, reconstructing a decision you'd already made once before.

The working life generates its own record. Most of it disappears.

Slide 19 of 63

Organisations are actually pretty good at tracking what got done — tickets closed, projects shipped, calendars full of meetings that happened. What we've built almost nothing for is retaining the why. The reasoning, the tradeoffs, the thing that was actually said in the room that led to a decision — that survives, if it survives at all, as a handful of different people's slightly different memories of the same meeting, decaying at different rates from the moment it ends. The working life generates its own record, continuously. Most of it just disappears.

AI didn't create this problem. It's the oldest cost there is in knowledge work, paid quietly, by everyone, for as long as there's been knowledge work to do. What's different now isn't the problem. It's that fixing it has gone from aspirational to actually being almost completely fixed.

I wish I could get people half as excited about that as they are about boiling a lake so OpenAI can show them what they would've looked like in 1985.

V - INTRODUCING COSTA

● ● ● Search... (x)

COSTA |

Briefing

Meeting Prep

Journal

Tasks

Notes

URLs

Slack

Images

Calendar

Transcripts

People

Tags

Profile

Sun 28 Sep 12:38

Briefing

Good morning, Alex.

Today

Two 1-1s today — Maya at 10 and Jordan at 15:30. The onboarding review with Lucas is at 14:00; worth reading his proposal before that (it's on your task list, 1h estimate).

Four tasks flagged for today. The two high-priority ones are linked to meetings: Maya's design review feedback (45m, before 10am) and Lucas's onboarding flow review (1h, before 2pm). Marcus's API error state response can wait until this afternoon.

People

Maya Chen — last spoke Tuesday. She's presenting the notification redesign at 10:00. Her last 1-1 flagged scope creep concerns; worth having a clear position ready before the meeting.

Jordan Park — 16:00 hiring panel. No open threads from last week; straightforward.

Focus

The Q2 principles deck is the highest-leverage output this week. It's blocking the headcount ask and shapes the growth squad's Q2 design direction. Two hours of clear thinking this morning beats three rushed hours this afternoon.

Calendar

Tue 26 Mar Today

● 09:00 Design standup

● 10:00 Maya Chen 1-1

● 11:00 Focus time

me

● 13:00 Leadership sync

● 14:30 Onboarding review

● 15:30 Jordan Kim 1-1

● 16:30 Hiring panel

Meeting Prep +

Tasks

Today: 2 · All: 11

☐ H Review Maya's redesign proposal

☐ H Send Q2 principles deck

☐ H Draft headcount rationale

☐ H Follow up DS contract renewal

☐ L Respond to Marcus

View all tasks +

ASSISTANT

YOU

Can you draft a Q2 design strategy doc?

COSTA

Done. It covers goals, team structure, and four strategic bets for Q2. I've kept it to one page of signal.

Q2 Design Strategy.md

Draft · 847 words

I pulled from your profile knowledge agent (career context and working style), your Q1 strategy doc, and Maya's 1-1 last week (she flagged the component audit as a Q2 dependency).

Can you add a section on the DS hiring plan?!

**A chief of staff, for one
person, every day.**

Slide 21 of 63

So — COSTA, properly, for a moment. It's a chief of staff, built for exactly one person, running every single day.

Not a startup.
Not a pitch.

Slide 22 of 63

It's not a startup, and this isn't a pitch, don't worry. But, I have hated almost every piece of software I've ever used in my career, and so by building my own software I get to turn that hatred back on myself, which feels like a deeply British thing to do and maybe I am a patriot after all.

One palette. Twenty-four hours.



Slide 23 of 63

Personal software, built to your own exact shape is one aspect of AI that I think is really cool, and the more weird little one-person apps that exist to solve one-person-shaped problems, the better.

Small example, because I think it says more than the pitch does. The whole interface shifts colour continuously across the day — midnight indigo, dawn orange, noon violet, round through dusk and back again. None of that's scientifically accurate — but accuracy was never really the point. The day's supposed to feel like a journey, not a dial switching between warm and cool. It's a completely pointless feature by any commercial measure. No business case survives contact with a product manager. But that's rather the point of personal software. Nobody else was ever going to care enough to build it.

I cared enough to build a second, smaller tool just to tune it properly. I regret nothing.

SL DYN EASE STARLIGHT

$$\Delta H(h) = 40 - \frac{h - 7}{14} \times 10$$

$$t = \begin{cases} \max\left(0, 1 - \frac{h}{7}\right) & h < 7 \\ \max\left(0, \min\left(1, \frac{h - 7}{11}\right)\right) & h \geq 7 \end{cases}$$

$$H_o = (H_a + \Delta H(h)) \bmod 360^\circ, \quad S_o = S_a, \quad L_o = L_a + t^2 (0.88 -$$

Full 24-hour extension of SL dyn ease and the active production formula for COSTA. The ease parameter t is bidirectional: it runs forward from 0 at dawn ($h=7$) to 1 at $h=18$ in the daytime half, and in reverse — from 1 at midnight down to 0 at dawn — in the overnight half. This ensures the opposite colour is highly luminous during the small hours as well as the afternoon. The hue drift ($+40^\circ$ at dawn, decreasing linearly to $+30^\circ$ at dusk) continues unclamped into overnight hours, reaching $+45^\circ$ at midnight. Continuity at $h=7$ is guaranteed by construction: both branches evaluate to $t=0$ at dawn.

Slide 24 of 63

It even has a name in the codebase — SL dyn ease starlight — and an actual formula behind it, which I'm not going to make you sit through, but which absolutely exists, has been peer reviewed by nobody, and solves a problem literally no user has ever had.

I liked it too much to keep it locked away inside the product, so it's on the public landing page too now — costawork.app, if you want to go and watch it drift for yourself later.

I want to use COSTA as a way to frame what I have learnt. I have been building and using this thing for seven months and it has taught me lots about what recall requires, once you stop treating it as an input for inference, and instead start treating it as the whole point.

**Shape of work.
Not shape of a chat window.**

Slide 25 of 63

COSTA is designed around the shape of work, not shape of a chat window. It isn't a smarter box to type questions into. It's built around the actual shape of my working day, which meant almost everything about it had to be different from the tools I was already using.

ONE LAYER VERSUS THREE

GENERIC AI TOOL

- **Interactive layer**

Ask a question • get an answer

- **Automatic layer**

If it exists, it lives in their product

- **Corpus**

Exists, but isn't yours to take with you

Your context, held in someone else's product.

COSTA

- **Interactive layer**

Assistant • full corpus as context

- **Automatic layer**

Brief • meeting prep • task extraction

- **Corpus**

1-1 notes • transcripts • tasks • calendar

Context accumulates. Both layers draw from it.

Slide 26 of 63

Underneath, there's a distinction I think about as the waterline. Generic AI tools only ever operate at the visible layer — the chat window, the bit you can see, with your context living inside someone else's product rather than something you hold. COSTA runs three layers deep: a persistent corpus that just accumulates, quietly, in the background; an automatic layer that processes and synthesises that corpus without me asking it to; and only then, on top, the part that actually looks like a conversation.

**The model is the same.
The corpus is the difference.**

Slide 27 of 63

The honest way to say the whole thing is this: the model is the same. The corpus is the difference. Nothing about the underlying AI is smarter than what anyone else is using. What's different is what it's allowed to know before it starts.

AGENT DIRECTORY STRUCTURE

```
context/agents/  
  
profile/ Profile Knowledge Agent  
  agent.json { "name": "Profile Knowledge Agent" }  
  career.md Role, background, professional trajectory  
  working-style.md Decision defaults, what energises and drains  
  
organisation/ Organisation Knowledge Agent  
  agent.json { "name": "Organisation Knowledge Agent" }  
  people.md Team structure, reporting lines, key relationships  
  
decisions/ Decisions Knowledge Agent  
  agent.json { "name": "Decisions Knowledge Agent" }  
  made.md Settled calls — the why, not just the what  
  open.md Live tradeoffs, not yet resolved
```

Slide 28 of 63

That corpus is structured into Knowledge Agents — directories of plain markdown, one per domain: career background, working style, who I work with, decisions already made, those that haven't been.

On demand

Chat interface

- 01 System prompt includes agent index + required-reads rules
- 02 Model reads task, identifies which agents apply
- 03 Model calls `read_file` for each relevant file
- 04 Content enters conversation context mid-loop

Model-directed • correct for general questions where relevance varies

Pre-loaded

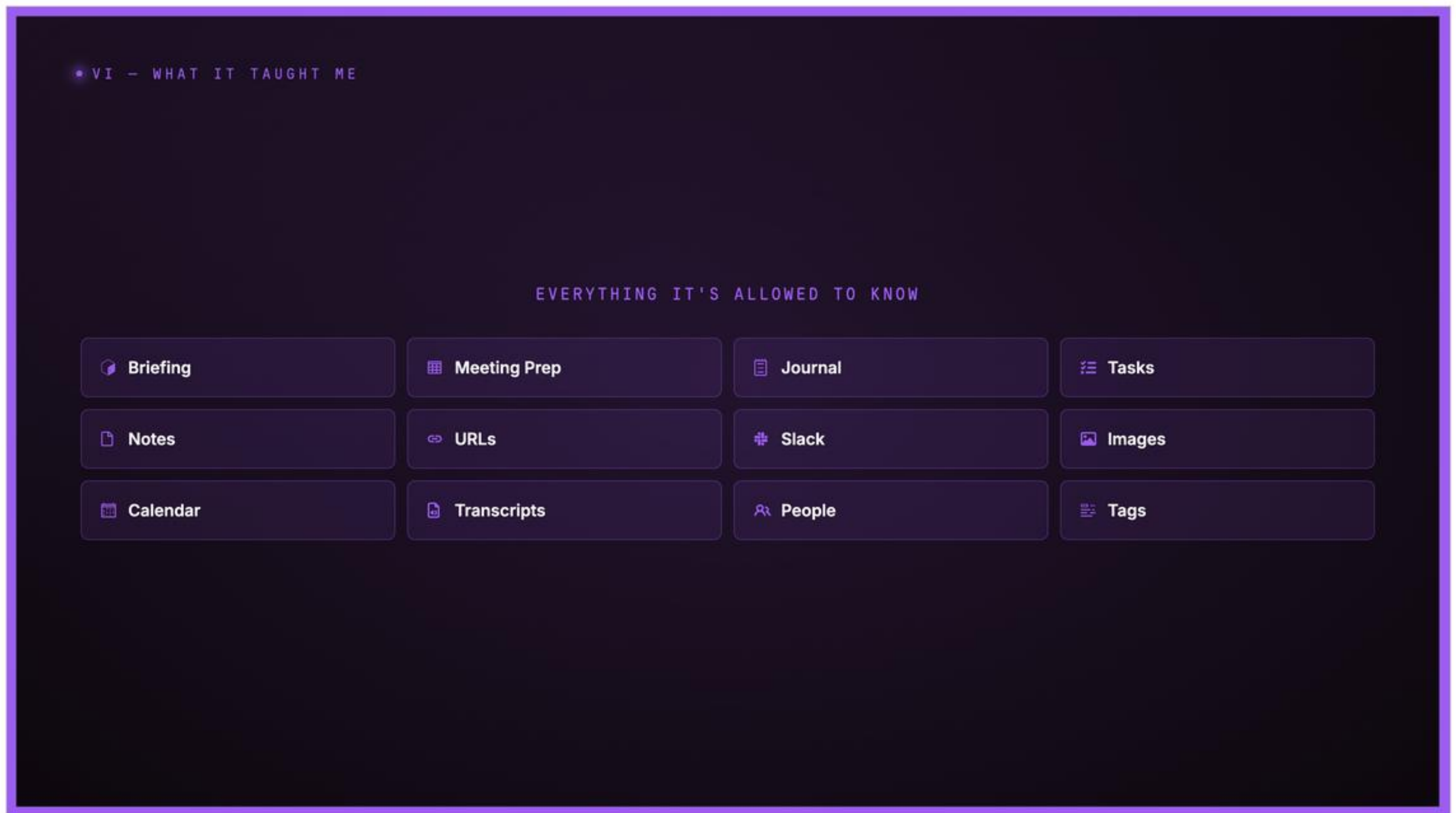
Brief • meeting prep

- 01 Server identifies which agents are always needed for this task
- 02 Agent files are read and injected into the system prompt
- 03 Model receives full agent context on the first turn
- 04 No `read_file` round-trip required

Server-directed • correct when the same agents are always needed

Slide 29 of 63

Two ways they get used. For an open conversation, the model decides for itself which agents to pull in, on demand. For a structured task — a brief, meeting prep — the relevant agents are loaded in ahead of time, so there's no lag while it works out what it needs. It even helps maintain itself — it can go back through transcripts and notes and propose updates to its own agents.



Slide 30 of 63

Where does all of that actually come from. Calendar, task history, notes I've written, a synced LinkedIn feed for anyone I work with — but the single richest source, by a distance, is meeting transcripts — roughly three-quarters of everything COSTA actually processes, by volume, comes from transcripts alone. That's context which used to be the most immediately forgotten thing in my entire week. Transcripts turn out to be unusually high-fidelity too, because a meeting is often the one place a discussion is actually decision-focused, rather than exploratory.

That Slack thread I mentioned gets solved directly too — there's a Slack integration that lets me forward a message or a whole thread straight in, the moment it matters.

PEOPLE

MC

Maya Chen
Head of Product Design · Meridian
Direct report · 8 mutual connections

SYNTHESIS

Maya joined 18 months ago and leads the design system work. Working relationship is strong — 1-1s are consistent and she's candid about blockers.

1-1 HISTORY

Mar 18	Component audit · team capacity
Mar 04	Q2 OKR planning · roadmap
Feb 18	Team structure · hiring plan

Slide 31 of 63

Person records work the same way, except they compound. They're not flat profiles — they're a graph, connective tissue between everything I've ever discussed with someone, every one-to-one, every decision, growing richer the longer I use it. Day one of this thing was fine. Month seven is categorically different, because patterns show up across months of accumulated context that no single conversation would ever reveal on its own.

So that's the architecture. What I actually want to spend a minute on, for a room like this one, is a couple of the design decisions sitting on top of it.



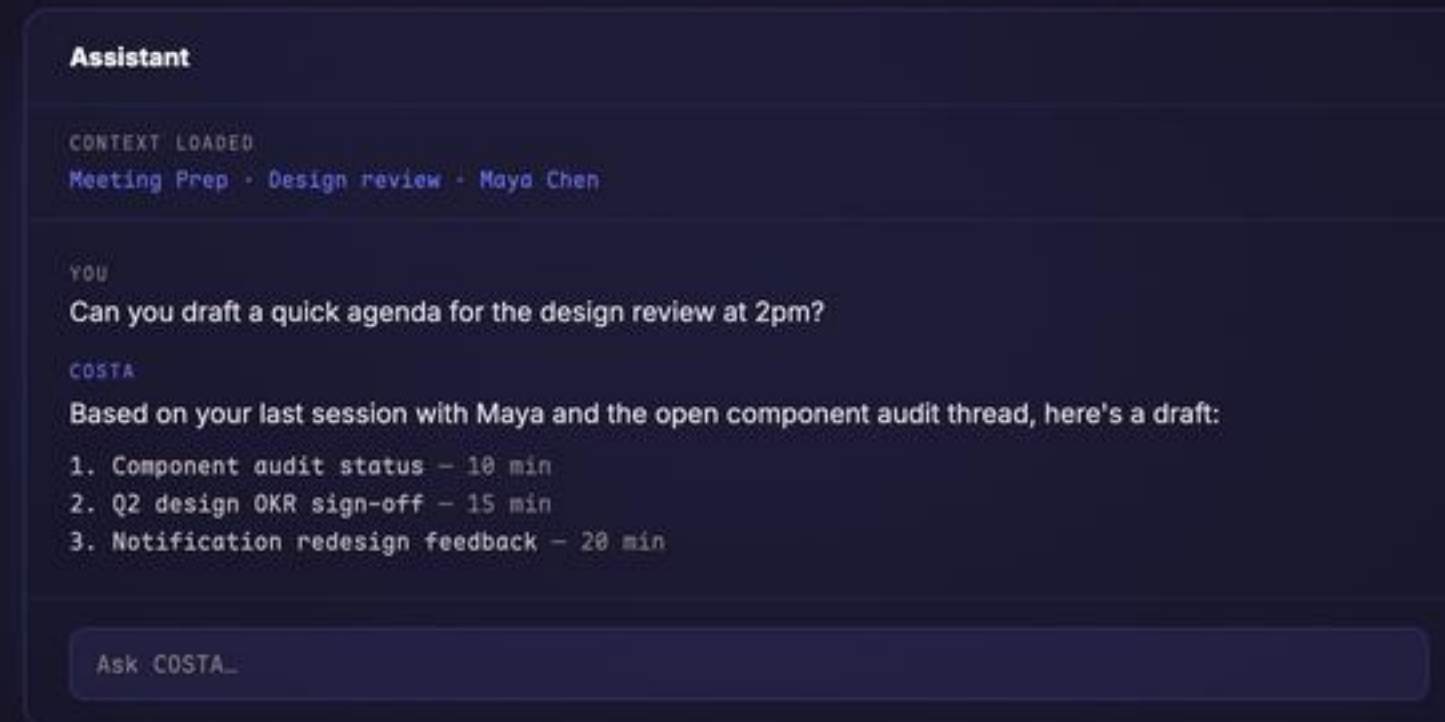
Slide 32 of 63

That podcast I mentioned at the start: the brief gets reformatted for audio, piped through a voice model, a jingle stuck on the front of it, and out comes a two or three minute episode, in a synthetic version of my own voice, telling me what the day actually looks like. That's a genuine design decision, not a party trick. Different formats fight information fatigue in a way that reading the same text twice never does, and a podcast reaches me in places a document can't.

My own voice.
Not another boss.

Slide 33 of 63

But the reason it's my voice, and not some generic assistant voice reading it back to me, is the actual point. Anyone else's voice telling me what to do today is just another boss. I've got enough of those already. My own voice saying it back to me is something closer to agency.



Slide 34 of 63

The chat interface does more than it looks like it does, too. It'll talk through a genuinely difficult decision with me — surface context I'd have sworn I'd never mentioned, push me to think from first principles instead of from habit. Some days that's just me ranting about a frustrating meeting at the end of the day, and it'll help put that meeting back into proper context — expanding the frame from one annoying hour out to the bigger, more exciting work it was actually a small part of.

Derived Actions.

Slide 35 of 63

And the output of one of those conversations can turn into an actual document. I've taken to calling those Derived Actions. A Derived Action comes out the other end of a real conversation I was actually part of, grounded in recall the whole way through, not a guess made alone in the dark.

Plain markdown.
My machine. No lock-in.

Slide 36 of 63

The lesson underneath all of it is that none of this works if you treat recall as a feature you add later. It has to be the architecture you commit to before you write anything else. And everything COSTA holds is stored as plain markdown, on my own machine — no vendor lock-in, nothing I couldn't read or edit myself in a text editor if the company behind whichever model I'm using disappeared tomorrow.

**It's my context.
I wanted to own it.**

Slide 37 of 63

It's my context. I wanted to actually own it.

**Every conversation
started cold.**

Slide 38 of 63

Building COSTA made me better at my job almost immediately, but it also created a gap between what I could do in COSTA, versus every other AI tool I was using.

Every new conversation started cold. Same base context, re-entered over and over, the model meeting me for the first time, every single time.

Averaged. Flattened.

Slide 39 of 63

And in the gap where my own knowledge should have sat, it reached for whatever it had absorbed off the open web instead — conference talks, Medium essays of variable quality, things that were nominally about design and weren't, the insane ramblings of Mark Zuckerberg, and that weird guy from Basecamp. Averaged. Flattened. Nobody's actual view, least of all mine.

coach

CONTENT

get_dlm_writing

get_dlm_links

get_design_worldview

get_stated_positions

get_taste

get_frameworks

INSTRUCTION

get_writing_style

get_voice_anti_patterns

get_ai_instructions

CONTEXT

get_career

get_working_style

get_personal_context

Slide 40 of 63

So I pulled the same pattern out of COSTA and made it portable. It's called coach — a standalone MCP server, queryable by any AI tool, not just mine. My own corpus, sitting in the gap where generic training data would otherwise go. Content tools holding what I've actually written and linked to, my design worldview, my taste, positions I've actually stated and stood behind. Instruction tools telling a model how I actually write, and just as usefully, what not to write.

mrfrisby.com/ask

Slide 41 of 63

There's a public version too — mrfrisby.com/ask, a toy more than a demo: a simpler way to poke at the same MCP server, if you want to see for yourself.

I wanted badly to believe it was working, but I was skeptical, so I built a proper blind evaluation — a generic model against the same agent, connected to coach, both answering the same design leadership questions, scored by a third model with no idea which side had the context.

12 / 13

prompts won by context

Slide 42 of 63

Against a blank, no-instruction model — a lower bar than a good prompt, but a real result on its own terms — context won twelve of thirteen.

"That isn't a framework I use or have written about — I don't want to walk you through something I'd be inventing on the spot and attributing to myself."

Slide 43 of 63

And on the one prompt that mattered most, baiting it to explain a framework I've genuinely never written about, the generic model invented one. Confidently, fluently, completely made up. Mine said: "That isn't a framework I use or have written about — I don't want to walk you through something I'd be inventing on the spot and attributing to myself." Which, incidentally, is also exactly how I talk, tedious lack of punctuation included.

• VII — A PORTABLE LAYER

• FULL REPORT

Can Personal Context Servers Improve AI Outputs?

[Link at the end](#)



Slide 44 of 63

The full methodology and the limitations — the blank baseline, the LLM judge, the sample size — are written up properly here, if anyone wants to check my working rather than take the headline number on faith.

**Recall failing gracefully
looks like honesty.**

Slide 45 of 63

So now I could state the theory I opened with confidently: Inference without recall doesn't fail quietly — it fails by making something up and sounding sure about it. Recall failing gracefully just looks like honesty.

What stops happening.

Slide 46 of 63

So what actually changes, day to day, once this is real instead of theoretical.

The first thing is what stops happening. I stop having to desperately fumble for old notes, or through old conversations. I'm going into meetings with those things surfaced and part of the basis for the conversation to come. I feel less harried, and better prepared.

**It's a relief to know my bad
memory isn't making me look
like a bad person.**

Slide 47 of 63

I used to worry a lot about forgetting commitments, or decisions, appearing thoughtless or careless. It's a relief to know that my bad memory isn't making me look like a bad person.

206h

estimated saved this quarter

Slide 48 of 63

COSTA tracks this too, because of course it does. This quarter, by its own estimate — based on time it can see it's saved across briefings, prep and drafts, week on week — it's saved me around two hundred hours.

The end result of that — being better prepared, and feeling more competent — is that I make better decisions, and help better decisions get made. Be those organisation, strategy or design decisions.

What energises you.
What drains you.

Slide 49 of 63

There's a quieter version of this too. COSTA also keeps track of how I feel at work — the things that give me energy, the things that take it away. I log a quick note most days on how it actually went, almost without thinking about it, and over months COSTA can show me the shape of that — when I actually do my best work. None of that's visible one day at a time. It only shows up once something's been keeping score for long enough.

So much of what we do everyday is about acting on carried context, that it feels wild that we made it to 2026 without having a way to carry it properly.

COSTA doesn't run itself.

Slide 50 of 63

That's the upside. Here's the part that's less magical.

COSTA doesn't run itself. It needs maintaining — it doesn't just stay still, it quickly gets worse, because context decays.

Proposed Actions

Slide 51 of 63

I've been telling you that recall beats inference, and I stand by that, but I also built an inference-based tool into COSTA, because I wanted to see if sufficient context to recall could aid inference. I recently added a feature called Proposed Actions — it effectively takes a look through my task list and decides which ones it can draft something for — an email, a document outline. Where the rest of COSTA is oriented around recall, Proposed Actions is exactly the inference work I've been arguing against outsourcing to AI.

**Right some of the time.
Wrong most of the time.**

Slide 52 of 63

And the answer is, kind of. It gets some things right some of the time, but it gets most things wrong, most of the time. That's fine — it isn't mission critical and COSTA is never acting on its own, only ever via me — but it's a good lesson. Even with incredibly deep and narrow context to recall, these tools are still not good at inference. This is the same bad UI drawing problem we talked about earlier.

There is also the limit of what gets collected, and how often it gets updated. Decisions change, perspectives evolve. Brains are good at that constant reshuffling of information, but systems have to be guided towards those same things, and that requires thoughtful maintenance.

Bounded. Unbounded.

Slide 53 of 63

This serves as a good reminder of what these tools are, and what they are not. Recall makes decision-making better because it is supported by and bounded by context, whereas inference is as unbounded as the LLM context from which it infers.

AI-native means expanded reach. Not new tools.

Slide 54 of 63

So — why does any of this matter beyond one man's slightly obsessive relationship with his own task list.

For design specifically, being AI-native was never really about new tools. It's about expanded reach — more options at the exploration stage, higher fidelity, faster, with the time that saves spent on the judgment work AI still can't do.

Taste. Judgment.

Slide 55 of 63

And that's the thing that links COSTA and our work as designers: once recall handles the information problem — once the research, the context, the precedent stops being the bottleneck — what's left to actually differentiate a designer is taste and judgment. The thing I made a joke about OpenAI needing, earlier, and meant every word of. Leave the part of the work you actually love doing alone, and let AI optimise everything else around it.

"This changes everything" vs. a smaller, more useful idea

Slide 56 of 63

None of what I've described today is the "this changes everything" story. It's smaller than that, and more useful for being smaller — a genuinely new capability, undersold because it isn't shiny and exciting in the same way a crappy prototype or a crappy poster seemingly are.

1. Collect your raw materials.

Slide 57 of 63

So, three things to take away, if you'll indulge me.

One. Collect your raw materials. Recall only works if there's something to recall from — your notes, your blog posts, your Slack rants. Not the polished output. The raw stuff, especially the stuff that was never meant to go anywhere. That's the thing no AI tool can supply on your behalf, because it isn't on the internet, and it was never meant to be. It's also yours, and you should treat it as the important and unique asset it is.

2. Consider building your own MCP server.

Slide 58 of 63

Two. Consider building your own MCP server. I know how that sounds. But it's simple enough that a tool which can't design for shit managed to build one for me. A handful of tools. Your own corpus, instead of the internet's, and then you can plug into Claude, into ChatGPT, even into Figma Make if you really want to test the limits of your patience.

3. Don't let design get recast as asset generation.

Slide 59 of 63

Three. Don't let “design” get recast as asset generation. We spent a generation telling everyone who would listen that design is not just about the output — that it's a process. If we're going to resist that being undone, we need to engage these emerging tools in that process, not just in the asset generation part.

**Design is what happens
between your ears.**

Slide 60 of 63

So — design is what happens between your ears. Not between calls to someone else's endpoint. AI can't do the thinking for you. What it can do properly, for the first time — is stop you having to guess at things you already had the information to actually know.

**Recall didn't need to be
invented. It needed to be
taken seriously.**

Slide 61 of 63

Recall didn't need to be invented, just optimised. All of that context has been withering in dark corners of our design processes all along, in the notebooks and the Slack threads and the confluence pages nobody read. It needed to be taken seriously, that's all. And whilst the world was busy generating badly designed posters, it may have accidentally plugged one of the biggest holes in knowledge work.

Thank you.

Slide 62 of 63

Thank you.



SLIDES & LINKS

**mrfrisby.com/
design-ai-summit**

The deck, the eval write-up, /ask, coach, and the newsletter — all at that one link.