The Agent Surface

Designing Products for Users Who Never Open the App

Alex Thomas, CTO, Applied Visions

Find us at the Applied Visions table, outside Ballroom A.

Connect on LinkedIn

linkedin.com/in/alexander-jt

applied visions

The thesis

People Will Bring Their Own AI

to your product, whether you plan for it or not.

The thesis

People Will Bring Their Own AI

to your product, whether you plan for it or not.

This talk

Getting the job done through your product

Productivity apps: your SaaS platform, your business software, your idea

The thesis

People Will Bring Their Own AI

to your product, whether you plan for it or not.

This talk

Getting the job done through your product

Productivity apps: your SaaS platform, your business software, your idea

Not this talk

Ads, upsells and onboarding tours

The parts of your screen an agent never sees

What's happening

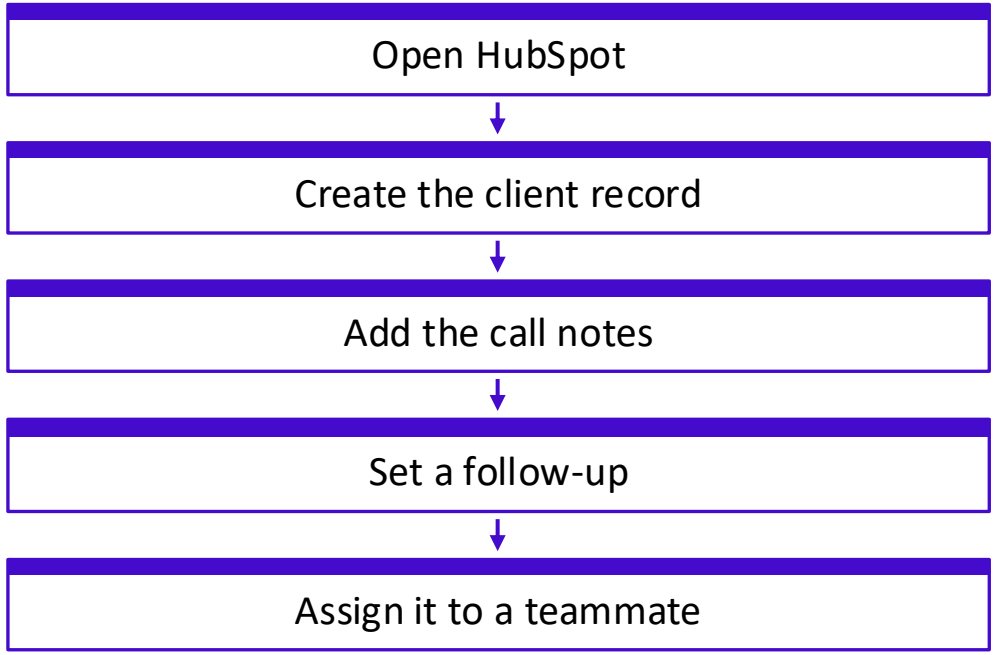
Your Next Power User Isn't a Person

Maybe you see a problem. Maybe an opportunity. So do I.

A tool I use every day and never see

HubSpot

I don't do this.

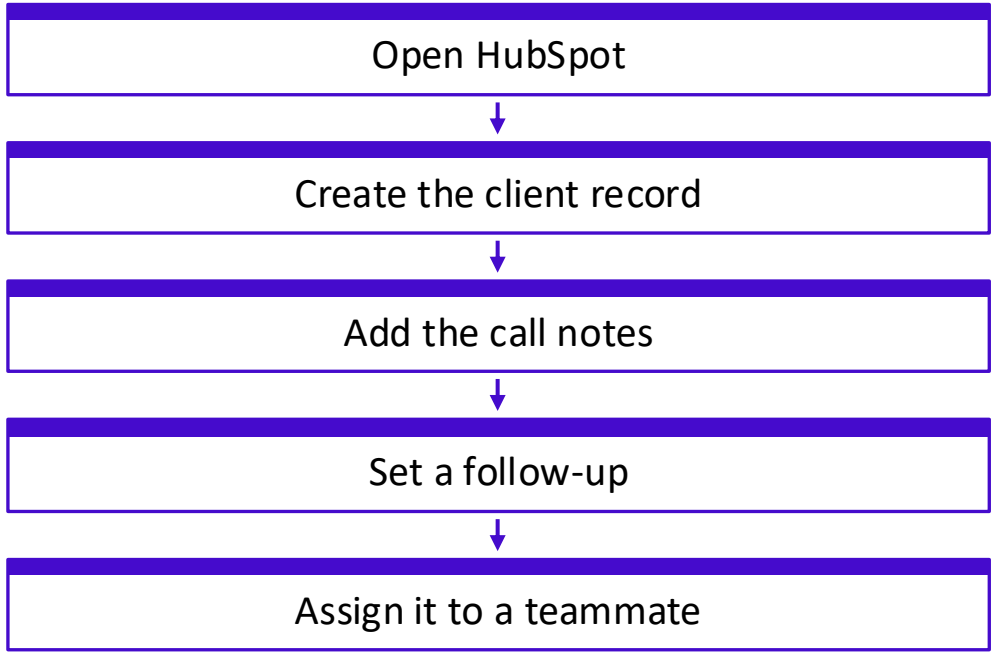


And so on, for every call.

A tool I use every day and never see

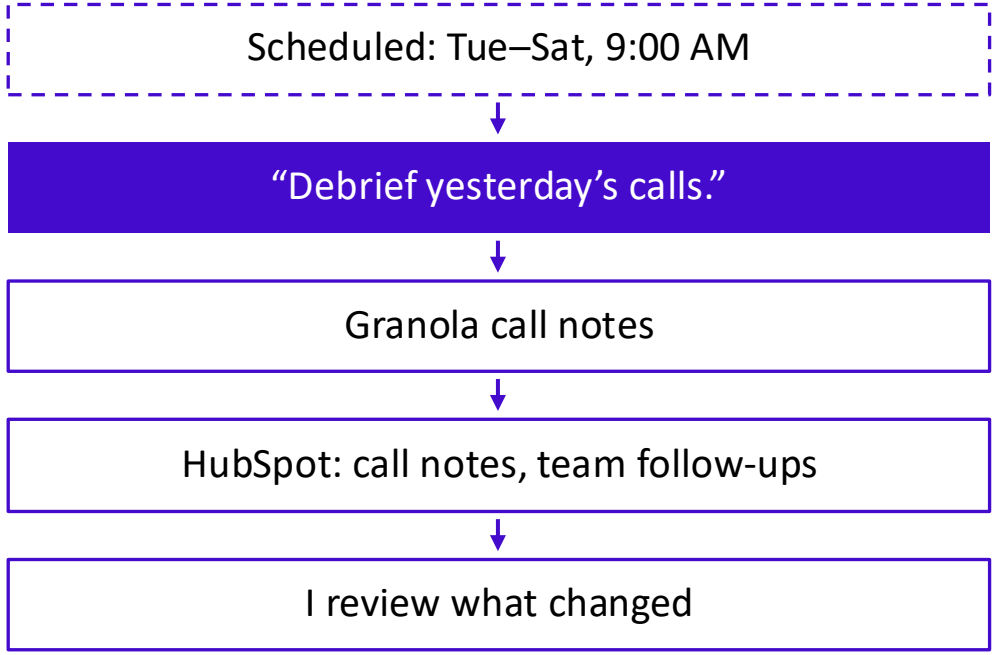
HubSpot

I don't do this.



And so on, for every call.

I do do this.



Each morning, for the previous workday's calls.

What's happening

Built In, or Brought In?

Built-in agent

Lives inside one product

Rovo in Jira, Breeze in HubSpot. Fine at that product. Blind to the rest of my day.

What's happening

Built In, or Brought In?

Built-in agent

Lives inside one product

Rovo in Jira, Breeze in HubSpot. Fine at that product. Blind to the rest of my day.

The agent I bring

Works across all my tools

Granola, HubSpot, Outlook, Teams. One request, every tool.

What's happening

Built In, or Brought In?

Built-in agent

Lives inside one product

Rovo in Jira, Breeze in HubSpot. Fine at that product. Blind to the rest of my day.

The agent I bring

Works across all my tools

Granola, HubSpot, Outlook, Teams. One request, every tool.

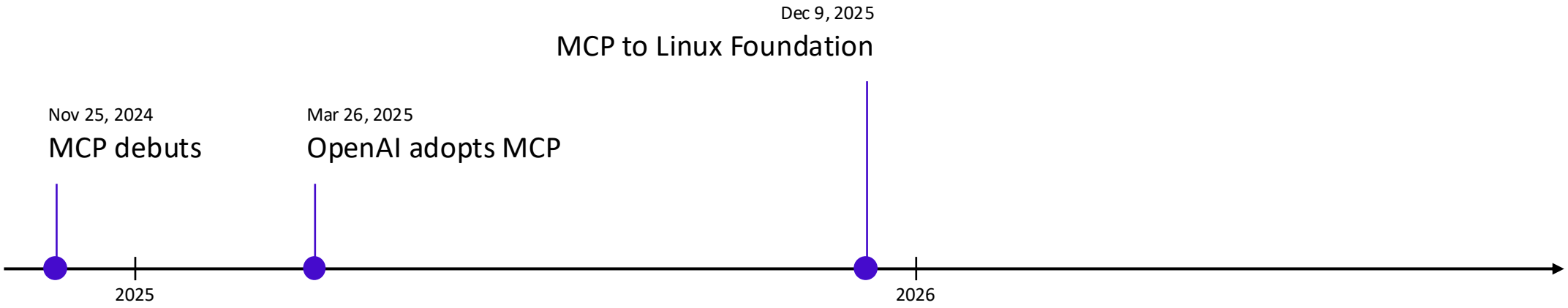
1.2B people use ChatGPT every week.

Most of your users already have an agent in hand.

Source: OpenAI at DevDay, Sep 29, 2026 (reported by The Verge).

What's happening

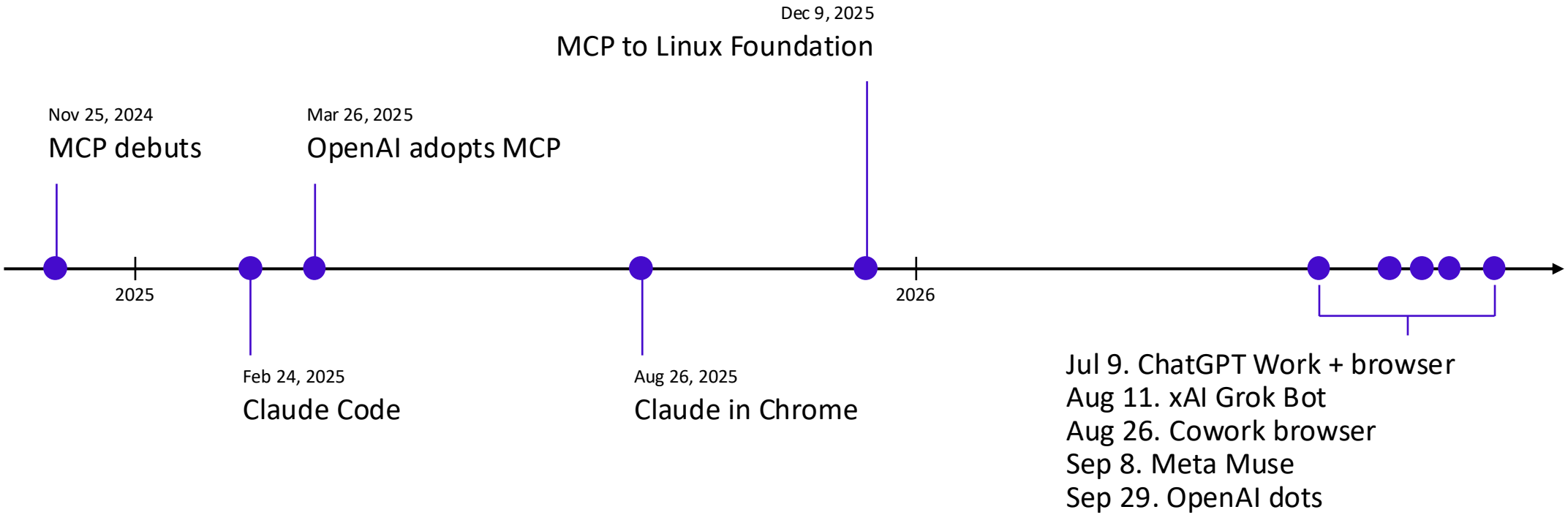
First You Opted In.



Sources: Anthropic; TechCrunch, Mar 26, 2025; Linux Foundation; OpenAI; xAI; Meta; HUMAN Security, State of Agentic Traffic, Sep 2026.

What's happening

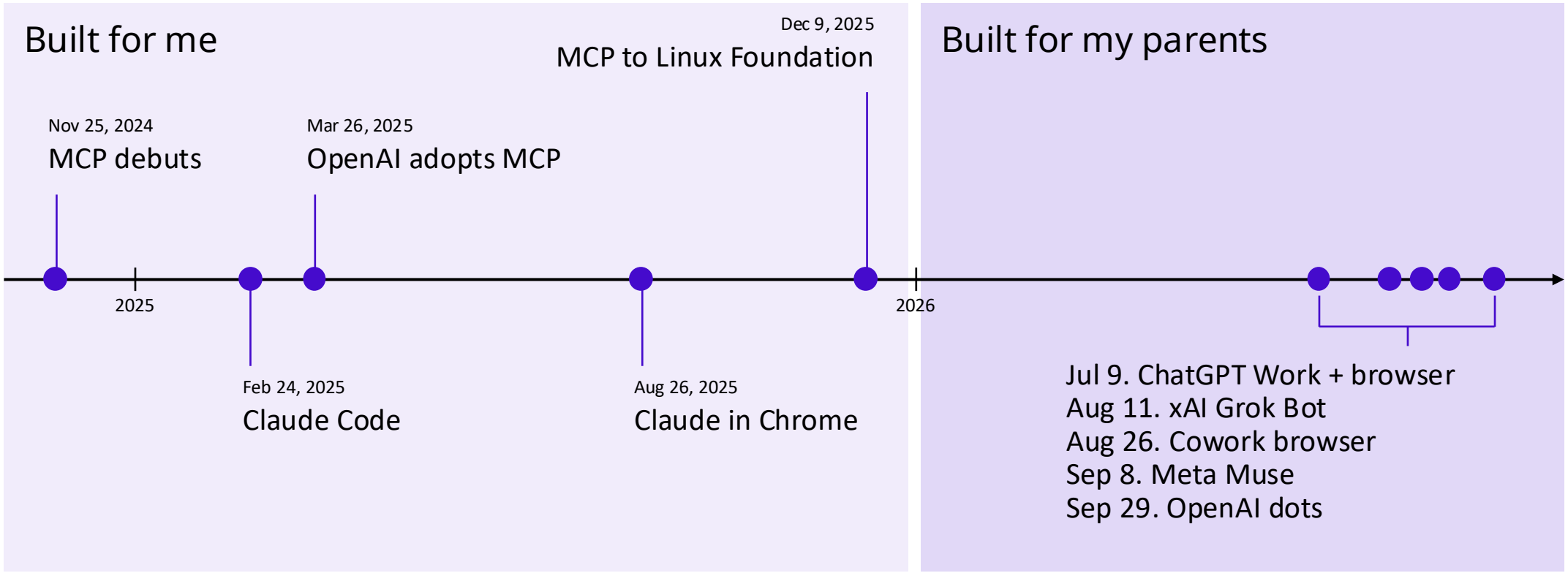
First You Opted In. Now They Just Show Up.



Sources: Anthropic; TechCrunch, Mar 26, 2025; Linux Foundation; OpenAI; xAI; Meta; HUMAN Security, State of Agentic Traffic, Sep 2026.

What's happening

First You Opted In. Now They Just Show Up.



Sources: Anthropic; TechCrunch, Mar 26, 2025; Linux Foundation; OpenAI; xAI; Meta; HUMAN Security, State of Agentic Traffic, Sep 2026.

What's happening

You Probably Can't See Them Coming

3×

Agentic browser traffic, August to September 2026

What's happening

You Probably Can't See Them Coming

3×

Agentic browser traffic, August to September 2026

40%

of that traffic came from one consumer agent: Meta's Muse

What's happening

You Probably Can't See Them Coming

3×

Agentic browser traffic, August to September 2026

40%

of that traffic came from one consumer agent: Meta's Muse

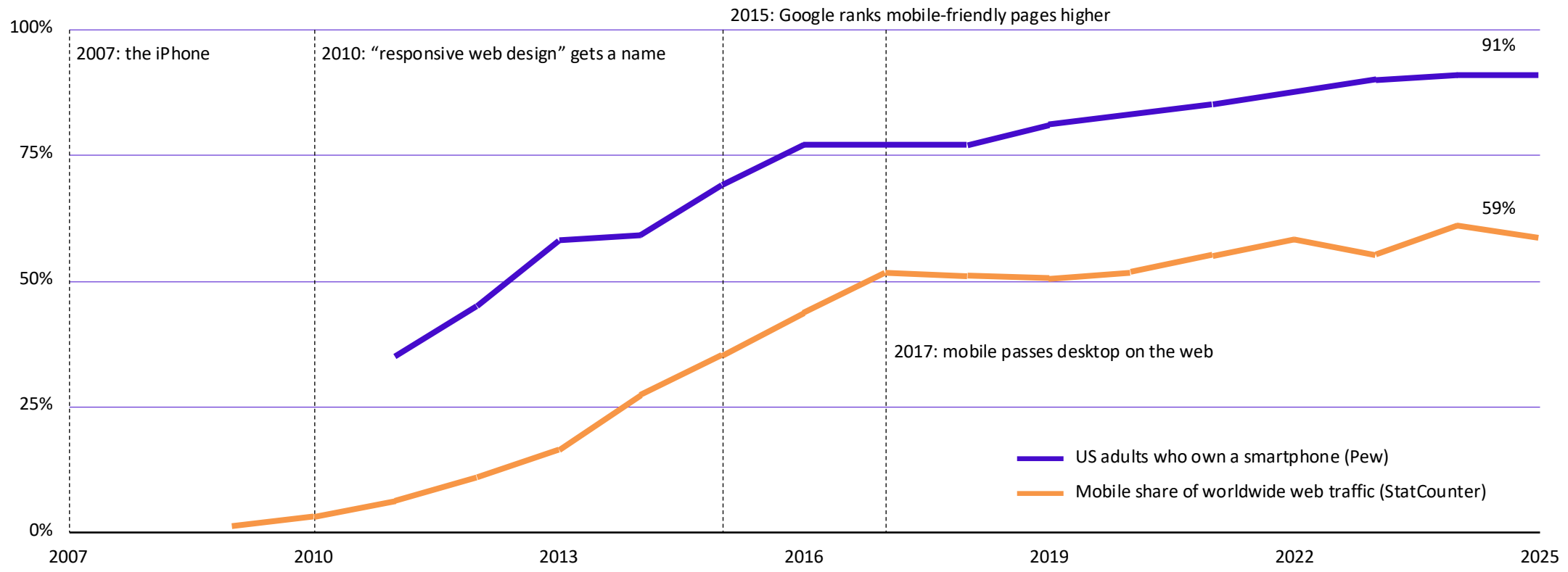
>50%

of companies studied saw Muse traffic within two weeks, and it didn't identify as an agent

Sources: HUMAN Security, State of Agentic Traffic, September 2026 (Oct 5, 2026): +209.7% month over month; Muse 40.3%. Cequence Security, Oct 7, 2026.

What's happening

Remember Mobile-First?



Sources: Pew Research Center, Mobile Fact Sheet (Nov 20, 2025). StatCounter Global Stats, desktop vs mobile vs tablet, yearly. Marcotte, A List Apart, May 25, 2010. Google, Apr 21, 2015.

What's happening

Mobile Had a Buy-In.

Mobile, 2007

$499

and a two-year contract

Four years later, 35% of US adults owned a smartphone

Sources: Apple iPhone launch pricing, June 2007. Pew Research Center, Mobile Fact Sheet (Nov 20, 2025).

What's happening

Mobile Had a Buy-In. Agents Don't.

Mobile, 2007

$499

and a two-year contract

Four years later, 35% of US adults owned a smartphone

Agents, 2026

Free

inside apps people already have

Muse: 40% of the agent traffic HUMAN sees, within weeks of launch

Not mainstream yet. But no buy-in means everyone gets to try it.

Sources: Apple iPhone launch pricing, June 2007. Pew Research Center, Mobile Fact Sheet (Nov 20, 2025). HUMAN Security, State of Agentic Traffic, Sep 2026.

The hard way

The Screen Is Already There. Why Do Anything?

Because through the screen is the hard way.

The hard way

Same Tasks, Fewer Steps

Driving the screen

16

steps per Office task

Word, Excel, PowerPoint; same model (o1)

The hard way

Same Tasks, Fewer Steps

Driving the screen

16

steps per Office task

Word, Excel, PowerPoint; same model (o1)

Calling APIs

6.6

steps per Office task

And more of the tasks got done

The hard way

Same Tasks, Fewer Steps

Driving the screen

16

steps per Office task

Word, Excel, PowerPoint; same model (o1)

Calling APIs

6.6

steps per Office task

And more of the tasks got done

On the web, calling APIs roughly doubled success: 15% to 29%.

Sources: Zhang et al. (Microsoft), arXiv 2503.11069, Table 3 (o1; success 16.3% → 24.5%). Song et al. (CMU), arXiv 2410.16464 (WebArena).

The hard way

Sessions Break.

Borrow the session

Fragile

Sessions expire
MFA and CAPTCHAs interrupt
Layouts move
Clicks fail silently

The hard way

Sessions Break. Passwords Leak.

Borrow the session

Fragile

Sessions expire
MFA and CAPTCHAs interrupt
Layouts move
Clicks fail silently

Hand over the password

Dangerous

The agent is you: no scope, no expiry, no audit
Revoking it means changing your password
MFA still gets in the way

The hard way

Sessions Break. Passwords Leak.

Borrow the session

Fragile

Sessions expire
MFA and CAPTCHAs interrupt
Layouts move
Clicks fail silently

Hand over the password

Dangerous

The agent is you: no scope, no expiry, no audit
Revoking it means changing your password
MFA still gets in the way

8×

AI requests to login pages,
January to June 2026

Source: DataDome, State of Bot & Agent Security 2026 (11.9M → 99.7M, Jan to Jun 2026).

The hard way

Atlassian Opened the Door

1M+

monthly users of Atlassian's MCP server and CLI, more than doubling in one quarter

2×

faster ARR growth for customers who connect through MCP

98%

of MCP users still open Jira in the same month

People with agents will prefer products with a good agent surface.

Source: Atlassian, Q4 FY26 letter to shareholders, Aug 6, 2026. Company-reported; customers who adopt MCP self-select.

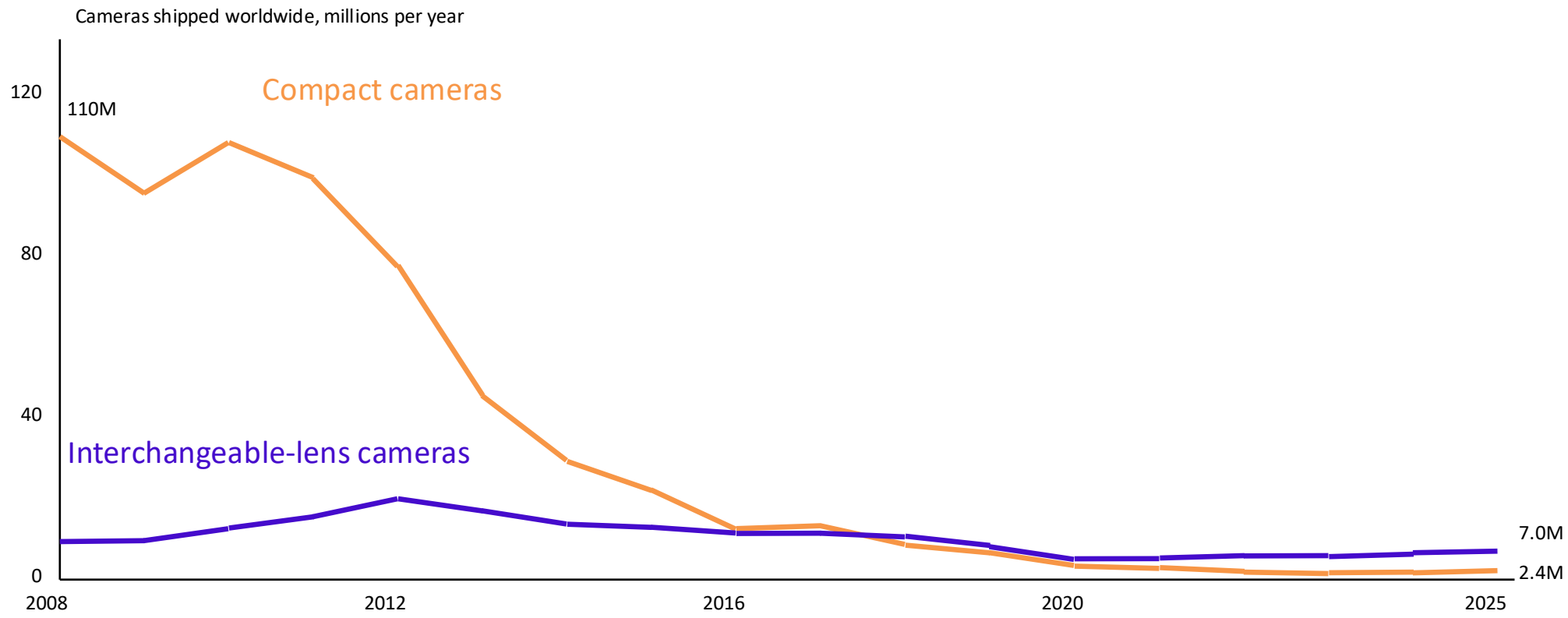
Why your software survives

Won't the Agents Just Replace You?

Why your software is still here.

Why your software survives

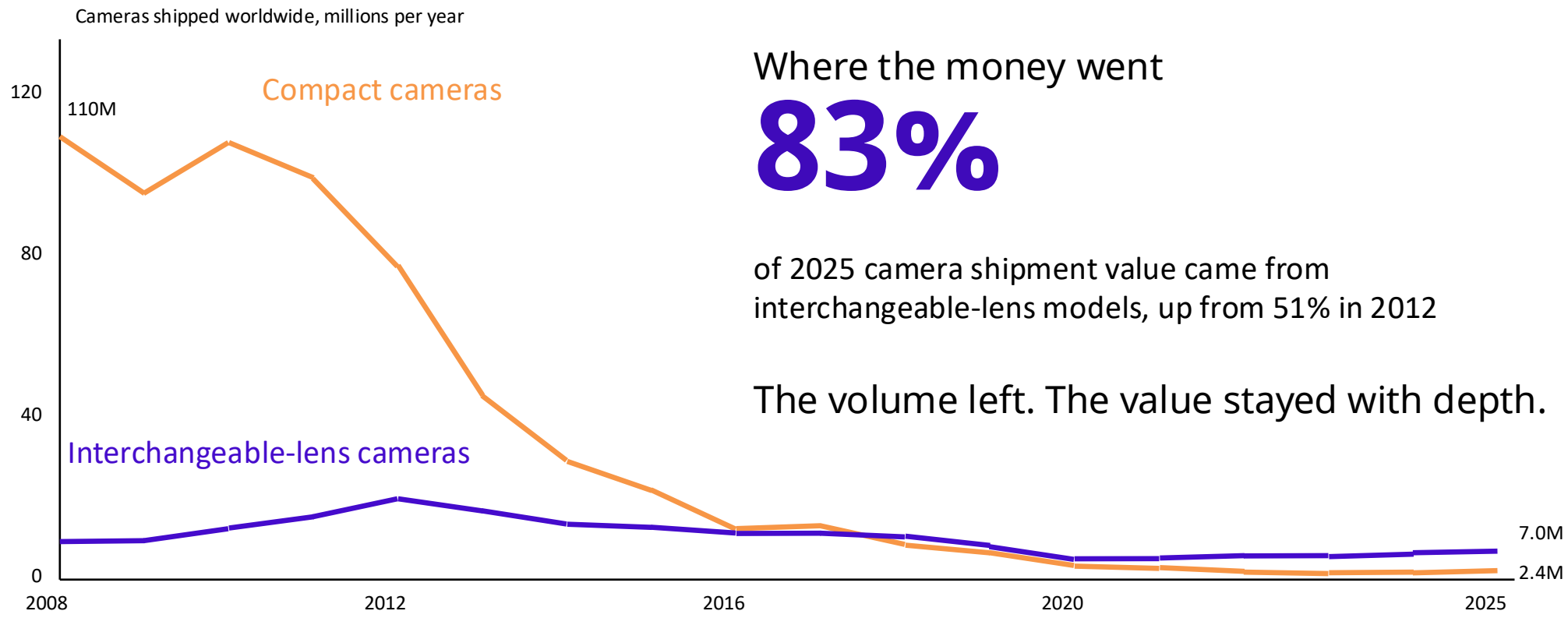
The Phone Ate the Point-and-Shoot



Source: CIPA worldwide shipments by member companies, 2008–2025 (annual Jan–Dec reports).

Why your software survives

The Phone Ate the Point-and-Shoot



Where the money went

83%

of 2025 camera shipment value came from interchangeable-lens models, up from 51% in 2012

The volume left. The value stayed with depth.

Source: CIPA worldwide shipments by member companies, 2008–2025; share of shipment value (yen), 2012 vs 2025.

Why your software survives

Good Enough for a Consumer.

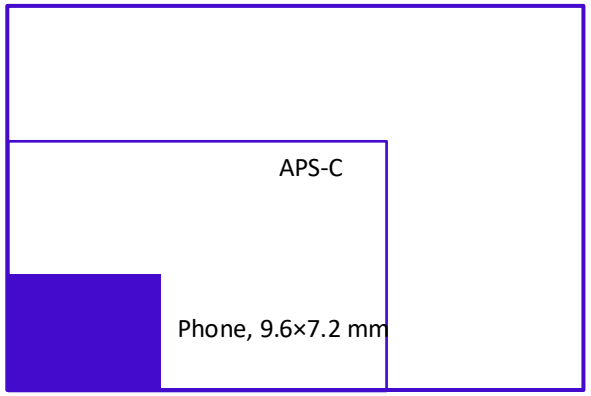
The phone: good enough for most people.



Phone, 9.6×7.2 mm

Why your software survives

Good Enough for a Consumer. Not for a Pro.



Full frame, 36×24 mm

12×

A full-frame sensor has 12× the area of a flagship phone's. More light, and lenses you swap for the job.

The phone: good enough for most people.

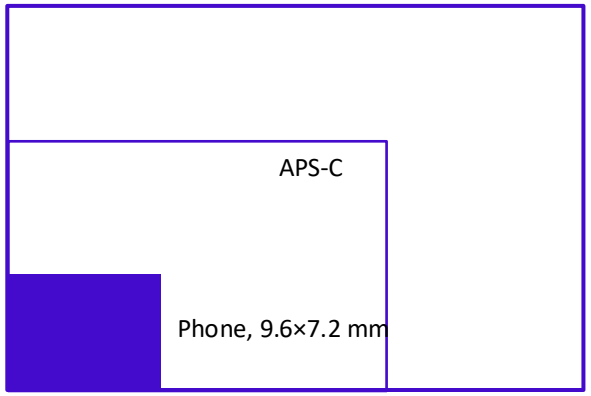
The mirrorless camera: what a pro needs.

Drawn to scale. Full frame 36×24 mm (864 mm²); APS-C 23.6×15.6 mm; 1/1.33-inch phone sensor 9.6×7.2 mm (69 mm², Galaxy S20 Ultra). Source: Wikipedia, Image sensor format.

Why your software survives

Good Enough for a Consumer. Not for a Pro.

Not actual size.



12×

A full-frame sensor has 12× the area of a flagship phone's. More light, and lenses you swap for the job.

The phone: good enough for most people.

The mirrorless camera: what a pro needs.

Drawn to scale. Full frame 36×24 mm (864 mm²); APS-C 23.6×15.6 mm; 1/1.33-inch phone sensor 9.6×7.2 mm (69 mm², Galaxy S20 Ultra). Source: Wikipedia, Image sensor format.

Why your software survives

Good at the Route. Bad at the Record.

Agents are good at the route

Turning intent into steps

Reaching any tool

Gluing systems together

Why your software survives

Good at the Route. Bad at the Record.

Agents are good at the route

Turning intent into steps

Reaching any tool

Gluing systems together

They're bad at the record

Remembering over time

Context rot, and a fresh start every session

Following rigid rules

Same task right 8 times running: under 25%

Giving the same answer twice

1,000 identical runs: 80 different answers

Sources: Chroma, “Context Rot,” Jul 2025. Yao et al., τ-bench, arXiv 2406.12045. Thinking Machines, “Defeating Nondeterminism in LLM Inference,” Sep 2025.

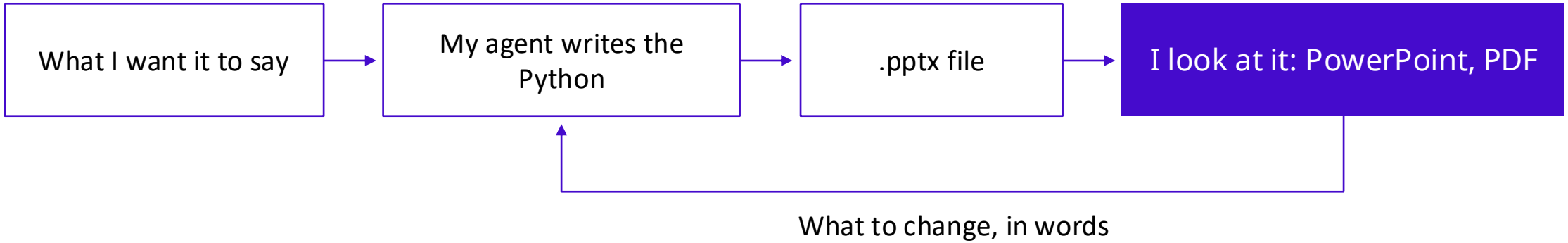
Expose your depth

Your API's Depth Is the Agent's Ceiling

So expose it.

Expose your depth

Iterated, Not Clicked



Plenty of iteration. None of it by clicking.

My last deck went through several versions in a week. The Blender video took six versions and a realism pass. Every round was a conversation, not a click.

Version counts from the lightning-deck build folder, Oct 8–12, 2026. To confirm before the talk: the pipeline as described.

A tool I'd never learn

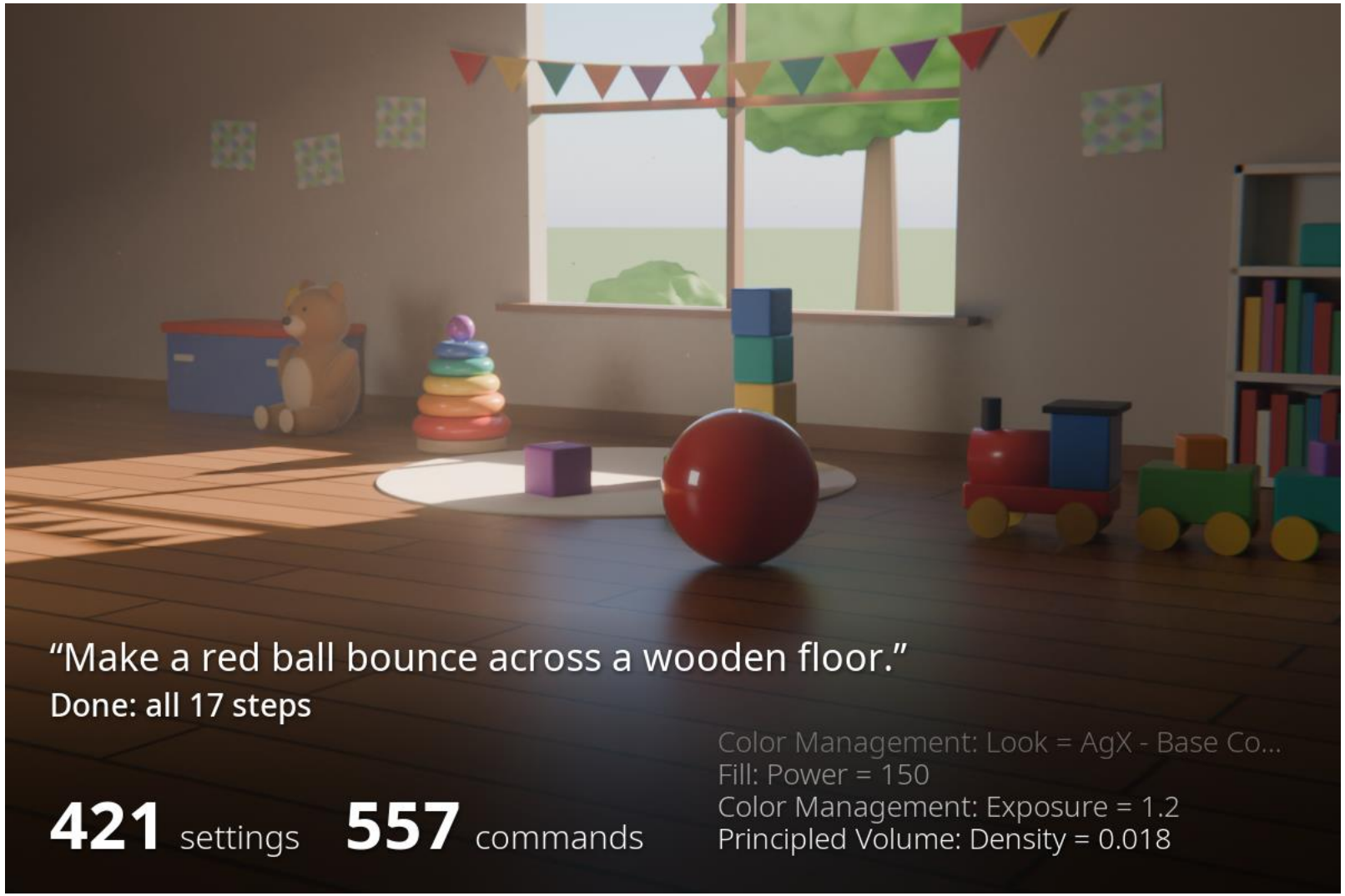
Blender

10,140

settings

2,491

commands



Hat tip: Jaclyn Konzelmann

Product Director, Google Labs

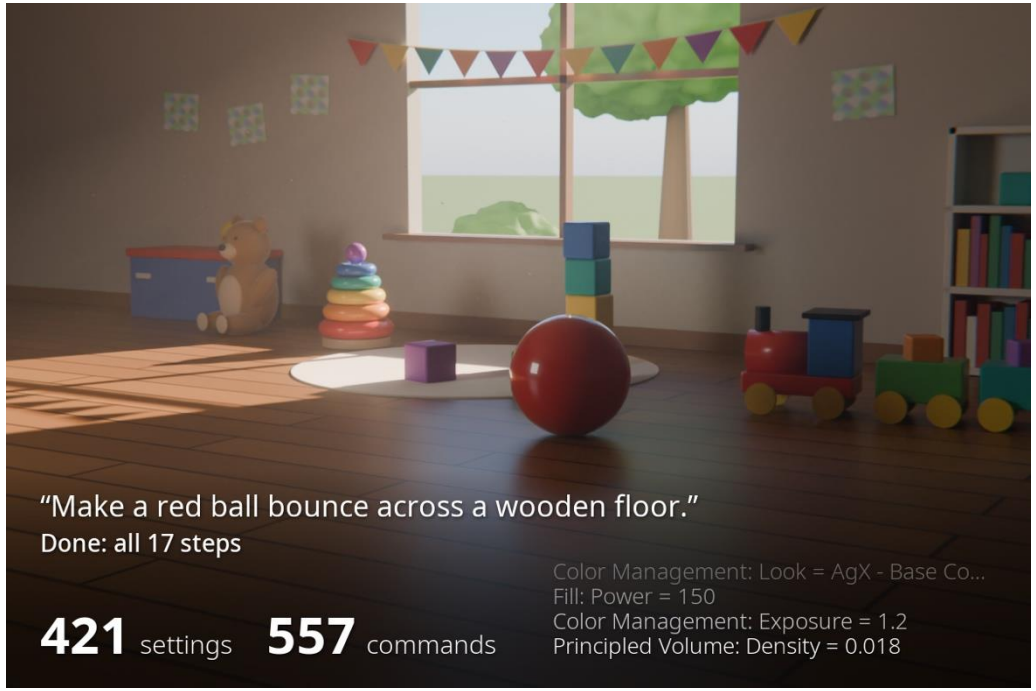
TAIC 2026

Every setting and command as it ran, built by Claude through Blender's Python API. Blender 5.2.2 LTS (bpy.ops, bpy.types), counted Oct 8, 2026.

Expose your depth

421 Settings Changed. I Touched None.

Where	Setting	Value
World: Sky	Sky Type	Multiple scattering
World: Sky	Aerosols	2.5
Light: Sun	Blackbody temperature	4,300 K
Haze	Volume anisotropy	0.75
Ball	Subsurface radius	(0.9, 0.2, 0.1)
Ball	Coat IOR	1.55
Ball	Thin-film thickness	380
Camera	F-stop	2.8
Camera	Aperture blades	6
Compositor	Glare type	Bloom
Compositor	Lens dispersion	0.015



Source: the agent's settings log for the lightning-talk build (421 settings, 557 commands, 17 steps; the sixth version of the video). Blender 5.2.2: 10,140 settings, 2,491 commands.

Expose your depth

Blender Has Python. What Do You Have?

Channel	Why it works	Where it hurts
Local MCP Runs on the user's computer	Reaches the machine: local files and desktop apps	Desktop app only. Hard to install unless packaged. IT can block it
Remote MCP Your server, added by URL	Works on web, desktop and phone. Nothing to install	Needs a public server and OAuth. At work, an admin adds it first
Connector Remote MCP, listed in a directory	One click to connect, and the agent suggests it when it fits	You pass directory review: OAuth, labeled tools, a test account
WebMCP Tools your web page registers	Nothing to install. Uses the session the user is already in	Only while the tab is open, in trial browsers. Hints, not security
A2A Agent-to-agent handoffs	Their agent hands a whole task to yours, like Atlassian's Rovo	You need your own agent on the other end. Mostly enterprise platforms

Sources: Claude and OpenAI connector, custom MCP and directory docs; Chrome for Developers, WebMCP; a2a-protocol.org; Atlassian Rovo A2A docs (2025–26).

Expose your depth

Other Ways In. Each Fits a Niche.

Channel	Why it works	Where it hurts
API Fits any web product, underneath	Any agent that can make a web request can use it. The rest wrap it	Built for developers: keys and docs, not tools an agent can pick from
CLI Fits developer and admin tools	Composable: filter the output before it floods the agent's context	Built for a person at a terminal. Runs with all of your permissions
Code library Fits file formats and app engines	The agent writes code against it: python-pptx for this deck, bpy for Blender	Needs somewhere to run code. Produces a file, not live data
Phone actions Fits mobile apps	Siri and Gemini act inside your app. Nothing extra to install	Only the phone's own assistant, and a separate build per OS
Commerce Fits online stores	Agents complete the purchase. You stay the seller of record	Only for buying, and two rival standards are still sorting it out

Sources: Apple App Intents and Android AppFunctions docs; Stripe ACP docs; Google, UCP announcement (Jan 11, 2026); Anthropic, “Code execution with MCP” (Nov 2025).

Make it witnessable

Control Isn't Enough

Give people something to witness.

Make it witnessable

It Saw the Shadow. I Judged the Shot.



Render → a screenshot to ChatGPT: “make this more photorealistic” → the critique goes back to my agent as intent → render again

Make it witnessable

The Loop Matters. The Taste Is Iffy.

Google Research and DeepMind, ICML 2024

SceneCraft

Take away the render-and-look loop and its score drops sharply.

Stanford, CVPR 2025

BlenderGym

The best models still struggle with edits human Blender users find easy.

ICML 2024

MLLM-as-a-Judge

Models compare two options like people do. They score and rank unlike people.

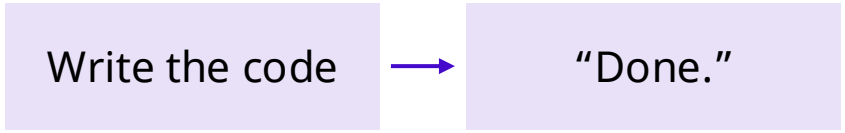
Let the agent compare. Let the person decide.

Sources: Hu et al., arXiv 2403.01248 (cumulative ablation). Gu et al., arXiv 2504.01786. Chen et al., arXiv 2402.04788.

Make it witnessable

Let the Agent See Its Own Work

Without a check



The page never rendered.

With a check



In your product: a render, preview or export the agent can call for itself.

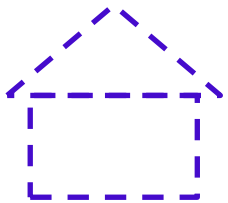
Source: Anthropic, code.claude.com/docs/en/best-practices.

Guardrails

You knew we'd make it here.

Guardrails

Your Users Still Do the Walkthrough



Their agent builds it

Blender: 421 settings

This deck: every slide

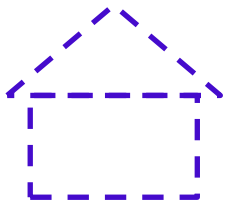
HubSpot: every record

Your job

Give it a door

Guardrails

Your Users Still Do the Walkthrough



Their agent builds it

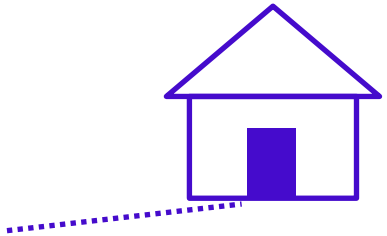
Blender: 421 settings

This deck: every slide

HubSpot: every record

Your job

Give it a door



The user walks through

Blender: check the render

This deck: open and check

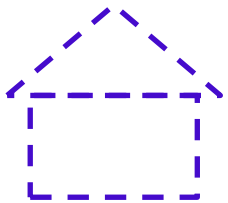
HubSpot: review follow-ups

Your job

A link to the real thing

Guardrails

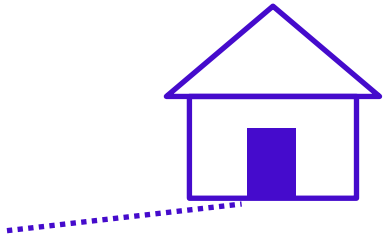
Your Users Still Do the Walkthrough



Their agent builds it

Blender: 421 settings
This deck: every slide
HubSpot: every record

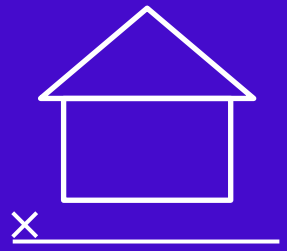
Your job
Give it a door



The user walks through

Blender: check the render
This deck: open and check
HubSpot: review follow-ups

Your job
A link to the real thing



The user signs off

Blender: judge the shot
This deck: give the talk
HubSpot: send follow-ups

Your job
Leave them the final say

Guardrails

Two of Three, Never All Three

A

Reads untrusted input

Email, web pages, notes anyone can edit

B

Touches private data

Your records, your customers, your money

C

Changes or sends things

Writes, payments, messages

An agent running on its own gets any two. All three? A person approves, or something else reliably checks.

Until prompt injection can be reliably caught.

Source: Meta AI, “Agents Rule of Two: A Practical Approach to AI Agent Security,” Oct 31, 2025.

Guardrails

You Don't Know Where Their Agent's Been

#1

Most exploited failure in a year of red-teaming: getting past the human check, through consent fatigue

New

Hidden text and off-screen elements that tell a screen-reading agent what to do

99

CVEs published for MCP-related software in 2025

Google's three principles: well-defined human controllers, carefully limited powers, observable actions.

Sources: Microsoft AI Red Team, Taxonomy of Failure Modes in Agentic AI Systems v2.0, Jun 4, 2026. Díaz, Kern, Olive, “Google's Approach for Secure AI Agents,” 2025.

Guardrails

Two of Three, Never All Three

A

~~Reads untrusted input~~

Their agent? Assume it. Always.

B

Touches private data

Your records, your customers, your money

C

Changes or sends things

Writes, payments, messages

Their agent has always read something untrusted. So it gets one more: your private data, or making changes. Never both on its own.

Until prompt injection can be reliably caught.

Source: Meta AI, “Agents Rule of Two: A Practical Approach to AI Agent Security,” Oct 31, 2025.

Guardrails

Let Agents Draft. People Commit.

Email

The agent drafts it

You hit Send

Guardrails

Let Agents Draft. People Commit.

Email

The agent drafts it

You hit Send

Shopping cart

The agent fills it

You type the card

Guardrails

Let Agents Draft. People Commit.

Email

The agent drafts it

You hit Send

Shopping cart

The agent fills it

You type the card

Anything destructive

The agent stages it

You commit it

Guardrails

Let Agents Draft. People Commit.

Email	Shopping cart	Anything destructive
The agent drafts it	**The agent fills it**	**The agent stages it**
You hit Send	You type the card	You commit it

The draft view shows everything before submit: what changes, what it costs, who it reaches.

Ask at the moments that matter. Not every step.

MCP specification, rev. 2026-07-28, Elicitation: URL mode for “sensitive interactions that must not pass through the MCP client.”

Access

How the Agent Gets In

Borrowing the session is fragile. It's also dangerous.

Access

Session Access Is Not Delegated Authority

A borrowed session

Read everything
Change anything
No end date
Looks exactly like me

Access

Session Access Is Not Delegated Authority

A borrowed session

Read everything
Change anything
No end date
Looks exactly like me

A scoped grant

My call-debrief agent

Reads my call notes
Updates HubSpot records and follow-ups
Expires, and I can revoke it
Never emails my customers

Same agent. Very different blast radius.

Access

Nothing to Invent. Just Adopt.

MIT and others, Jan 2025

Authenticated delegation

Extends OAuth and OpenID Connect so a person can delegate scoped, auditable authority to an agent.

MCP specification, Jul 2026

OAuth 2.1 authorization

The protocol's authorization is built on OAuth 2.1, with protected-resource metadata and resource indicators.

Microsoft, spring 2026

Entra Agent ID

Generally available: agents get their own identities, with purpose-built OAuth flows.

Sources: South et al., arXiv 2501.09674. modelcontextprotocol.io, authorization, rev. 2026-07-28. Microsoft Learn, What's new in Entra Agent ID.

Access

Give the Agent Its Own Door

Call debrief wants to:

✓ Read your Granola call notes

✓ Update HubSpot records and follow-ups

✗ Email your customers

✗ Delete records

Expires in 30 days · revoke any time

- Its own identity, not the person's session
- Scoped grants, per task
- Expiry and revocation
- An audit trail: who did what, for whom

Do this
Give the agent its own door

Illustration: a scoped grant for my call-debrief agent.

What to do

Three Things to Do

1

Put your depth behind a well-scoped API

Reach the deep features. Authorization lives on the server, not in the UI.

What to do

Three Things to Do

1

Put your depth behind a well-scoped API

Reach the deep features. Authorization lives on the server, not in the UI.

2

Give the agent its own door

Whatever fits your product. Scoped, expiring and auditable. Not the person's session.

What to do

Three Things to Do

1

Put your depth behind a well-scoped API

Reach the deep features. Authorization lives on the server, not in the UI.

2

Give the agent its own door

Whatever fits your product. Scoped, expiring and auditable. Not the person's session.

3

Make results witnessable

Receipts and URLs, and drafts that people commit.

Agents Drive.
People Judge.
Software Makes It Reliable.

Alex Thomas, CTO, Applied Visions



Connect on LinkedIn
linkedin.com/in/alexander-jt
alext@avi.com

