

THE COMPLETE

# AI Generalist Roadmap 2026

MAY EDITION | 5 LEVELS | ZERO CODE | ~2 YEARS TO MASTER

**142K**

TECH LAYOFFS YTD  
2026

**20%**

DEV JOB DROP AGES  
22-25

**43%**

MULTI-AI-SKILL  
PREMIUM

**\$700B**

HYPERSCALER AI  
CAPEX

by **Outskill**

# Table of Contents

<b>01</b>	Why This Matters Now	<b>23</b>	Level 4 — Agents & Automation
<b>02</b>	The May 2026 Landscape	<b>24</b>	Dynamic Workflows — New in May
<b>03</b>	Key Model Releases — April & May 2026	<b>25</b>	Level 4 — The Agent Tool Landscape
<b>04</b>	The Model Selection Framework	<b>26</b>	Level 4 — Designing Agents That Work
<b>05</b>	Google I/O 2026 Spotlight <b>NEW</b>	<b>27</b>	Level 4 — Projects to Ship
<b>06</b>	The 5-Level Roadmap	<b>28</b>	Level 4 — Resources & Papers
<b>07</b>	Level 1 — Foundations: The PRD Method	<b>29</b>	Level 5 — Vibe Coding
<b>08</b>	Level 1 — The Starter Stack	<b>30</b>	The Vibe Coding Stack — May 2026
<b>09</b>	Level 1 — Prompting Principles	<b>31</b>	Cursor Composer 2.5 Spotlight
<b>10</b>	Level 1 — Projects to Ship	<b>32</b>	Level 5 — Vibe Coding Principles
<b>11</b>	Level 1 — Resources & Papers	<b>33</b>	Level 5 — Projects to Ship
<b>12</b>	Level 2 — Context & Connections	<b>34</b>	Level 5 — Resources & Papers
<b>13</b>	Level 2 — RAG in May 2026	<b>35</b>	The Claude Ecosystem in May 2026
<b>14</b>	Level 2 — MCP, the Protocol That Won	<b>36</b>	The Claude Power Stack — 9 Surfaces
<b>15</b>	Level 2 — Projects to Ship	<b>37</b>	AI Pricing Cheat Sheet — May 31, 2026
<b>16</b>	Level 2 — Resources & Papers	<b>38</b>	The 2026 Job Market
<b>17</b>	Level 3 — Multimodal Creation	<b>39</b>	The Policy Response <b>NEW</b>
<b>18</b>	The Video AI Shift <b>NEW</b>	<b>40</b>	Case Study: Virgin Voyages
<b>19</b>	Level 3 — Voice + Audio in 2026	<b>41</b>	What's Coming Next — June–Aug 2026
<b>20</b>	Level 3 — Image + Doc	<b>42</b>	Resources & Communities
<b>21</b>	Level 3 — Projects to Ship	<b>43</b>	The Generalist's Commitment
<b>22</b>	Level 3 — Resources & Papers	<b>44</b>	Close — You Don't Need Permission

---

# Why This Matters Now

---

The AI generalist isn't a role. It's the new floor.

- **Meta cut 8,000 (10% of workforce)** on May 20. Intuit cut 3,000 (17%). Coinbase cut 700 (14%).
- These are profitable companies. They're not cutting to survive — they're cutting to fund a **\$700B** AI infrastructure buildout in 2026.
- Stanford HAI's 2026 AI Index: software dev employment for workers **ages 22–25 is down ~20% since 2024**.
- LinkedIn 2025 Workforce Report: AI skills now appear in **42% of software job descriptions**, up from 8% in 2022.
- The pattern: junior coding roles shrink, AI-skilled roles command a **43% premium**, and “AI-fluent generalists” are the new senior hire.

*“Half of AI-attributed layoffs get quietly rehired offshore at 40–60% of US cost.”*

— HR Executive, Q1 2026

## What This Roadmap Does

This is a complete 5-level path from beginner to senior AI generalist. Every level has a definition of done, a starter stack, prompting principles, and three projects to ship. The framework hasn't changed in 2026 — the tools have. May 2026 alone added Claude Opus 4.8, Gemini 3.5 Flash, Cursor Composer 2.5, and DeepSeek V4-Pro at permanent 75%-off pricing. This edition catches you up.

# The May 2026 Landscape

What changed in 30 days

## MODELS

- **Claude Opus 4.8** (May 28) — #1 on Artificial Analysis Index (61.4); 88.6% SWE-bench Verified.
- **Gemini 3.5 Flash** (May 19) — beats 3.1 Pro on coding/agentive at ~40% the price.
- **DeepSeek V4-Pro** made its 75% discount *permanent* (May 22): \$0.435/\$0.87 per 1M tokens.
- **Cursor Composer 2.5** (May 18) — custom RL post-trained on Kimi K2.5 base.
- Plus: Qwen 3.7-Max preview, SubQ 1M-Preview, xAI Grok Build CLI beta.

## AGENTS & PLATFORMS

- **Google Antigravity 2.0** — standalone agent platform with CLI + SDK.
- **Gemini Spark** — persistent personal agent inside the Gemini app (US, Ultra tier).
- **Claude Platform on AWS** — first-of-its-kind, Anthropic operates the service.
- **Dynamic Workflows** in Claude Code — one agent fans out to hundreds of parallel subagents.
- 20+ new Anthropic legal MCP connectors.

## INDUSTRY

- **Sora is dead.** Web/app shut April 26; API ends Sept 24.
- Anthropic in early IPO talks (Goldman, JPM, Morgan Stanley) targeting **October 2026**; declined an \$800B+ valuation round.
- California Gov. Newsom's **May 21 EO** — first state-level response to AI-driven displacement.
- KPMG (May 19) + PwC (May 14) + Stainless acquisition (May 18) all closed Anthropic deals this month.

## The pattern is converging

Three things are now true simultaneously: model capability is still climbing fast, frontier prices are dropping (4x cheaper in 12 months for equivalent quality), and the agent abstraction has shipped at scale across every major lab. **The arbitrage window for AI generalists is wider than it has ever been.** The next 12 months will narrow it.

# Key Model Releases — April & May 2026

Every frontier release in the last 60 days

DATE	MODEL	WHY IT MATTERS
May 28	Claude Opus 4.8	#1 on Artificial Analysis Intelligence Index (61.4); 88.6% SWE-bench Verified; \$5/\$25 per 1M tokens, Fast variant \$10/\$50; 1M context default.
May 22	DeepSeek V4-Pro	Made 75% discount <i>permanent</i> : \$0.435 in / \$0.87 out per 1M tokens. Cheapest frontier-class reasoning model on the market.
May 20	Qwen 3.7-Max (preview)	Alibaba's strongest closed-weight model; multimodal-first; preview only.
May 19	Gemini 3.5 Flash	\$1.50/\$9 per 1M; ~4× faster output (~280 tok/s); beats 3.1 Pro on coding and agentic benchmarks.
May 19	Gemini Spark + Antigravity 2.0	Google's consumer agent (Spark) and developer agent platform (Antigravity) ship together.
May 18	Cursor Composer 2.5	+11.0 on CursorBench v3.1 vs Composer 2; custom RL on Kimi K2.5 base.
May 14	xAI Grok Build CLI (beta)	xAI's developer-facing CLI for agentic workflows.
May 5	SubQ 1M-Preview	First commercial non-Transformer architecture at frontier scale; 12M context; 52× faster attention.
Apr 24	DeepSeek V4	Open-weight reasoning model. Triggered the V4-Pro pricing reset.
Apr 24	Kling 3.0	4K video, multi-shot consistency, native audio. Took #1 on video arena.
Apr 23	GPT-5.5 "Spud"	OpenAI's mass-consumer model; strong multimodal; native realtime.
Apr 16	Claude Opus 4.7	Superseded by 4.8 in just 41 days — signal that Anthropic's cadence has tightened.
Apr 7	Claude Mythos (Glasswing)	Restricted frontier model for cybersecurity research only; no public access.

Also released: Cohere Command A+ (218B sparse MoE), ZAYA1-8B (first frontier model on AMD Instinct), StepAudio 2.5 Realtime.

# The Model Selection Framework

How to pick the right model in May 2026

JOB TO BE DONE	PRIMARY PICK	FALLBACK	WHY
Long-horizon agent / coding	Claude Opus 4.8	Gemini 3.5 Flash	88.6% SWE-bench; Dynamic Workflows.
High-volume agentic	Gemini 3.5 Flash	Opus 4.8 Fast	\$1.50/\$9 at ~280 tok/s.
Cheap reasoning / open-weight	DeepSeek V4-Pro	Qwen 3.7-Max	\$0.435/\$0.87 permanent.
1M+ context	Opus 4.8 (1M)	SubQ 1M-Preview (12M)	Default 1M on Opus; SubQ for massive corpora.
Multimodal input	Gemini 3.5 Flash	GPT-5.5	Best native multimodal-per-dollar.
Mass consumer brainstorming	GPT-5.5	Gemini 3.5 Flash	Strongest distribution + memory.
Frontier / restricted	Claude Mythos	—	Glasswing program only.
Voice / realtime	StepAudio 2.5 Realtime	GPT-5.5 Realtime	Lowest latency, strong RLHF for voice roleplay.

**Rule of thumb:** Opus 4.8 when errors cost money. Gemini 3.5 Flash for everything else. DeepSeek V4-Pro when token cost dominates the equation.

# Google I/O 2026 Spotlight NEW

Google's "agent-first" pivot

On May 19 at the Shoreline Amphitheatre, Google reorganized its entire AI stack around **agents**. Five announcements matter for the generalist:

## 1. Gemini 3.5 Flash

\$1.50/\$9 per 1M tokens — ~40% cheaper than 3.1 Pro and faster on every dimension. 1,048,576 token input / 65,536 output. Beats 3.1 Pro on coding, agentic, and multimodal benchmarks. Output throughput up ~4x.

## 2. Gemini Spark

Persistent personal agent inside the Gemini app, built on 3.5 Flash + Antigravity. US-only on AI Ultra (\$100/mo). No API.

## 3. Antigravity 2.0

Standalone agent platform with CLI, SDK, managed execution, enterprise support, and audit logging. Routes through **Gemini Enterprise Agent Platform** (renamed from Vertex AI Agent Builder).

## 4. AI Mode in Search

Now default globally. Search results lead with an AI-generated answer for most queries.

## 5. AI-Enabled Mouse Pointer

DeepMind experiment — cursor proactively suggests next actions. Limited rollout.

**Coming:** Gemini 3.5 Pro full GA in June 2026 with 2M context. Will be Google's direct response to Opus 4.8.

# The 5-Level Roadmap

Beginner to senior AI generalist in ~2 years

LEVEL	FOCUS	DURATION	WHAT YOU'LL BE ABLE TO DO
L1	Foundations	2–3 months	Solo workflows. Write a PRD. Prompt with intent. Ship 3+ recurring workflows.
L2	Context & Connections	3–4 months	RAG + MCP. A Claude/Gemini that knows your business and takes action.
L3	Multimodal Creation	3–4 months	Image, video, voice, doc. Replace agency-level production.
L4	Agents & Automation	4–6 months	Multi-step autonomous workflows. AI as colleague, not tool.
L5	Vibe Coding	4–6 months	Production software with AI as the primary author.

Total runway: 16–23 months. Levels 2 and 3 can overlap. The framework is sequential by capability, not by calendar.

LEVEL 1

# Foundations: The PRD Method

~2–3 months · Solo workflows, written specs, intentional prompting

## The core skill: write a spec, then ship.

Every senior PM and engineer in 2026 already does this — they write a one-page PRD (Product Requirements Document) before prompting. The loop is **Problem** → **Requirements** → **Deliver**. Skip the spec, you get garbage. Write the spec, you get gold.

## Why this works

Modern frontier models are reasoning systems. They reward structure. A 60-second written PRD turns “help me with my email” into a workflow that saves 30 minutes a day, every day, forever.

## You're done when...

- You have 3+ recurring workflows saving 30+ minutes/day.
- You can write a 1-page PRD in under 10 minutes.
- You stop saying “AI is bad at X” and start saying “my prompt was bad.”

# Level 1 — The Starter Stack

What to install/subscribe to in your first week

## Chat (pick one, \$20/month)

- **Claude Pro** — backed by Opus 4.8. Best for reasoning, writing, code.
- **ChatGPT Plus** — backed by GPT-5.5. Best for mass-market features, memory.
- **Gemini Advanced** — backed by 3.5 Flash. Best for Google integrations, multimodal.

## Image generation

- **Midjourney v8** for art-direction quality.
- **GPT Image** for fast in-chat iteration.
- **Imagen 4** for photoreal and brand-safe output.

## Voice

- **ElevenLabs** for voice generation.
- **Whisper** (free, local or API) for transcription.

## Browser

- **Claude in Chrome** (beta) — browser-side agent that can read pages and take actions.

If you can only pay for one thing: **Claude Pro**. Opus 4.8 is currently the strongest all-around foundation model for the kinds of workflows L1 is about.

# Level 1 — Prompting Principles

Five rules that make every model 10× better

## 1. Show, don't tell

Paste 2 examples of what you want. Frontier models pattern-match on examples far better than on adjectives.

## 2. Constrain the format

“Return a markdown table with columns: Name, Risk, Mitigation, Owner.” Vague format = vague output.

## 3. Assign a specific role

Not “you are a lawyer” — “you are a senior tax attorney specializing in cross-border M&A for Delaware C-corps.” Specificity activates the right priors.

## 4. Use extended thinking

Opus 4.8, GPT-5.5, and Gemini 3.5 Flash all support extended thinking. Turn it on for anything non-trivial. It's free quality.

## 5. Save your wins

Every prompt that produces a great result goes into a personal library (a Notion doc, an Apple Note, anything). Reuse and refine. You're building an asset, not having a conversation.

**Anti-pattern:** treating chat models like Google. They're reasoning systems, not retrieval systems. Give them context and constraints, not keywords.

---

# Level 1 — Projects to Ship

---

Three workflows that pay back the entire roadmap

## Project 1: Your Weekly Digest

Paste your inbox, Slack DMs, and 3 newsletter excerpts every Sunday. Get back a Monday-morning brief: what shipped, what's blocked, what needs a decision this week. **Saves: 60+ min/week.**

## Project 2: The Interview Prep Machine

Paste the job description, your resume, and 4 of the company's earnings calls. Get back: 20 likely questions, your best answer to each, and 3 stories you should rehearse. **Saves: 4–6 hours per interview cycle.**

## Project 3: The Doc Collapser

30+ page PDF in, 1-page summary out: TL;DR, 5 key risks, 5 questions for the author, and an exec-ready paragraph. **Saves: 2–3 hours per long doc.**

**Bonus:** set up a Claude Project (or Custom GPT) for each of these. Once built, they're always one click away.

# Level 1 — Resources & Research Papers

Foundations stack — updated May 2026

## Guides & Documentation

### Anthropic Prompt Engineering

[docs.claude.com/en/docs/build-with-claude/prompt-engineering](https://docs.claude.com/en/docs/build-with-claude/prompt-engineering)

### What's New in Opus 4.8

[docs.claude.com/en/about-claude/models/whats-new-claude-4-8](https://docs.claude.com/en/about-claude/models/whats-new-claude-4-8)

### OpenAI Introducing GPT-5.5

[openai.com/index/introducing-gpt-5-5](https://openai.com/index/introducing-gpt-5-5)

### OpenAI Prompt Engineering

[platform.openai.com/docs/guides/prompt-engineering](https://platform.openai.com/docs/guides/prompt-engineering)

### Gemini 3.5 Flash Announcement

[blog.google/technology/google-deepmind/gemini-3-5-flash](https://blog.google/technology/google-deepmind/gemini-3-5-flash)

### Google NotebookLM

[notebooklm.google.com](https://notebooklm.google.com)

### Ollama (run LLMs locally)

[ollama.com](https://ollama.com)

### LM Studio

[lmstudio.ai](https://lmstudio.ai)

### LLM Stats — live leaderboard

[llm-stats.com](https://llm-stats.com)

### Artificial Analysis Intelligence Index

[artificialanalysis.ai](https://artificialanalysis.ai)

### DeepSeek V4-Pro on HF

[huggingface.co/deepseek-ai/DeepSeek-V4-Pro](https://huggingface.co/deepseek-ai/DeepSeek-V4-Pro)

## Research Papers

### Attention Is All You Need (Vaswani, 2017)

[arxiv.org/abs/1706.03762](https://arxiv.org/abs/1706.03762)

### Chain-of-Thought Prompting

[arxiv.org/abs/2201.11903](https://arxiv.org/abs/2201.11903)

### Few-Shot Learners (GPT-3)

[arxiv.org/abs/2005.14165](https://arxiv.org/abs/2005.14165)

### Survey of Prompt Engineering

[arxiv.org/abs/2402.07927](https://arxiv.org/abs/2402.07927)

### LoRA — Low-Rank Adaptation

[arxiv.org/abs/2106.09685](https://arxiv.org/abs/2106.09685)

### Hallucination Mitigation Survey

[arxiv.org/abs/2401.01313](https://arxiv.org/abs/2401.01313)

### OWASP LLM01:2025 Prompt Injection

[genai.owasp.org/llm-top-10-2025/llm01-prompt-injection](https://genai.owasp.org/llm-top-10-2025/llm01-prompt-injection)

### Sentence Transformers

[sbert.net](https://sbert.net)

TOOLS: ChatGPT (GPT-5.5) | Claude (Opus 4.8) | Gemini 3.5 Flash | Perplexity | DeepSeek V4-Pro | NotebookLM | Ollama | LM Studio

LEVEL 2

# Context & Connections

~3–4 months · RAG + MCP. AI that knows your business and takes action.

## The unlock

L1 is talking to a brilliant generalist who knows nothing about you. L2 is talking to a brilliant generalist who has read everything you've written and can act on your behalf. Two technologies make that possible:

- **RAG (Retrieval-Augmented Generation)** — static knowledge from your documents, notes, and codebases.
- **MCP (Model Context Protocol)** — live tools and APIs: calendar, Gmail, Notion, Drive, CRM, your codebase.

Real workflows need both. Static facts + live actions = a useful colleague.

## You're done when...

- You have a domain-specific Claude/Gemini that knows your business.
- It takes 3+ types of action without copy-paste.
- You can teach a non-technical teammate to use it in 15 minutes.

---

# Level 2 — RAG in May 2026

---

The state of the retrieval stack

## Managed RAG (pick one)

- **Claude Projects** — upload up to ~200K tokens of context per project, accessible inside Claude.
- **NotebookLM** — Google's research-grade notebook with citations and audio overviews.
- **ChatGPT Custom GPTs** — for users already in the OpenAI ecosystem.

## Build-your-own (developer path)

- **LlamaIndex / LangChain** for orchestration.
- **Pinecone / Weaviate / pgvector** for vector storage.
- **Cohere Rerank / Voyage** for retrieval quality.
- Opus 4.8's 1M-token default context often beats RAG entirely for docs under ~750 pages.

**The 2026 rule:** try long-context first, fall back to RAG. Models got cheap enough that loading the whole doc is now often the right answer.

# Level 2 — MCP, the Protocol That Won

Model Context Protocol — the standard for AI-to-tool connections

MCP is an open standard introduced by Anthropic, now adopted by OpenAI, Google, and Microsoft. It's the USB-C of AI tools: one protocol, any AI, any service.

## What MCP gives you

- **One protocol** for connecting any AI client to any tool.
- **Pre-built servers** for Gmail, Calendar, Drive, GitHub, Slack, Notion, Linear, Asana, Stripe, Postgres, and dozens more.
- **Custom servers** — build your own in ~50 lines of code for any internal API.

## Where MCP runs in May 2026

- Claude Desktop, Claude Code, Claude in Chrome.
- Anthropic's 20+ legal-vertical MCP connectors (added May 2026).
- Cursor, Windsurf, Continue.dev — all support MCP servers natively.
- Google's Antigravity 2.0 and Gemini Enterprise Agent Platform support MCP for tool calls.

**The point:** learn MCP once, plug AI into every system you use. It's the closest thing the AI ecosystem has to a universal standard.

---

## Level 2 — Projects to Ship

---

From talking AI to acting AI

### **Project 1: “Chat With My Company”**

Create a Claude Project with 12 months of your team's notes, docs, and specs. Ask it questions like “what did we decide about pricing in Q3?” or “what's our position on the partnership with X?” Becomes the institutional memory you wish your wiki was.

### **Project 2: My CRM With a Brain**

Connect Claude to your CRM (Salesforce, HubSpot, Attio) via MCP. Now ask: “which deals are slipping this month and why?” Get a real answer, with sources, in 30 seconds.

### **Project 3: Personal Knowledge Base**

Connect Notion + Drive + Apple Notes into a single Claude Project. Result: a personal knowledge graph that answers in your voice, with citations. Bonus: feed 6 months of journal entries and ask it to summarize patterns.

# Level 2 — Resources & Research Papers

Context, MCP & voice — updated May 2026

## MCP & Agent Protocols

### MCP Official Docs

[modelcontextprotocol.io](https://modelcontextprotocol.io)

### Claude MCP Connectors Setup

[docs.claude.com/en/docs/agents-and-tools/mcp](https://docs.claude.com/en/docs/agents-and-tools/mcp)

### MCP GitHub

[github.com/modelcontextprotocol](https://github.com/modelcontextprotocol)

### Agentic AI Foundation

[linuxfoundation.org/projects/aaif](https://linuxfoundation.org/projects/aaif)

### Introducing MCP — Anthropic

[anthropic.com/news/model-context-protocol](https://anthropic.com/news/model-context-protocol)

### What is MCP? — IBM

[ibm.com/topics/model-context-protocol](https://ibm.com/topics/model-context-protocol)

### Code Execution with MCP

[anthropic.com/research/code-execution-mcp](https://anthropic.com/research/code-execution-mcp)

### MCP vs A2A — Complete Guide (2026)

[dev.to/pocket\\_tools/mcp-vs-a2a-complete-guide](https://dev.to/pocket_tools/mcp-vs-a2a-complete-guide)

## RAG & Vector Infrastructure

### Pinecone

[pinecone.io](https://pinecone.io)

### LlamaIndex

[llamaindex.ai](https://llamaindex.ai)

### LangChain

[langchain.com](https://langchain.com)

### Function Calling Guide

[promptingguide.ai/capabilities/function-calling](https://promptingguide.ai/capabilities/function-calling)

### What is RAG? — IBM

[ibm.com/topics/retrieval-augmented-generation](https://ibm.com/topics/retrieval-augmented-generation)

## Voice Agents

### ElevenLabs Conversational AI

[elevenlabs.io/conversational-ai](https://elevenlabs.io/conversational-ai)

### Deepgram Voice AI

[deepgram.com/learn/voice-ai-agents](https://deepgram.com/learn/voice-ai-agents)

### OpenAI Whisper v4

[github.com/openai/whisper](https://github.com/openai/whisper)

### Hume EVI 2

[hume.ai](https://hume.ai)

### Sesame

[sesame.com](https://sesame.com)

### Multi-Turn Conversations — OpenAI

[platform.openai.com/docs/guides/conversation-state](https://platform.openai.com/docs/guides/conversation-state)

TOOLS: Claude + MCP | Slack | Notion | Google Drive | Gmail | Asana | Pinecone | LangChain | ElevenLabs

LEVEL 3

# Multimodal Creation

~3–4 months · Stop typing. Start showing.

## The unlock

L1 and L2 are text-first. L3 is cross-modal: see, hear, generate across formats. What this makes possible:

- Whiteboard photo → structured code in 60 seconds.
- 90-minute meeting → action items, decisions, and 3 follow-up emails.
- 4K product launch videos with synced audio in an afternoon.
- PDF tables with merged cells, footnotes, and rotated text — correctly extracted.

Multimodal isn't a feature in 2026 — it's the default mode of frontier models.

# The Video AI Shift NEW

Sora is dead. Here's who won.

OpenAI shut down Sora's web/app on **April 26, 2026**; API ends **September 24**. Reported \$15M/day operating cost vs \$2.1M lifetime revenue. Disney walked from a \$1B licensing deal. The market reset overnight.

## The May 31 Arena Leaderboard

RANK	MODEL	ELO	STRENGTH
1	Kling v3 (Kuaishou)	2127	Multi-shot consistency, audio sync, 4K.
2	Seedance 2.0 Fast (ByteDance)	1993	Image-to-video, native physics.
3	Happy Horse 1.0	1962	Best physical motion realism.
—	Veo 3.1 (Google)	Top tier	Cinematic, 4K + audio.
—	Runway Gen-4.5	Pro favorite	Camera control, character consistency.
—	Grok Imagine 1.0	Speed king	~17 sec/clip generation.

## Pricing snapshot

Kling 3.0: ~\$0.07–0.10/sec (65% below Sora 2). Veo 3.1: premium tier. Runway Gen-4.5: \$12–\$95/mo subscription. Sora 2 was \$0.75/sec (5× Veo Fast) — one of several reasons it died.

**Pick rule:** cinematic → Veo 3.1. Volume + cost → Kling 3.0. Director-style control → Runway Gen-4.5. Image-to-video → Seedance 2.0.

---

# Level 3 — Voice + Audio in 2026

---

Generation, transcription, realtime

## Generation

- **ElevenLabs v3** — voice cloning, 30+ languages, emotional control.
- **Suno v5 / Udio v3** — music generation, full vocals, multi-genre.
- **StepAudio 2.5 Realtime** — lowest-latency open model, RLHF for voice roleplay.

## Transcription

- **Whisper v4** — open-source, runs locally.
- **AssemblyAI Universal-2** — paid, best diarization.
- **Gemini 3.5 Flash** — cheapest native audio understanding.

## Realtime conversation

- **GPT-5.5 Realtime API** — the production default.
- **OpenAI Voice Mode** — consumer.
- **Sesame + Hume EVI 2** — emotionally aware voice agents.

## Real workflows you can ship today

- **Voice notes** → **structured docs** via Granola, Reflect, or Voibe.
- **Meetings** → **minutes** via Granola, Otter, or Zoom AI Companion.
- **Customer support voice bots** via Vapi or Retell.

---

# Level 3 — Image + Doc

---

The other half of multimodal

## Image generation

- **Midjourney v8** — art direction quality, best aesthetic.
- **Imagen 4 Ultra** — photoreal, brand-safe, strongest on people.
- **Flux 2** — open-weight, fastest local generation.
- **GPT Image** — in-chat speed and iteration.
- **Recraft v3** — vector output, brand consistency.

## Document understanding

- **Claude Opus 4.8** — native PDF, screenshots, tables; 1M context handles full SEC filings.
- **Gemini 3.5 Flash** — best multimodal-input-per-dollar.
- **Anthropic File API** — structured file ingestion at scale.

## OCR & scanning

- **Mistral Document AI** — production OCR with layout.
- **Reducto / Unstructured** — pipeline-grade document parsing.
- For one-offs: just use Opus 4.8 directly. It's better than most pipelines.

---

# Level 3 — Projects to Ship

---

Replace the agency

## Project 1: Product Launch Video Kit

Veo 3.1 for the cinematic 30-second hero. Kling 3.0 for 6 social cuts. Suno for the theme music. ElevenLabs for voiceover in 4 languages. **Total time: one afternoon. Was: \$20K agency engagement.**

## Project 2: Meeting-to-Everything Pipeline

Zoom recording → transcript → action items + 3 follow-up emails + Notion page + Slack post — all auto-generated. Wire it up with Granola or Gemini 3.5 Flash + MCP.

## Project 3: Whiteboard-to-PRD

Photo of a whiteboard → Opus 4.8 → structured PRD with sections (Problem / Users / Requirements / Risks / Open Questions). End the “I’ll type this up later” debt forever.

# Level 3 — Resources & Research Papers

Multimodal creation stack — updated May 2026

## Generation Tools

**Midjourney v8 Docs**

[docs.midjourney.com](https://docs.midjourney.com)

**Kling v3**

[klingai.com](https://klingai.com)

**Seedance 2.0 Fast**

[jimeng.com](https://jimeng.com)

**Veo 3.1 — Google DeepMind**

[deepmind.google/technologies/veo](https://deepmind.google/technologies/veo)

**Runway Gen-4.5**

[runwayml.com](https://runwayml.com)

**Imagen 4 Ultra**

[deepmind.google/technologies/imagen](https://deepmind.google/technologies/imagen)

**Flux 2 — Black Forest Labs**

[bfl.ai](https://bfl.ai)

**Ideogram**

[ideogram.ai](https://ideogram.ai)

**ElevenLabs v3**

[elevenlabs.io](https://elevenlabs.io)

**Suno v5**

[suno.com](https://suno.com)

**Udio v3**

[udio.com](https://udio.com)

## Research Papers & Reference

**DDPM (Ho et al., 2020)**

[arxiv.org/abs/2006.11239](https://arxiv.org/abs/2006.11239)

**Latent Diffusion (Stable Diffusion)**

[arxiv.org/abs/2112.10752](https://arxiv.org/abs/2112.10752)

**Text-to-Image Survey**

[arxiv.org/abs/2303.07909](https://arxiv.org/abs/2303.07909)

**ControlNet**

[arxiv.org/abs/2302.05543](https://arxiv.org/abs/2302.05543)

**Stable Diffusion Guide — HF**

[huggingface.co/docs/diffusers/stable\\_diffusion](https://huggingface.co/docs/diffusers/stable_diffusion)

**Runway Gen-4 Intro**

[runwayml.com/blog/introducing-runway-gen-4](https://runwayml.com/blog/introducing-runway-gen-4)

**AI Video Generators 2026 Compared**

[aimlapi.com/blog/best-ai-video-generators-2026](https://aimlapi.com/blog/best-ai-video-generators-2026)

TOOLS: Kling v3 | Veo 3.1 | Midjourney v8 | ElevenLabs v3 | Suno v5 | Runway Gen-4.5 | Flux 2 | Ideogram | Seedance 2.0

LEVEL 4

# Agents & Automation

~4–6 months · Stop prompting. Start delegating.

## What an agent actually is

An agent is a workflow where the *model* decides what to do next — not a human, not a deterministic if/else. The model plans, calls tools, observes the result, and iterates until done.

## What changed in 2026

- Reliability crossed 80% on multi-step real-world tasks.
- Anthropic shipped **Dynamic Workflows** (May 28) — agents that spawn parallel subagents.
- Google's **Antigravity 2.0** made agent infrastructure enterprise-grade.
- **Gemini Spark** brought a persistent agent to consumers.

## The shift

From “AI is a tool I prompt” to “AI is a colleague I delegate to.” Generalists who internalize this in 2026 ship 5–10× more than peers who don't.

# Dynamic Workflows — New in May NEW

Anthropic's answer to long-horizon agentic work

Shipped May 28 alongside Opus 4.8. Research preview inside Claude Code. The primitive: one agent fans out to hundreds of subagents, then merges their work.

## How it works

1. You give one goal: "Audit our 2M-line codebase for SQL injection vulnerabilities."
2. A **planner agent** scopes the work and parallelizes by module.
3. Hundreds of **subagents** spin up, each owning one module.
4. Results stream back in real time.
5. A **merger agent** consolidates findings, dedupes, and ranks by severity.

All inside a single Claude Code session. No orchestration code to write. No infra to provision.

This is the same primitive Virgin Voyages used to go from 50 to 1,500 agents in 6 months (a **2,900% increase**). Same primitive that powers Antigravity 2.0 and OpenAI's Operator. **Dynamic Workflows is the agent abstraction that finally works at scale.**

---

# Level 4 — The Agent Tool Landscape

---

What to use, when

## General-purpose agent platforms

- **Claude Agent SDK + Claude Code** — strongest for coding-adjacent agents.
- **Google Antigravity 2.0 + Gemini Enterprise Agent Platform** (renamed from Vertex AI Agent Builder).
- **OpenAI Agents SDK + Responses API** — production multi-agent orchestration.
- **LangGraph / CrewAI / AutoGen** — open-source orchestration frameworks.

## Vertical agent platforms

- **Manus** — general consumer agent.
- **Devin / Cognition** — software engineering.
- **Decagon / Sierra / Cresta** — customer support.
- **Harvey / Hebbia** — legal & finance.

## Consumer agents

- **Gemini Spark** — persistent personal agent (Gemini app).
- **ChatGPT Agent Mode** — web-browsing autonomous agent.
- **Claude in Chrome** — browser agent (beta).

## Enterprise infrastructure

- **Claude Platform on AWS** — Anthropic operates the service.
- **Anthropic Managed Agents** — fully hosted agent execution.
- **Stainless** — SDK auto-generation (acquired by Anthropic May 18).

---

# Level 4 — Designing Agents That Work

---

Four principles from production deployments

## 1. Scope ruthlessly

“Personal assistant” fails. “Weekly expense-report categorizer” ships. The smaller the scope, the higher the reliability, the faster the iteration loop.

## 2. Give tools, not instructions

Don't describe how to do a task — expose the right tools and let the model figure it out. Frontier models in 2026 are better planners than most prompt engineers.

## 3. Plan failure modes

Human-in-the-loop checkpoints. Max-cost ceilings. Max-step limits. Rollback paths. Every production agent has all four.

## 4. Operational readiness beats benchmarks

Virgin Voyages did not pick agents by SWE-bench score. They measured **time-to-resolution**, **error rate**, and **CSAT impact**. Pick your metrics before you pick your model.

*“We went from 50 agents to 1,500 in 6 months. The ones that mattered weren't the smartest — they were the most narrowly scoped.”*

— Virgin Voyages, May 2026

---

# Level 4 — Projects to Ship

---

Three agents that prove the level

## **Project 1: Inbox Triage Agent**

Runs overnight. Output every morning: 5 drafted replies, 12 archives, 8 deletes, 3 flagged for you. Build with Claude Agent SDK + Gmail MCP. Iterate the prompt over 2 weeks until you trust it.

## **Project 2: Weekly Competitive Intel Agent**

Every Monday: agent visits 5 competitors' websites, changelogs, LinkedIn pages, and job postings. Returns a 1-page memo: what shipped, what hiring signals, what to watch.

## **Project 3: Codebase-Roaming Agent**

Claude Code + Dynamic Workflows. Goal: "Document our codebase." Returns: architecture overview, 10 code smells, test coverage map. This is the agent you ship to prove L4 to a CTO.

---

# Level 4 — Resources & Research Papers

---

Agents & automation — updated May 2026

## Automation Platforms

**Claude Cowork (desktop)**

[claude.com/product/cowork](https://claude.com/product/cowork)

**Claude for Chrome**

[claude.com/claude-for-chrome](https://claude.com/claude-for-chrome)

**Claude Computer Use Docs**

[docs.claude.com/en/docs/agents-and-tools/computer-use](https://docs.claude.com/en/docs/agents-and-tools/computer-use)

**n8n**

[n8n.io](https://n8n.io)

**Make.com**

[make.com](https://make.com)

**Zapier**

[zapier.com](https://zapier.com)

**CrewAI**

[crewai.com](https://crewai.com)

**Lindy AI**

[lindy.ai](https://lindy.ai)

**Relevance AI**

[relevanceai.com](https://relevanceai.com)

## Foundational Reading

**Dynamic Workflows — Anthropic (May 28)**

[anthropic.com/news/dynamic-workflows](https://anthropic.com/news/dynamic-workflows)

**Building Effective Agents**

[anthropic.com/research/building-effective-agents](https://anthropic.com/research/building-effective-agents)

**Multi-Agent System — IBM**

[ibm.com/topics/multi-agent-system](https://ibm.com/topics/multi-agent-system)

**AI Agent Orchestration — IBM**

[ibm.com/topics/ai-agent-orchestration](https://ibm.com/topics/ai-agent-orchestration)

**LangChain Docs**

[docs.langchain.com](https://docs.langchain.com)

**LangChain Academy**

[academy.langchain.com](https://academy.langchain.com)

**CrewAI GitHub**

[github.com/joaomdmoura/crewAI](https://github.com/joaomdmoura/crewAI)

**n8n Beginner's Guide**

[freecodecamp.org/news/a-beginners-guide-to-automation-with-n8n](https://freecodecamp.org/news/a-beginners-guide-to-automation-with-n8n)

**What is RPA? — IBM**

[ibm.com/topics/rpa](https://ibm.com/topics/rpa)

TOOLS: [Claude Cowork](#) | [Claude Chrome](#) | [n8n](#) | [Make.com](#) | [Zapier](#) | [Relevance AI](#) | [Lindy](#) | [CrewAI](#)

LEVEL 5

# Vibe Coding

~4–6 months · Describe what you want. AI writes the code. You ship it.

## Where the term came from

Andrej Karpathy coined “vibe coding” for the workflow where you describe what you want, the AI writes code, you review and refine, and ship. By May 2026 it's no longer a meme:

- **92% of devs** use an AI coding assistant daily (Stack Overflow 2025).
- **Cursor** raised at **\$9.9B** in May.
- **Composer 2.5 / Claude Code / Devin** can close PRs end-to-end.
- **Non-coders** are shipping production apps via Lovable, Bolt, and v0.

## The honest framing

This is the highest-leverage level. It's also the level most likely to look different in 18 months. Pick the principles, not the tools — the next iteration will differ in execution, not in principle.

---

# The Vibe Coding Stack — May 2026

---

Pick by who you are

## For non-technical founders, PMs, designers

- **Lovable** — best full-stack: auth, DB, payments, Stripe, Supabase. Ships in ~35 min.
- **Bolt.new** — fastest (~28 min); technical founders often prefer it.
- **v0 by Vercel** — highest output quality (9/10); React/Next.js + shadcn/ui.
- **Replit** — full infrastructure, DB, deploy in one IDE.

## For developers

- **Cursor** — 9.2/10. The default IDE. Composer 2.5 since May 18.
- **Claude Code** — Pro (\$20) / Max (\$100 or \$200). Dynamic Workflows native.
- **Windsurf** — most production-ready; now OpenAI-owned.
- **GitHub Copilot** — usage-based “AI credit” pricing starting June 1.

## Free path

Voibe + Cline + your existing API credits. Windsurf's free tier (25 Cascade credits/month) is the most generous starting point.

---

# Cursor Composer 2.5 Spotlight

---

Why Cursor stopped being just an IDE

Composer 2.5 is Cursor's in-house model, built on a Moonshot Kimi K2.5 base with proprietary reinforcement-learning post-training on actual developer workflows from the Cursor user base.

## The gains over Composer 2

- **+11.0** on CursorBench v3.1
- **+6.1** on SWE-Bench Multilingual
- **+7.6** on Terminal-Bench 2.0

## Why it matters strategically

Cursor is no longer just an IDE wrapper. Anysphere now owns a piece of the inference stack — faster, better multi-file refactors, and a defensible moat as the company heads toward an IPO.

## Trade-off

Opus 4.8 is still better for one-shot complex reasoning. Composer 2.5 is the fastest path to “code that runs” for in-flow IDE work. Use both; let the surface decide.

---

# Level 5 — Vibe Coding Principles

---

Five rules that separate ship from spaghetti

## 1. PRD before code

30 minutes of writing saves 4 hours of “fix one bug, introduce two more.” The PRD discipline from L1 pays its biggest dividends here.

## 2. Watch the burn rate

A messy 4-hour Claude Code session can cost \$50+. Set a budget per project. Cancel runs that aren't converging.

## 3. Version control like your life depends on it

Commit every working state. Branch every experiment. AI-authored code is fast to write and fast to break.

## 4. Security is your problem

The *Lovable VibeScamming* incident, and Guardio Labs' finding that AI coding assistants introduce prompt-injection backdoors, are not edge cases. Check every output before deploy — especially for fintech, healthcare, or government work.

## 5. Don't ship what you can't explain

You should be able to describe what each function does in one sentence. If you can't, you don't understand the code, and you can't maintain it. Use AI as a junior dev, not as a black box.

---

# Level 5 — Projects to Ship

---

Real software, your name on it

## Project 1: Full-Stack SaaS in a Weekend

Lovable or Bolt. Build something with auth + DB + payments + a public URL. Charge \$5/month. Get one paying customer. That's the bar.

## Project 2: The Internal Tool IT Won't Build

Every company has 50 of these. Replit or v0. Build the one your team has complained about for 6 months. Ship it Monday.

## Project 3: Open a PR on a Real Repo (Challenge)

Claude Code or Composer 2.5 on a real open-source bug. Get the PR reviewed and merged. **Single best proof-of-level signal for L5.**

# Level 5 — Resources & Research Papers

Vibe coding stack — updated May 2026

## Vibe Coding Tools

**Cursor (Composer 2.5)**

[cursor.com](https://cursor.com)

**Bolt.new**

[bolt.new](https://bolt.new)

**Lovable.dev**

[lovable.dev](https://lovable.dev)

**Replit Agent 3**

[replit.com](https://replit.com)

**v0 by Vercel**

[v0.dev](https://v0.dev)

**Windsurf**

[codeium.com/windsurf](https://codeium.com/windsurf)

**Claude Code Docs**

[docs.claude.com/en/docs/agents-and-tools/claude-code](https://docs.claude.com/en/docs/agents-and-tools/claude-code)

**GitHub Copilot**

[github.com/features/copilot](https://github.com/features/copilot)

**Grok Build CLI (xAI)**

[x.ai](https://x.ai)

**OpenAI Codex**

[openai.com/codex](https://openai.com/codex)

## Hosting & Backend

**Vercel**

[vercel.com](https://vercel.com)

**Supabase**

[supabase.com](https://supabase.com)

**Firebase**

[firebase.google.com](https://firebase.google.com)

**Google AppSheet**

[cloud.google.com/appsheet](https://cloud.google.com/appsheet)

## Foundational Reading

**Low-Code vs No-Code — IBM**

[ibm.com/blog/low-code-vs-no-code](https://ibm.com/blog/low-code-vs-no-code)

**Vibe Coding Tools Compared (May 2026)**

[till-freitag.com/en/blog/vibe-coding-tools-comparison](https://till-freitag.com/en/blog/vibe-coding-tools-comparison)

**Cursor vs Bolt vs Lovable**

[lovable.dev/guides/cursor-vs-bolt-vs-lovable-comparison](https://lovable.dev/guides/cursor-vs-bolt-vs-lovable-comparison)

**15 Best Vibe Coding Tools**

[index.dev/blog/vibe-coding-tools](https://index.dev/blog/vibe-coding-tools)

**Best Vibe Coding Tools — Roadmap.sh**

[roadmap.sh/vibe-coding/best-tools](https://roadmap.sh/vibe-coding/best-tools)

**TOOLS:** [Cursor 2.5](#) | [Bolt](#) | [Lovable](#) | [Replit](#) | [v0](#) | [Windsurf](#) | [Codex](#) | [Claude Code](#) | [Copilot](#) | [Grok Build CLI](#)

---

# The Claude Ecosystem in May 2026

---

Why Anthropic is the operator's bet

- **Opus 4.8 (May 28):** #1 on Artificial Analysis Intelligence Index (61.4) — beats GPT-5.5 (60.2) and the prior Opus 4.7 (57.3).
- **Benchmarks:** 88.6% SWE-bench Verified; 84% Online-Mind2Web; 96.7% USAMO 2026 (+27 over 4.7); first model >10% on the Legal Agent Benchmark (all-pass tier).
- **Revenue:** ~\$30B annualized as of early April 2026.
- **IPO:** Early talks with Goldman, JPM, and Morgan Stanley targeting October 2026. Declined an \$800B+ valuation round.
- **Cloud:** Claude Platform on AWS — first-of-its-kind; Anthropic operates the service end-to-end; all features available day one.
- **MCP:** 20+ new legal-vertical connectors plus 12 practice-area plugins.
- **Enterprise:** KPMG (276K-person workforce, May 19), PwC (May 14), Stainless acquisition (May 18).
- **SMB:** Claude for Small Business launched May 13.
- **Geography:** Milan office opened May 27.
- **Stance:** Ad-free permanently. 81K-person multilingual study on Claude usage patterns published May.

# The Claude Power Stack — 9 Surfaces

Where Claude shows up in May 2026

SURFACE	WHAT IT DOES	STATUS
Claude (web / mobile / desktop)	Conversational chat, projects, files	GA
Claude Code	Terminal-based coding + Dynamic Workflows	GA
Cowork	Desktop file & task automation for non-developers	Beta
Claude in Chrome	Browser-side agent	Beta
Claude in Excel	Spreadsheet agent	Beta
Claude in PowerPoint	Presentation generation	Beta
Claude in Word	Document drafting	Beta
Claude in Slack	Team chat assistant	GA
Claude Design	Design-system to code (launched Apr 17)	Research Preview

Behind all of these: **Opus 4.8** for hard work, **Sonnet 4.6** for fast turns, **Haiku 4.5** for cheap volume. Connected by **MCP**, the **Claude Agent SDK**, and **Managed Agents**.

**The bet:** AI lives as a layer inside every productivity surface, not as a separate app you switch to. May 2026 is the month that bet became visibly real.

# AI Pricing Cheat Sheet — May 31, 2026

Subscriptions and API rates at a glance

## Consumer chat (\$/month)

- **Claude Pro** \$20 · **Max** \$100 or \$200
- **ChatGPT Plus** \$20 · **Pro** \$200
- **Gemini Advanced** \$20 · **AI Ultra** \$100 or \$200
- **Cursor Pro** \$20 · **Business** \$40/user

## API pricing (per 1M tokens, input / output)

MODEL	INPUT	OUTPUT	NOTES
Claude Opus 4.8	\$5	\$25	1M context default
Claude Opus 4.8 Fast	\$10	\$50	2.5× speed
Claude Sonnet 4.6	\$3	\$15	1M context since Feb
GPT-5.5	\$5	\$25	
Gemini 3.5 Flash	\$1.50	\$9	\$0.15 cached
Gemini 3.1 Pro	\$2.50	\$15	Superseded
DeepSeek V4-Pro	\$0.435	\$0.87	Permanent 75% off
Qwen 3.7-Max	TBA	TBA	Preview only

## Vibe coding tools

Lovable / Bolt / v0 / Replit / Cursor: \$20–\$50/mo + credit consumption. GitHub Copilot: usage-based “AI credit” pricing starting June 1. Claude Code is included in Pro and Max.

# The 2026 Job Market

What the data actually shows

## The cuts

- **142K** tech layoffs YTD 2026.
- **\$700B** in 2026 capex from Amazon, Microsoft, Alphabet, Meta — nearly 2× 2025.
- Meta cut **8,000 (10%)** on May 20. Intuit **3,000 (17%)**. Coinbase **700 (14%)**.
- Software postings down **36.4%** vs February 2020; down **6.7%** year-over-year.

## The premium

- AI engineer base salary **\$206K** (2025), **+7%** in Q1 2026.
- Multi-AI-skill workers earn a **+43%** premium.
- Mid-level AI engineer comp **+9.2%** (Motion Recruitment, 2026).
- AI-specialization job postings **+49% YoY**.

## The split

- Stanford HAI: 22–25-year-old dev employment **down ~20%** since 2024.
- BLS still projects **17% growth** in SWE through 2033 (327,900 new jobs).
- AI skills in **42%** of software JDs (LinkedIn 2025) — up from 8% in 2022.

**The rehire pattern:** roughly half of AI-attributed layoffs are quietly rehired offshore at 40–60% of US cost. The story is “AI replaced this work” — the reality is often “AI made offshoring cheaper.”

# The Policy Response NEW

Governments are starting to react

## California — May 21, 2026

Governor Newsom signed an executive order directing state agencies to study AI displacement, reform the WARN Act for AI layoffs, and expand unemployment insurance eligibility. **180-day deadline** for policy recommendations. **October 15 deadline** for collective bargaining review.

## European Union

AI Factories program (**€20B InvestAI**) begins first deployments in June 2026. Sovereign compute distributed across 7 member states.

## US federal

Stricter pre-release safety testing (Glasswing-model precedent). OSTP weighing displacement-linked tax credits. BLS now tracks “AI-displaced” as a distinct layoff category.

**What this means for the generalist:** the “move fast and break things” era is winding down. Operational readiness, audit logging, and human-in-the-loop are now table stakes — not optional polish.

# Case Study: Virgin Voyages

2,900% agent growth in 6 months

50

AGENTS IN NOV 2025

1,500

AGENTS BY MAY 2026

2,900%

GROWTH IN 6 MONTHS

## What they didn't do

- Pick agents by SWE-bench score.
- Replace humans wholesale.
- Centralize all “AI strategy” under one team.

## What they did

- Empowered teams to scope **one agent each**.
- Measured operational readiness: time-to-resolution, error rate, CSAT — not benchmarks.
- Used **narrow scope as a constraint**: each agent does exactly one thing.
- Kept human-in-the-loop on every customer-facing and financial decision.

**The lesson:** build twenty narrow agents, measure them, kill the failures, scale the winners. The generalist version of this is the same. Start narrow.

---

# What's Coming Next — June–Aug 2026

---

Confirmed releases, rumors, and watch-list

## Confirmed

- **Gemini 3.5 Pro** full GA June 2026 (announced at I/O). 2M context window.
- **Grok 5** — xAI confirmed in-training for Q2 2026 release.
- **EU AI Factories** — first deployments in June 2026.
- Code with Claude conferences: London (May 20–21, done), Tokyo (June 5–6).
- **Sora API permanent shutdown** — September 24, 2026.

## Rumored (speculative)

- **Claude Sonnet 4.8** — a string referencing it leaked in the Claude Code source map in May (Bun bundler oversight). Anthropic has not announced. Sonnet 4.6 remains current.
- The **Manifold pre-May 24 release market** resolved **NO** at **3%**.
- No firm signal on GPT-5.6 or GPT-6.0.
- **Claude Mythos** GA still restricted to Project Glasswing — cybersecurity research only.

## Watch list

- Anthropic's October 2026 IPO materializing.
- Newsom EO becoming a template for other state governments.
- Composer 2.5 displacing direct Anthropic-API usage among individual developers.

---

# Resources & Communities

---

Where to keep learning after this PDF

## Newsletters

**The Pragmatic Engineer**

[pragmaticengineer.com](https://pragmaticengineer.com)

**Ben's Bites**

[bensbites.com](https://bensbites.com)

**The Neuron**

[theneurondaily.com](https://theneurondaily.com)

**Latent Space**

[latent.space](https://latent.space)

**Stratechery (Ben Thompson)**

[stratechery.com](https://stratechery.com)

## Communities

**r/LocalLLaMA**

[reddit.com/r/LocalLLaMA](https://reddit.com/r/LocalLLaMA)

**r/ClaudeAI**

[reddit.com/r/ClaudeAI](https://reddit.com/r/ClaudeAI)

**r/OpenAI**

[reddit.com/r/OpenAI](https://reddit.com/r/OpenAI)

**r/singularity**

[reddit.com/r/singularity](https://reddit.com/r/singularity)

**Hacker News**

[news.ycombinator.com](https://news.ycombinator.com)

**MCP Discord**

[modelcontextprotocol.io](https://modelcontextprotocol.io)

## Courses

**Outskill AI Generalist (flagship)**

[outskill.com/generalist](https://outskill.com/generalist)

**Maven**

[maven.com](https://maven.com)

**DeepLearning.AI short courses**

[deeplearning.ai](https://deeplearning.ai)

## Documentation

**Claude docs**

[docs.claude.com](https://docs.claude.com)

**OpenAI platform docs**

[platform.openai.com/docs](https://platform.openai.com/docs)

**Google AI for Developers**

[ai.google.dev](https://ai.google.dev)

**Model Context Protocol**

[modelcontextprotocol.io](https://modelcontextprotocol.io)

---

# The Generalist's Commitment

---

Three habits and three traps

## Three habits

### 1. Ship every week.

Not “learn every week.” A bad workflow shipped beats a great workflow described.

### 2. Switch models monthly.

The leader changes every 60 days. Pay for 2–3 frontier subscriptions and rotate. You're calibrating your taste, not picking a favorite.

### 3. Teach what you learn.

Slack channel, lunch & learn, LinkedIn post, blog. Teaching forces you to compress — and compression is where understanding lives.

## Three traps

- **Tweaking prompts instead of shipping projects.** The prompt is the rough draft. The project is the asset.
- **Trusting benchmarks for your work.** SWE-bench doesn't know your codebase. Run your own evals on your own tasks.
- **Treating “AI” as the skill.** The skill is the *workflow*. AI is just the tool inside it.

*“The generalist of 2026 isn't the person who knows the most about AI. It's the person who's shipped the most with it.”*

— Outskill

---

# You don't need permission. You need **a project.**

L1 2–3mo · L2 3–4mo · L3 3–4mo · L4 4–6mo · L5  
4–6mo

Total: 16–23 months, zero to senior.

*“Pick one weekly workflow. Write a one-page PRD. Open Claude, Gemini, or ChatGPT. Ship v1. Iterate next week. That's the whole roadmap — everything else is variations.”*

© Outskill 2026 · Updated May 31, 2026 · Next edition: June 30, 2026