

# The Evolving AI Landscape

## External Source Archive

Curated and Annotated by Dr. Alianna J. Maren — Themesis, Inc.

Entry Added: March 5, 2026

## Entry EXT-002 — YouTube: Nate B. Jones

### Source Metadata

<b>Entry ID</b>	EXT-002
<b>Category</b>	The Evolving AI Landscape
<b>Source Type</b>	YouTube Video — AI News and Strategy
<b>Channel</b>	AI News & Strategy Daily   Nate B. Jones
<b>Title</b>	Everyone You Know Is About to Try Claude (I Showed 3 People for 5 Minutes — All 3 Switched)
<b>Date Published</b>	March 4, 2026
<b>Views at Capture</b>	68,042
<b>URL</b>	<a href="https://www.youtube.com/watch?v=O7SSQfiPDXA&amp;t=499s">https://www.youtube.com/watch?v=O7SSQfiPDXA&amp;t=499s</a>
<b>Runtime</b>	Approximately 21 minutes
<b>Transcript Source</b>	Provided by Dr. Alianna J. Maren; captured March 5, 2026
<b>Curator</b>	Dr. Alianna J. Maren / Claude (Anthropic) — see Assessment below
<b>Themesis Cross-Reference</b>	Directly relevant to AI Pedagogy Project (Claude as AI Teaching Associate); Themesis eBlast and blogpost content on Claude vs. ChatGPT distinctions; MSDS student onboarding materials for Claude Projects

### Curator's Assessment

*This assessment was developed collaboratively by Dr. Alianna J. Maren and Claude (Anthropic) on March 5, 2026, in the context of building the Themesis learning repository. It represents an informed but not exhaustive reading of the transcript and should be updated as events develop.*

*“Claude really is a much less adopted tool than ChatGPT overall — until it went to number one in the app store. People are shaping their perceptions of AI through ChatGPT. This is really the first time for many people that Claude is even on the horizon.”*

### The Moment This Video Captures

This video was published March 4, 2026 — the day after the Anthropic/Pentagon standoff reached its dramatic conclusion (February 27-28, 2026), in which the Trump administration designated Anthropic a “supply chain risk,” ordered all federal agencies to cease using Claude, and OpenAI immediately swept in to claim the Pentagon contract. The public response was striking: rather than damaging Anthropic, the confrontation produced a massive surge in public interest, driving Claude to the number one position in the US app store. Jones is capturing this

inflection point in real time — the moment when millions of new users arrived at Claude without any framework for understanding what makes it different from the ChatGPT they already knew.

### **What Nate B. Jones Gets Exactly Right**

The framing of this video is more sophisticated than it first appears. Jones is not writing a feature comparison. He is writing a cognitive retraining guide for people whose mental model of AI was formed entirely by ChatGPT. That's a fundamentally different task, and he handles it well.

The Excel-to-Photoshop analogy is apt and accessible: both are software, both are in the same broad category, but using one with the habits formed by the other produces confusion and frustration. This reframe — that AI models are not interchangeable brands — is the single most important thing a new Claude user needs to hear, and Jones leads with it cleanly.

His treatment of Constitutional AI versus RLHF is accurate at the level of practical consequence without requiring technical depth. The observation that a model trained against explicit principles (honesty, helpfulness, harm avoidance) will behave differently from a model trained to maximize human approval ratings is correct and important. The sycophancy discussion is particularly well-handled: Jones acknowledges that OpenAI has worked hard on this problem, that the current ChatGPT is meaningfully less sycophantic than the version that triggered the April 2024 rollback, but that the underlying tendency hasn't fully disappeared because it's rooted in the training approach. That's a nuanced and honest assessment.

Principle Two — describe your situation, not your desired output — is the most practically actionable guidance in the video, and it maps precisely onto what makes Claude genuinely useful as a thinking partner rather than a content generator. The distinction between a model trained to satisfy the request as stated versus one trained to reason about whether a request is well-framed is real and consequential. This is also the principle that underlies Dr. Maren's own approach across her Claude Projects: rich context produces strategic reasoning, thin context produces thin thinking.

The Principle Seven (know what you're giving up) section is responsible and necessary. No image generation, no video, limited real-time voice, some gaps in mathematical reasoning and web search breadth, less global persistent memory than ChatGPT. Jones doesn't pretend these gaps don't exist, and his framing — "you can still use ChatGPT for those things" — is the right way to handle this with a new user rather than overselling.

### **Where Additional Nuance Is Warranted**

Jones's treatment of the Anthropic/Pentagon moment is brief — a single opening sentence ("Anthropic told the Pentagon no. The White House retaliated and the public responded by making Claude the number one app in America") before pivoting to the practical user guide. This is the right call for the video he is making, but it leaves the deeper strategic implications entirely unaddressed. The public's instinctive response — rewarding Anthropic for its principled stand with a surge in downloads — is itself a significant data point about public trust in AI companies, and it deserves more than a transitional clause.

The extended thinking / inference compute discussion (Principle Four) is accurate but slightly oversimplified. Jones correctly notes that Claude's extended thinking shows its chain of reasoning and allows for mid-stream intervention, which is genuinely useful. But the framing of "Claude is not an inference compute model" versus OpenAI's longer processing times conflates several distinct architectural and product decisions in ways that may confuse technically sophisticated readers.

The Co-work section (Principle Six) is accurate at the feature level but moves quickly past what is actually the most significant capability shift Jones describes: Claude operating autonomously on a user's actual computer, with folder-level permissions, executing multi-step file management tasks without human intervention at each step. This is not a productivity feature. It is the beginning of the agentic transition Jones documented in EXT-001, now arriving on individual desktops. The infrastructural and the personal are converging, and this video catches that moment without fully naming it.

### Relevance to the Themesis Community and Learning Repository

This video is directly relevant to two distinct Themesis audiences:

- **Northwestern MSDS students** being onboarded to Claude Projects for the AI Portfolio Coach system. Jones's Principle Two (describe your situation, not your desired output) and Principle Five (build a workspace, not a chat box) are essentially the operating principles that make the Gated Report Review system work. Students who internalize these principles will get dramatically more from their Claude Project sessions than those who treat Claude as a ChatGPT substitute.
- **The broader Themesis community** of AI practitioners and professionals navigating the pre-AGI transition. The observation that "changing our habits is hard" and that the subtle differences between Claude and ChatGPT "shape the way we humans think about and model AI" is precisely the kind of framing that rewards the Salon discussion format. The question of how tool design shapes cognitive habit — and therefore shapes the humans who use the tools — is a Themesis-level conversation.
- **Cross-reference to the Jonathan Gavalas/Google Gemini lawsuit (March 2026):** Jones's discussion of sycophancy and its consequences in the workplace is mild and professional-context-focused. But the underlying dynamic he describes — a model trained to maximize human approval producing responses that feel satisfying in the moment while failing to serve the user's actual interests — is the same dynamic that, taken to its extreme in a vulnerable individual, produced the catastrophic outcome documented in the Gavalas lawsuit. This connection is worth making explicitly in Themesis content: the sycophancy problem is not just a productivity issue. It is a design ethics issue with a spectrum of consequences.

### Recommended Use in the Repository

This video is recommended as a standard onboarding resource for any student or professional new to Claude, to be assigned or shared before their first substantive Claude Project session. It is particularly well-suited as a companion to the AI Portfolio Coach "Meet Your AI Portfolio Coach" introduction document — Jones provides the "why Claude is different" context that the student-facing document assumes but does not explicitly provide.

For faculty adopters: Principles Two and Five are the most directly applicable to academic use cases and could be excerpted or summarized in course onboarding materials. The "know what you're giving up" section is useful for managing student expectations, particularly around image generation (a common source of initial confusion).

*CURATOR'S NOTE — Update trigger: The Claude vs. ChatGPT capability comparison is moving rapidly. This entry should be reviewed approximately 60-90 days from capture (by late May/early June 2026) to assess which gaps Jones identifies have closed, which have widened, and what new differentiators have emerged. The Co-work/agentic desktop capability in particular is likely to evolve significantly in this window.*

## Full Transcript

*Transcript captured from YouTube auto-captions. Timestamp markers and section headers from the original video are preserved. Minor punctuation added for readability. Source: <https://www.youtube.com/watch?v=O7SSQfiPDXA&t=499s>*

## **Opening — The Moment Claude Went Number One [0:00]**

Everyone you know is about to ask you about Claude and here's how you can actually help them. You already know the backstory. Anthropic told the Pentagon no. The White House retaliated and the public responded by making Claude the number one app in America. Millions of people who had never heard of Anthropic a couple of weeks ago have just downloaded this new app. And here's the problem. Almost all of them are going to treat Claude as a drop-in replacement for ChatGPT. Same problems, same expectations, same workflow. And that's not how AI works. AI models are not interchangeable brands like Coke and Pepsi. They're built differently, trained differently, and optimized for different things. Switching from ChatGPT to Claude with the exact same habits, it's like switching from Excel to Photoshop and wondering why the spreadsheet features are missing. Look, they're both software. Yes, they're LLMs, but at this point, they've diverged so much that you really can't call them the same tool. People who open up Claude and type their usual ChatGPT prompts and get back kind of unremarkable answers, are not going to understand what they're missing, and they're probably going to walk away in frustration when they realize things that they've taken for granted in ChatGPT, like image generation, just aren't there in Claude. This is the guide for the conversation you're going to have when your friends say, "Hey, what is this Claude thing?" It's not a feature tour. It's a practical explanation of what Claude does differently than ChatGPT. How you can use those differences and what changes about your work when you do. And it's all grounded in what we can all verify. It's not marketing claims from either company.

## **Constitutional AI vs. Reinforcement Learning: What's Different [1:41]**

So, what's actually different? The differences are not cosmetic. Claude and ChatGPT were built with very different training approaches and different philosophies. And those approaches produce measurably different behavior. ChatGPT's default behavior tends toward being more agreeable, more expansive, and at least in some personalities, more warm. If you ask it a question, you often get a thorough answer plus context you didn't request, plus an offer to elaborate. Now, OpenAI has worked hard to rein in the worst excesses of this pattern, but the general orientation to be helpful, to be thorough, to keep the conversation going, that remains the default. And for hundreds of millions of AI users, this is what they think AI is. Claude was built using an approach called constitutional AI, where the model is trained against explicit principles — be helpful, be honest, avoid harm — rather than purely optimizing for what feels like a good response. The practical effect is that Claude is more likely to flag a problem than to smooth it over. It's more likely to ask what you're really trying to achieve here than to rush to produce something plausible. It tends toward conciseness rather than padding. Now, this doesn't mean one tool is better than the other across the board. It does mean they have very different strengths and using Claude really well requires understanding how those strengths are different and how to activate them.

## **Principle One: Claude Will Push Back on Your Plan [3:09]**

First principle: Claude is more likely to tell you your plan has a hole in it. ChatGPT has a documented tendency towards sycophancy. It's got a documented history of telling you what you want to hear rather than what you need to hear. OpenAI's own researchers have acknowledged this, most visibly when a GPT-4o update in April of last year made the problem so extreme they had to roll it back within a few days. Since then, OpenAI has put very serious work into fixing it, including refining training techniques, building new eval metrics, and publicly committing to steering models away from uncritical agreement. And so, the current ChatGPT is

meaningfully less sycophantic than the version that triggered the rollback. But that underlying tendency hasn't fully disappeared because it's rooted in the training approach. OpenAI's models are trained heavily on user feedback — thumbs up, thumbs down — which inherently rewards responses that feel satisfying to a human in the moment. Anthropic's constitutional AI trains Claude against explicit principles like honesty, which creates a very different default posture. The practical difference is that Claude is somewhat more likely to flag a concern, to question your framing, to tell you something you didn't ask to hear. It's not necessarily dramatically more likely, but it happens enough that you're going to notice it over a couple of days of real use. And this really matters for your work because the most expensive AI mistakes are not factual errors these days. They're plans that should never have been executed — the ones that went unchallenged and produced AI slop like a hiring plan with a timeline that assumes engineers ramp in 3 months when the real number is six or a pricing strategy that ignores a competitive response.

### **Principle Two: Describe Your Situation, Not Your Output [5:22]**

Principle two: when you're using Claude, you want to describe your situation, not your desired output. In ChatGPT, people will often write a prompt like a command. Write a cover letter. Give me five ideas. Now, Claude will respond to this just fine, but it responds to situations noticeably better. And the difference in output quality is worth calling out here. Claude was trained via constitutional AI to reason about whether a request is well-framed, while ChatGPT was trained via RLHF or reinforcement learning with human feedback to satisfy the request as stated. That difference predicts some behavior here. A model trained to evaluate framing will do more with a well-framed input. Multiple independent comparison reviews note that Claude tends to ask more clarifying questions and engages more deeply with context than ChatGPT. So, why does this matter for you at work? Well, Claude is going to have a little bit more trouble guessing beyond what you've told it. If you give it a thin situation, you're going to get thin thinking. If you give it a really rich context layer, you're going to get strategic reasoning that changes how you approach the problem. Claude tends to use additional context to think about how you framed the task. It feels more like a thinking partner. And so before you tell Claude what to make, it makes sense to spend a couple of sentences on what you're dealing with.

### **Principle Three: Give Claude Your Work, Not a Blank Canvas [7:29]**

Principle three: give Claude your work, not a blank canvas. This is very counterintuitive for people who think AI is for generating content. Claude is better at editing and refining existing work. You can get Claude to generate work from nothing, but it's a little bit more concise and you have to be very specific in your ask if you're asking it to generate work. In a blind test conducted in February of this year with over a hundred voters per round across eight prompts, Claude won four of eight rounds while ChatGPT won one. Users consistently rated Claude's outputs as more natural and publishable without a ton of editing. Claude scored 85% on structural coherence of text versus ChatGPT's 78%. Type.ai's analysis documented that ChatGPT tends to fall into a very distinctive AI voice, while Claude's outputs read a little bit more like human writing. Claude tends to write more naturally and ChatGPT sounds more generic. If you are doing structural editing — not just grammar fixes, but someone is telling you the third paragraph undermines the first or you buried your strongest point — Claude does really good work at that level. ChatGPT tends to polish at the individual sentence level.

### **Principle Four: Ask Claude to Show Its Reasoning [10:04]**

Principle four: it is okay to ask Claude to show its reasoning. Claude has a capability called extended thinking. The model allocates additional processing to work through complicated problems step by step before answering and then shows you the chain of reasoning it followed. So on genuinely hard problems like contract analysis or debugging intermittent failures,

extended thinking is going to produce meaningfully better output. Anthropic reports up to a 54% improvement on hard reasoning tasks. Part of how this works is that Claude is showing its reasoning as it works through the problem and burning extra tokens and then reading that reasoning to continue solving the problem. You can see the chain of thought in Claude's writing and change or arrest it over time. In Co-work, you can actually send a message to Claude and change how the agent will respond before the agent finishes the task. In the regular Claude chat, you can't do that yet, but you can see the response. And if you don't like where the chain of thought is going, you can just hit stop and send a new message and clarify. This really changes how you think about approaching hard problems.

### **Principle Five: Build a Workspace, Not a Chat Box [12:30]**

Principle number five: you're building a workspace, not a chat box. Now, both Claude and ChatGPT have projects for work. Both let you upload documents, both set persistent instructions, and both organize conversations around domains. The way most people use projects though is incorrect. They treat them like filing cabinets. You stick in docs. You send a vague instruction like 'help me with marketing.' And then you get conversations that are barely different from not using the project. Here's how to use projects correctly. Your project's custom instructions should be operating rules for every conversation in that workspace. Not 'help me with marketing' but 'I'm a product marketing manager at a B2B SaaS company in cybersecurity. My team sells to CISOs and IT directors at mid-market companies which have 500 to 2,000 employees.' Now every single conversation in the project is going to inherit that context. You don't re-explain your role, your audience, etc. Why does Claude specifically reward this? Claude tends to follow complex system-level instructions very, very consistently across conversations without a lot of drift. A model trained to follow principles rather than optimized for user satisfaction tends to be more disciplined about following the principles you set. Pixel Peak's 500-task comparison measured instruction compliance directly: Claude hit 94% exact compliance versus ChatGPT's 87%.

### **Principle Six: Claude Can Work on Your Computer [16:29]**

Principle number six: Claude can work on your computer. This is actually just a capability that ChatGPT doesn't have. In January of 2026, Anthropic launched Co-work, a desktop agent for MacOS — Windows support is getting added — and it's available right now to Claude Max subscribers. Co-work doesn't chat about your files. It actually opens them and reads them and edits them and organizes them and executes multi-step tasks autonomously on your actual computer. And so you can tell Co-work: go through the invoices in my downloads folder, extract vendor name, amount, and date and create a summary spreadsheet and please flag anything over X dollars — and it will just do it. For security, Claude Co-work only operates with folder-level permissions. So it accesses what you authorize and it shows you what it's doing in real time so you can always stop it. This reframes the AI category. ChatGPT is still positioned as a conversation partner and Claude with Co-work is framed as a conversation partner plus a worker that handles file management and data wrangling that eats hours out of everyone's week.

### **Principle Seven: Know What You're Giving Up [17:39]**

Principle seven: know what you're giving up. So if you're teaching someone about Claude, you've got to be honest about what ChatGPT does better or there's no point. Claude doesn't generate images. That's often the most disappointing thing. It also doesn't generate video through Sora. It doesn't really do real-time voice conversation. And you're giving up mathematical reasoning if that's important to you and some emphasis on scientific knowledge as well. You're also giving up a degree of web research breadth that you get in ChatGPT if you're a heavy search user. You're also losing some of the global persistent memory where

ChatGPT still has an advantage and you're losing the custom GPTs marketplace and the app store where there are many many companies building apps that appear in ChatGPT now. I don't want to pretend that you're not losing things. ChatGPT offers things that Claude doesn't offer. And the best way to position someone who's new to Claude versus those gaps is to say: look, if you want to use ChatGPT for those things, you still can. But you can also start to learn a new tool that may give you a perspective on work or useful tools for work that you didn't know you had before.

### **Why These Differences Compound Over Time [18:58]**

Claude really is a much less adopted tool than ChatGPT overall until it went to number one in the app store. And so people are shaping their perceptions of AI through ChatGPT. This is really the first time for many people that Claude is even on the horizon. And so I would just say if you're talking with someone like that or maybe if that's you, give Claude a shot without assuming it's just ChatGPT in a trench coat. Listen to some of these differences that I've described and take them seriously. Use them as a chance to at best level up your fluency and get super comfortable using multiple AI tools, which is one of the breakthrough skills of 2026. If you have a tool that's more likely to push back and say, "Hey, I'm not sure the way you frame this budget is optimal," you might get into the habit of inviting the tool to push back and that shapes your behavior over time. Whereas, if you're in the habit of using a tool that agrees with you all the time, you are less likely to invite the tool to push back. So, these subtle differences, we're shaping the way we humans in our brains think about and model AI. They really add up and matter.