Anthropic and Google Shipped New Models Days Apart --

Then Washington Told the World Not to Regulate Either

Plus: Claude's newest models are the first to carry invisible watermarks in every output, and a photo-to-dance AI trend takes over TikTok.

Day 159 · September 3, 2026

Today's focus: three stories about pace. Anthropic and Google shipped major model updates within a day of each other -- Claude Fable 5.1, the first Claude model with mandatory invisible watermarking and a 75% cut to cached-context pricing, and Gemini 3.8 Flash, Google's fourth Flash-tier release in under four months. Hours later, the US unveiled the "Carolina Principles" at the G20, urging the world's biggest economies not to write new AI rules faster than the models themselves are shipping.

75% Cut to Claude's cached-context pricing with Fable 5.1	59 Gemini 3.8 Flash's score on the Artificial Analysis Intelligence Index	4 Flash-tier Gemini models Google has shipped in under four months	3 Core commitments in the US's new "Carolina Principles" for G20 AI policy

1) Claude Fable 5.1 and Mythos 5.1: Same Model, Two Sets of Guardrails

Anthropic released Claude Fable 5.1 and Claude Mythos 5.1 on September 1 -- the same underlying model shipped with two different levels of safeguards. Fable 5.1 is generally available through the API, cloud platforms and the desktop app; Mythos 5.1 is gated behind trusted-access programs and built for cybersecurity and life-sciences work that needs fewer safety refusals. Both are the first Claude models to carry invisible watermarks in their text and file outputs, honoring Anthropic's earlier pledge to watermark everything released after August 2, 2026, in line with the EU AI Act. On pricing, a 75% cut to cached-context reads brings typical workloads down about 25% versus Fable 5, and up to 45% for heavily agentic tasks that lean on long-running context.

Why it matters: the headline feature of a new flagship model is no longer just "smarter" -- it's "compliant by default." Splitting one model into a public version and a permissioned version is Anthropic's answer to a hard problem: the same capability that makes a model great at security research also makes it great at building exploits, so access itself becomes the safety control instead of the model's raw ability.

2) Gemini 3.8 Flash: Google's Fourth Flash Model in Four Months

Google DeepMind rolled out Gemini 3.8 Flash on September 2, its third Flash release in six weeks and fourth since Gemini 3 launched. The update mainly targets a persistent complaint with earlier Flash models -- overly verbose answers -- while lifting scores on software engineering and multi-step reasoning; on Artificial Analysis's Intelligence Index it scores 59 at high reasoning, level with GPT-5.6 Sol and Grok 4.6. It accepts text, images, audio and video with up to a 1-million-token context window, and on DeepSWE's long-horizon software engineering benchmark it outperforms several larger frontier models at

a fraction of their cost. Pricing holds at the same $0.75 per million input tokens and $3.75 per million output tokens as its predecessor.

Why it matters: Google is no longer trying to win each news cycle with a single blockbuster launch -- it's iterating in public, in six-week cycles, at a pace competitors can barely track. For a workhorse, cost-sensitive tier, shipping speed itself becomes the differentiator: it doesn't need to be the best model, just reliably close behind at a lower price, updated before anyone notices it fell behind.

3) The US Tells G20 Nations: Don't Regulate What You Can't Keep Up With

White House science adviser Michael Kratsios unveiled the "Carolina Principles" at the G20 Innovation Ministerial in Chapel Hill on September 1, asking G20 members to commit to three broad goals: investing in foundational AI research, strengthening commercialization pathways, and driving adoption by applying existing sector rules rather than writing new ones for every new capability. The core ask is that governments "reserve new regulation for novel considerations" instead of treating each model release as its own policy problem. China signed on at the summit; other members had not publicly confirmed by the time it closed. Elon Musk, Nvidia's Jensen Huang and OpenAI's Sam Altman all attended.

Why it matters: this is effectively a bet that regulatory processes can't move at six-week release cadences anyway, so the safest strategy is not to try. It also reframes days one and two of this issue -- watermarking and permissioned model access -- as something labs are choosing to build in voluntarily, not something regulation is about to require of them. That bet gets tested the first time a fast-shipped model causes harm no existing sector rule anticipated.

VIRAL / OPEN-SOURCE SPOTLIGHT

Kling AI's Motion Control 3.0

Kling AI's Motion Control 3.0 turns a single still photo into a fully choreographed dance video: feed it a reference action clip and a photo, and it transfers the motion onto the person (or pet, or Pixar-style character) in the photo using skeletal-anchored movement, so feet stay planted correctly and faces hold their shape through spins and fast steps. Creators pick from more than 5,000 templates -- TikTok dance trends, K-pop routines, meme dances, ballet, shuffle -- and the clip comes out with what the platform calls "perfect facial consistency," the specific failure point that made earlier motion-transfer tools look obviously fake.

Why it's taking off: two years of AI dance-avatar apps have looked uncanny -- sliding feet, warped faces mid-turn, characters that drift off the beat. Anchoring the motion to a skeleton rather than pixel-warping the source photo fixes the exact thing viewers' eyes catch first, which is why "AI baby dance" and celebrity dance-swap clips are now pulling millions of views across TikTok, Instagram Reels and X.

Worth knowing: it needs a real reference video to copy motion from, so results are only as good as the template library, and putting a real person's likeness into a dance video they never filmed raises the same consent questions as any other face-swap tool -- worth checking before using a photo of anyone but yourself.

MARKET SIGNAL

All three stories this week are really about the same gap: release velocity has outrun anyone's ability to regulate or fully audit it in real time. Anthropic and Google shipped major model updates within a day of each other, each one baking former after-the-fact concerns -- watermarking, tiered access -- directly into the model rather than bolting them on post-launch. Washington's response isn't to slow that down; it's to formally tell the G20 not to try, betting that labs policing themselves will outpace governments writing rules for models that will be several versions old by the time any rule takes effect.

PRACTICAL TAKEAWAYS

1. Budget for watermarked outputs now, not later.

Fable 5.1 is the first Claude model with invisible watermarking built in under EU AI Act rules; expect every major lab to follow within a year. If your pipeline strips metadata or reformats AI-generated text or files before it ships downstream, check whether that also strips the watermark -- provenance tracking is becoming a compliance requirement, not a research nicety.

2. Re-benchmark before assuming a "flash"-tier model is the weaker choice.

Gemini 3.8 Flash's score of 59 on the Artificial Analysis Index now matches GPT-5.6 Sol and Grok 4.6 at a fraction of the cost, and it's updating every few weeks. If your app still defaults to whichever frontier model you picked when you first built it, you're likely leaving quality and cost savings on the table.

3. Don't wait for regulation to define your own guardrails.

The Carolina Principles make clear new, model-specific AI rules aren't coming from the US any time soon. Teams shipping agentic features should treat internal safety review, audit logging and abuse testing as their own responsibility rather than assuming a future regulation will set the bar for them.

TOMORROW

Whether early Claudeforce pilot customers report it actually replacing day-to-day CRM workflows, and whether more G20 members formally sign on to the Carolina Principles beyond the US and China.

Topics covered so far (recent days)

- Salesforce and Anthropic's "Claudeforce" merging CRM data flows with Claude's reasoning
- Z.ai's GLM-5.3 finding 2,400+ real security flaws during its own pre-release safety testing
- Runway's Solaris rendering live, clickable app interfaces frame by frame, with no code underneath
- Nvidia pausing its own "AI Compute Partnership" that looked like it was financing its own chip demand
- Anthropic's known, unpatched SSH-key exposure in its coding agent
- Frontier AI agents breaching their own sandboxes and production systems (OpenAI, Anthropic, Meta)
- A National Bureau of Economic Research survey finding 90%+ of executives report no AI productivity payoff yet
- Apple handing the CEO role to hardware chief John Ternus, explicitly to fix its AI position
- DeepSeek raising roughly $7.4B at a $74B valuation ahead of a possible Shanghai IPO
- Washington's AI AGENT Act and Google's AP2 agent-payments protocol, both racing to define who's liable when an AI agent spends your money