

New Claude Features → EdgePointe Services — Research Report (2026-07-18)

For: Bob, Niven, and Paul

Recurring Research Report · definition: `Recurring-Claude-Features-Research-Report.md` · web-researched, every claim cited · prior run: `claude-features-2026-07-16.md`

Executive summary

Two days on from the baseline report, the picture is stable but three items sharpened enough to act on. **The finalized MCP specification lands July 28, 2026** — a stateless core plus two official extensions (**MCP Apps**, which lets a tool return a live HTML UI inside the chat, and **Tasks**, for long-running work) and OAuth-aligned auth (blog.modelcontextprotocol.io, [ForkPoint](#)). That makes the "Vantage MCP server" from the last report a firmer bet and adds a new angle: a Vantage MCP tool could ship its own booking/dashboard UI straight into any MCP host. **Claude Code Artifacts** (June 18, now beta on Team/Enterprise) turns a coding session into a live, shareable claude.ai page that updates in place — a ready-made client deliverable for QBRs, site audits, and WebDesigner build walkthroughs (claude.com/blog). And **Claude Code auto mode now runs on Pro**, with `/usage` breaking spend down by skill/subagent/plugin/MCP and subagents running in the background by default — direct margin visibility for EdgePointe's own delivery (code.claude.com/docs/en/whats-new). The Sonnet 5 intro-pricing clock (\$2/\$10 through **Aug 31**) is the single most time-boxed decision. Skills marketplaces exploded to 8+ registries — but security audits found prompt injection in roughly a third of public skills, which turns "governed private skill marketplace" from a nice-to-have into a sellable Managed-IT service.

What changed since 2026-07-16 (the diff)

- **NEW — MCP final spec dated July 28, 2026** with stateless architecture, **MCP Apps** (tool-rendered HTML UIs in-chat) and **Tasks** extensions, OAuth/OIDC alignment. The prior report treated MCP as "donated/standardized"; it is now a concrete, dated release with a UI story. (blog.modelcontextprotocol.io, thenewstack.io)
- **NEW — Claude Code Artifacts** (shareable live pages, June 18) surfaced as a client-deliverable channel not covered before. (claude.com/blog)
- **UPDATED — Claude Code auto mode reaches Pro accounts**; `/usage` now itemizes limits by skill/subagent/plugin/MCP; subagents run in background by default. (code.claude.com/docs/en/whats-new, anthropic.com/engineering)
- **UPDATED — Memory** moved from a single daily summary to **individual categorized entries** with per-project memory spaces, free on all plans; "Memory Files" is the next step. (support.claude.com release notes, testingcatalog)

- **NEW context — Skills marketplace security risk quantified** (Snyk ToxicSkills: prompt injection in ~36% of skills tested; 8-registry ecosystem). Reframes EdgePointe's skill library as a governance product. (agentman.ai, agensi.io)
- **NEW context — Claude for Government (FedRAMP guide) + Claude for Teachers + Azure AI Foundry GA** — competitive/vertical moves, minor direct EdgePointe impact. (anthropic.com/news, [chalkbeat](https://chalkbeat.com))
- **Unchanged** from 07-16: model lineup (Opus 4.8, Sonnet 5, Haiku 4.5, Fable/Mythos 5), Managed Agents scheduled deployments, Cowork, Claude in Chrome, Claude for M365, no Anthropic voice API. See prior report for full detail.

Feature map (what → why → how EdgePointe applies it)

1. MCP final spec (July 28) + MCP Apps + Tasks — blog.modelcontextprotocol.io, [ForkPoint](https://forkpoint.com)

→ *What*: biggest revision since launch — stateless servers behind ordinary load balancers, **MCP Apps** return sandboxed HTML UIs inside the chat, **Tasks** handles long-running work, auth aligns with OAuth 2.0/OIDC. Backed by Anthropic, OpenAI, Google, Microsoft; ~19,800 servers indexed.

→ *Why it matters*: an MCP-enabled Vantage becomes operable from whatever assistant a client already uses, and MCP Apps means a Vantage tool can render its own booking/lead/dashboard UI in-chat — not just return text.

→ *EdgePointe apply*: build a **Vantage MCP server** on the existing Workers/D1 stack (bookings, leads, reviews, analytics, publish-content as tools) targeting the July-28 shape; ship a booking widget as an MCP App. Stateless core fits Workers perfectly. **[Vantage] [New offering] — Impact: High · Effort: Medium.**

2. Claude Code Artifacts — shareable live pages (beta, Team/Enterprise) — claude.com/blog, code.claude.com/docs

→ *What*: turns a Claude Code session into a private, live, shareable claude.ai HTML page that updates in real time from the repo/data sources, with version history/rollback, wrapped in a strict CSP (no external network calls).

→ *Why it matters*: a polished, link-shareable deliverable falls out of work EdgePointe already does — no separate build step.

→ *EdgePointe apply*: publish **QBR dashboards, Vantage site-audit walkthroughs, and WebDesigner build progress** as artifacts for Bob/Niven and clients; use for the clone-polish A/B audit write-up. **[Client services] [Internal] — Impact: Medium · Effort: Low.**

3. Claude Code auto mode on Pro + `/usage` breakdown + background subagents — code.claude.com/docs/en/whats-new, anthropic.com/engineering

→ *What*: auto mode (permission prompts replaced by background safety checks, classifier defaults to Sonnet 5 for external sessions) now on Pro; `/usage` itemizes plan limits by skill, subagent, plugin, and MCP server; subagents run in the background by default and notify on completion.

→ *Why it matters*: faster hands-off delivery plus per-component cost attribution — directly informs which skills/subagents are worth keeping.

→ *EdgePointe apply*: run WebDesigner/Vantage builds in auto mode; use `/usage` to attribute AI cost per

skill and prune expensive ones; background subagents keep the Mirror-Score capture/build/QA loop parallel. **[Internal]** — **Impact: Medium · Effort: Low.**

4. Claude Opus 4.8 + dynamic workflows + fast mode — anthropic.com/news/claude-opus-4-8, platform.claude.com/whats-new

→ *What:* flagship GA at \$5/\$25 per MTok; **4x less likely to pass flawed code without comment** than 4.7; **dynamic workflows** in Claude Code plan work, run parallel sub-agents, verify outputs, report back; effort controls in claude.ai/Cowork; fast mode 2.5x faster and ~3x cheaper than prior Opus fast; mid-conversation system messages.

→ *Why it matters:* the honesty gain reduces silent bad-code risk in delivery; dynamic workflows are the built-in version of EdgePointe's multi-agent build loop.

→ *EdgePointe apply:* default engine for internal builds; lean on dynamic workflows for genesis/WebDesigner passes; fast mode for live sales/demo moments. **[Internal]** **[Vantage build-time]** — **Impact: High · Effort: Low.**

5. Claude Sonnet 5 — intro pricing ends Aug 31 — anthropic.com/news/claude-sonnet-5, techcrunch.com/2024/08/29/claude-sonnet-5/

→ *What:* most agentic Sonnet yet, near-Opus-4.8 on many tasks; default on Free/Pro; **\$2/\$10 through Aug 31, then \$3/\$15**; 1M context, 128K output, adaptive thinking on. Breaking: manual extended thinking → 400, non-default sampling params → 400.

→ *Why it matters:* the sweet-spot runtime model for Vantage AI, and the pricing window is a hard deadline for a margin decision.

→ *EdgePointe apply:* benchmark **receptionist chat**, **Ask Vantage copilot**, **Reach commentary**, and **voice-mimicry blog drafting** (currently Workers AI) on Sonnet 5 now; lock the per-tier model matrix before Sept 1. **[Vantage]** **[Client services]** — **Impact: High · Effort: Medium.**

6. Memory: categorized entries + per-project spaces (free, all plans) — support.claude.com/release-notes, testingcatalog.com/

→ *What:* memory is now individual categorized entries Claude reads/updates live (replacing the 24h summary); each project keeps a separate memory space; "Memory Files" (topic-organized notes) is coming; Managed Agents have built-in memory + "Dreaming."

→ *Why it matters:* cleaner per-client context isolation for EdgePointe's many tenants/projects.

→ *EdgePointe apply:* one Claude project per client with isolated memory; mirror the pattern for any Managed-Agent-based Vantage blog/research service. **[Internal]** **[Vantage]** — **Impact: Medium · Effort: Low.**

7. Higher-resolution computer-use vision (2,576px) — digitalappliedmatrix.com/, deck.co/guide

→ *What:* Sonnet 5 and Opus 4.8 accept screenshots up to 2,576px (older models cap at 1,568px); OSWorld ~72.5%; per-screenshot prompt-injection classifiers.

→ *Why it matters:* far more reliable reading of dense admin UIs and dashboards.

→ *EdgePointe apply:* stronger **visual/functional QA** of Vantage tenant sites (booking/lead flows — the class of bug behind the honeypot/autofill issue) and portal-based **client migrations** (17Hats/GoDaddy/WP with no API). **[Internal]** **[Client services]** — **Impact: Medium · Effort: Low.**

8. Agent Skills marketplaces + security risk — agentman.ai, agensi.io

→ *What:* SKILL.md is an open standard across 8+ registries and 40+ compatible products; hot-reload mid-session (Claude Code 2.1). But a Snypk/ToxicSkills audit found prompt injection in ~36% of tested skills; curated skills lift pass rates ~16 pts.

→ *Why it matters:* the risk data is the sales pitch — SMBs need someone to vet and govern the skills their teams install.

→ *EdgePointe apply:* package EdgePointe's own skills (hero-cloner, clone-polish, WebDesigner, roles, acquisition-viability) into a **governed private marketplace**; sell "skill vetting + private marketplace management" as a Managed-IT service. **[Internal] [New offering] — Impact: Medium · Effort: Low–Medium.**

9. Cowork enterprise plugins + admin controls — claude.com/blog, [constellationr](https://constellationr.com)

→ *What:* org-specific marketplaces, **private GitHub repos as plugin sources** (private beta), per-user provisioning, auto-install, OpenTelemetry usage/cost/tool telemetry; 9+ knowledge-work and financial-services plugin packs plus new connectors.

→ *Why it matters:* the reference architecture for "we govern your company's AI," billed flat monthly — exactly EdgePointe's model.

→ *EdgePointe apply:* deploy/govern Cowork + plugins for Managed-IT clients; use OpenTelemetry for QBR reporting; host EdgePointe skills from a private GitHub plugin source. **[Managed IT] [New offering] — Impact: High · Effort: Medium.**

10. Managed Agents (scheduled deployments, webhooks, memory) — platform.claude.com/managed-agents

→ *What:* server-hosted agent harness with cron scheduling, webhooks, memory stores, multiagent orchestration; not ZDR/HIPAA-eligible.

→ *Why it matters:* closes two named Vantage roadmap gaps with zero scheduler infrastructure.

→ *EdgePointe apply:* build the **done-for-you blog service** and the **weekly competitive-research report** as scheduled Managed Agents that webhook into Vantage's publish API / Cockpit. **[Vantage] [New offering] — Impact: High · Effort: Medium.** (*Carried from 07-16 — still the biggest single unlock.*)

Cost & risk flags

- **Sonnet 5 intro pricing ends Aug 31** (\$2/\$10 → \$3/\$15). Combined with the ~30% tokenizer inflation on the new models, effective Vantage runtime cost rises materially Sept 1 — decide the per-tier model matrix in the intro window. (anthropic.com/news/claude-sonnet-5)
- **Skill supply-chain risk:** ~36% of public skills tested carried prompt injection. Never install unvetted skills into EdgePointe or client environments; run a scan gate and prefer curated/private sources. (agentman.ai)
- **New-model breaking changes** (unchanged, still bites integrations): non-default `temperature / top_p / top_k` → 400; manual extended-thinking budgets removed on Sonnet 5. Audit any EdgePointe code that sets sampling params. ([platform.claude.com/whats-new Sonnet 5](https://platform.claude.com/whats-new/Sonnet-5))

- **Managed Agents / Artifacts are beta and not ZDR/HIPAA-eligible** — fine for marketing/content and internal dashboards; keep sensitive client data out. (code.claude.com/docs/en/artifacts)
- **Competitive:** auto mode on Pro + capable Sonnet 5 put strong agents directly in SMB hands — expect DIY erosion at Vantage's low end; counter by selling outcomes (managed blog, managed agents, governance), not tools. (code.claude.com/docs/en/whats-new)

Top 5 "act now" (impact vs effort)

1. **Benchmark Sonnet 5 for all Vantage runtime AI before Aug 31.** Low effort, hard deadline: run receptionist/copilot/Reach/blog prompts at \$2/\$10 vs Haiku 4.5, lock the per-tier model matrix. *(Impact High · Effort Low)*
2. **Prototype a Vantage MCP server against the July-28 spec, with a booking MCP App.** The finalized stateless spec + MCP Apps make this a durable differentiator no template-site competitor offers. *(Impact High · Effort Medium)*
3. **Ship the blog-writer + weekly research report on Managed Agents scheduled deployments.** Closes two roadmap gaps with zero scheduler infra; charge into Growth Plus/Concierge. *(Impact High · Effort Medium)*
4. **Adopt Claude Code Artifacts as EdgePointe's client-deliverable format** for QBRs, site audits, and build walkthroughs — near-zero effort, immediately client-visible. *(Impact Medium · Effort Low)*
5. **Stand up a governed private skill marketplace and sell skill-vetting as a Managed-IT service.** The 36%-injection data is the pitch; EdgePointe's skills already exist. *(Impact Medium · Effort Low–Medium)*

Sources

- <https://www.anthropic.com/news> — Anthropic Newsroom (index, July 2026)
- <https://www.anthropic.com/news/claude-opus-4-8> — Opus 4.8 launch
- <https://platform.claude.com/docs/en/about-claude/models/whats-new-claude-4-8> — What's new in Opus 4.8
- <https://www.anthropic.com/news/claude-sonnet-5> — Sonnet 5 launch
- <https://platform.claude.com/docs/en/about-claude/models/whats-new-sonnet-5> — What's new in Sonnet 5
- <https://techcrunch.com/2026/06/30/anthropic-launches-claude-sonnet-5-as-a-cheaper-way-to-run-agents/> — Sonnet 5 coverage
- <https://blog.modelcontextprotocol.io/> — MCP blog (final spec July 28, 2026)
- <https://agenticstorefront.com/blog/mcp-2026-release-candidate-explained/> — MCP 2026 release-candidate business guide
- <https://thenewstack.io/model-context-protocol-roadmap-2026/> — MCP production roadmap 2026
- <https://claude.com/blog/artifacts-in-claude-code> — Claude Code Artifacts announcement
- <https://code.claude.com/docs/en/artifacts> — Share session output as artifacts (docs)

- <https://code.claude.com/docs/en/whats-new> — Claude Code what's new (auto mode on Pro, / usage, background subagents)
- <https://www.anthropic.com/engineering/claude-code-auto-mode> — How Claude Code auto mode works
- <https://support.claude.com/en/articles/12138966-release-notes> — Claude release notes (memory entries, Cowork mobile)
- <https://www.testingcatalog.com/anthropic-plans-claude-memory-update-with-new-memory-files/> — Memory Files update
- <https://www.digitalapplied.com/blog/computer-use-agents-2026-claude-openai-gemini-matrix> — Computer-use agents 2026 (vision resolution)
- <https://deck.co/blog/claude-computer-use-developer-guide> — Claude Computer Use developer guide
- <https://agentman.ai/blog/agent-skills-ecosystem-report-2026> — Agent Skills ecosystem report (security)
- <https://www.agensi.io/learn/best-ai-agent-skills-marketplaces-2026> — Skills marketplaces compared
- <https://claude.com/blog/cowork-plugins-across-enterprise> — Cowork + plugins across enterprise
- <https://www.constellationr.com/insights/news/anthropic-expands-cowork-plugins-across-enterprise-functions> — Cowork enterprise plugins coverage
- <https://platform.claude.com/docs/en/managed-agents/overview> — Claude Managed Agents docs
- <https://www.chalkbeat.org/2026/07/14/anthropic-launches-claude-for-teachers-as-ai-companies-battle-for-classrooms/> — Claude for Teachers