

The AI Buyer's Brief

How to choose an AI model and plan — tokens, pricing, and fit — without the hype.

Five serious platforms now compete for the same \$20–\$200 of your monthly budget: ChatGPT, Claude, Google Gemini, Perplexity, and Manus. None is best at everything. The right choice depends on the work you repeat weekly — not the demo that impressed you once. This brief gives you the pricing, the token economics, and a neutral framework to decide, test, and re-decide.

THE SPENDING LADDER — FOUR RATIONAL STOPPING POINTS

\$0

EXPLORE

Free ChatGPT, Claude, Gemini, or Perplexity. Enough for occasional questions, learning what AI is good at, and forming a weekly habit before spending anything.

\$8–\$20

COMMIT

One paid assistant (or a budget tier like ChatGPT Go at \$8 or Google AI Plus at \$7.99). Higher limits, stronger models, files, projects, memory.

\$40

PAIR

Two complementary \$20 plans — different strengths, not more of the same. Often more strategic range than a single \$100 plan.

\$100–\$200

CAPACITY

Buy only when one ecosystem clearly fits and you repeatedly hit its limits doing valuable work. Pay for proven constraint, not imagined potential.

THE DEFAULT MOVE

Start with one \$20 plan for 30 days. Upgrade only when capacity constrains real output.

Do not pay \$100 to avoid making a \$20 decision — and do not stay free to avoid making any decision at all.

THREE QUESTIONS THAT SETTLE 90% OF THIS DECISION

1. What do I repeat weekly?

AI plans reward repeatable workflows — research, writing, analysis, building, scheduling. A plan that serves your top three recurring tasks beats one that dazzles on tasks you do twice a year.

2. What can't I do reliably on free?

Free tiers in 2026 are genuinely capable. Pay only for the specific gap: higher limits, stronger models, longer documents, agents, or scheduled work you actually hit walls on.

3. What does my data require?

Consumer plans may train on your content by default (opt-outs vary); business tiers typically don't. If you handle client data, check training, retention, and connector policies before you subscribe.

Tokens: how AI is actually priced

Subscriptions hide it; APIs expose it. Either way, every plan limit and every invoice traces back to tokens.

A token ≈ 3–4 characters

Roughly 0.75 English words. A 1,000-word document is about 1,300 tokens. “Unbelievable” may be 3 tokens; “cat” is 1.

Context window

The model’s working memory per conversation, now commonly 200K–1M tokens (≈ 150–750K words). Long chats and big files consume it — and consume your usage limits faster.

Input vs. output

You pay separately for what you send (input) and what the model writes (output). Output costs 3–6× more — generating text is computationally harder than reading it.

Caching & batch

APIs discount repeated prompts (cached input, often –90%) and non-urgent jobs (batch, typically –50%). Providers pass these savings to developers, not chat subscribers.

API RATES — WHAT RAW MODEL ACCESS COSTS (USD PER 1M TOKENS)

PROVIDER	MODEL	INPUT	OUTPUT	NOTES
OpenAI	GPT-5.6 Sol / GPT-5.5 (flagship)	\$5.00	\$30.00	1M context; cached input –90%; batch –50%
	GPT-5.6 Terra / GPT-5.4 (mid)	\$2.50	\$15.00	Half flagship cost; strong for routine work
	GPT-5.4 nano (economy)	\$0.20	\$1.25	Classification, extraction, high-volume tasks
Anthropic	Claude Fable 5 (frontier)	\$10.00	\$50.00	Top-tier reasoning; priced above Opus
	Claude Opus 4.8 (flagship)	\$5.00	\$25.00	Matches GPT flagship input; cheaper output
	Claude Sonnet (mid)	\$2–\$3	\$10–\$15	Intro \$2/\$10 through Aug 31, 2026, then \$3/\$15
	Claude Haiku 4.5 (economy)	\$1.00	\$5.00	Fast, cheap; everyday and agentic subtasks
Google	Gemini 3.1 Pro (flagship)	\$2.00	\$12.00	Price aggressor at the frontier; 1M context
	Gemini 3.5 Flash (mid)	\$1.50	\$9.00	Launched May 2026; beats 3.1 Pro on coding
Perplexity	Sonar / Sonar Pro (search)	\$1–\$3	\$1–\$15	Plus per-1K-request search fees (\$5–\$14)

PERSPECTIVE — WHAT \$1 OF API BUYS

- On an economy model (\$0.20/\$1.25): roughly 600,000+ words of drafting — several novels’ worth.
- On a flagship (\$5/\$25–\$30): about 25,000–30,000 words of high-quality output — a serious report or two.
- Why it matters: a \$20 subscription is a flat-rate bundle of tokens, tools, and interface. Heavy, narrow, repeatable workflows can be cheaper via API; broad daily use is almost always cheaper on a subscription.

The \$20 tier, compared five ways

All five cost ≈\$20/month. They are not interchangeable — each wins when its surrounding workflow is your workflow.

ChatGPT Plus

\$20/mo

The generalist operating system.

CHOOSE IT FOR

Broadest toolset: chat, images, projects, custom GPTs, scheduled tasks, Codex coding, strong mobile continuity. GPT-5.6 flagship access included.

KNOW GOING IN

Model routing and per-feature limits vary; the breadth can feel complex; Free/Go tiers now carry ads (Plus removes them).

Claude Pro

\$20/mo · \$200/yr

The long-form thinking partner.

CHOOSE IT FOR

Nuanced writing, long documents, synthesis, complex instructions; Projects, Research, Claude Code and Cowork for agentic file/browser work.

KNOW GOING IN

One shared usage pool — long chats, big files, and heavy tools drain it quickly; no simple per-task cap to plan around.

Google AI Pro

\$19.99/mo

The Google-ecosystem multiplier.

CHOOSE IT FOR

Gemini 3.1 Pro, Deep Research, NotebookLM, Veo video, 1M context, 2 TB storage, Workspace tie-ins — plus a YouTube Premium Lite bundle.

KNOW GOING IN

Value concentrates inside Google's ecosystem; if your work isn't Gmail/Docs/Drive-centered, much of the bundle sits idle.

Perplexity Pro

\$20/mo · \$200/yr

The cited-research engine.

CHOOSE IT FOR

Unlimited Pro Search with citations, ~20 Deep Research/day, a model picker across GPT, Claude, and Gemini, and the free Comet browser.

KNOW GOING IN

An answer engine, not a workflow engine — weaker as your only system for long projects, personal context, or complex creation.

Manus Pro

from \$20/mo

The delegated-task agent.

CHOOSE IT FOR

Autonomous execution: research runs, slide and website builds, browser operation, scheduled tasks. 4,000 credits/mo at \$20; 8,000 at \$40.

KNOW GOING IN

Credit burn is opaque — one deep task can consume 200–1,000+ credits; costs are hard to predict. A specialist, not a daily driver.

The complete price map — \$0 to \$200

Individual plans, July 2026. Team/enterprise seats and annual discounts noted where they change the math.

CHATGPT (OpenAI)

Free	\$0	GPT-5.5, limited use; ads (US)
Go	\$8	Higher caps, core model; may show ads
Plus	\$20	GPT-5.6 flagship, full toolset; no ads
Pro	\$100	≈ 5× Plus usage (launched Apr 2026)
Pro	\$200	≈ 20× usage; GPT-5.5 Pro reasoning mode
Business	\$25/seat	\$20 annual; no training on business data

CLAUDE (Anthropic)

Free	\$0	Expanded Feb 2026: Projects, connectors
Pro	\$20	\$200/yr; full capabilities, Claude Code
Max 5x	\$100	5× Pro usage per session; higher outputs
Max 20x	\$200	20× Pro; for all-day heavy use
Team Std	\$25/seat	\$20 annual; shared workspaces
Team Prem	\$125/seat	\$100 annual; power seats, Claude Code

GOOGLE GEMINI

Free	\$0	Daily limits; basic Workspace tie-ins
AI Plus	\$7.99	More 3.1 Pro access; 200 GB storage
AI Pro	\$19.99	3.1 Pro, Deep Research, Veo, 2 TB, YT Lite
AI Ultra	\$99.99	≈ 5× Pro; 20 TB (cut from \$249.99 at I/O)
AI Ultra 20x	\$199.99	Top tier; ≈ 20× Pro limits
Workspace	bundled	Gemini now included in business tiers

PERPLEXITY

Free	\$0	Unlimited basic search; ~5 Pro searches/day
Education	\$10	Full Pro for verified students
Pro	\$20	\$200/yr; unlimited Pro Search, model choice
Max	\$200	\$2,000/yr; Model Council; 10K credits
Ent. Pro	\$40/seat	Governance, SSO, data isolation
Ent. Max	\$325/seat	Heavy agentic + research capacity

MANUS

Free	\$0	300 credits/day; Lite agent; 1 concurrent task
Pro	\$20	4,000 credits/mo; full agent models
Pro+	\$40	8,000 credits/mo; 7-day trial
Pro Max	\$200	40,000 credits/mo; heavy delegation
Team	\$20/seat+	Pooled credits; 2-seat minimum

Annual billing saves ~17% on all paid tiers

ONE \$100+ PLAN VS. TWO \$20 PLANS

Go premium when:

you repeatedly hit one platform’s limits, your valuable work lives in that ecosystem, and the extra capacity clearly returns more than the \$60–\$80/mo difference.

Pair two \$20s when:

you want two different strengths (e.g., one for research, one for creation), rarely hit limits, and value a second opinion plus resilience if one service degrades.

Which plan fits which person

Start from the work you repeat, not the brand you’ve heard most about.

BY DOMINANT WORK TYPE

Create + organize daily	ChatGPT Plus	Broadest everyday toolkit; scheduling, images, custom assistants
Write, think, synthesize	Claude Pro	Long-form reasoning, long documents, careful instructions
Live in Google Workspace	Google AI Pro	Gmail/Docs/Drive integration, NotebookLM, storage bundle
Research with sources	Perplexity Pro	Cited, current-web answers; multi-model access under one fee
Delegate whole tasks	Manus Pro	Agentic execution — slides, sites, browser work (watch credits)
Narrow repeatable pipeline	Pay-as-you-go API	A few dollars of tokens can beat any subscription

STAY FREE (FOR NOW) IF...

- You use AI a few times a week and can tolerate limits, resets, or reduced tool access.
- You mostly ask questions that free tiers answer well — explanations, quick drafts, sourced lookups.
- You have not yet formed a weekly habit. Do not pay for “someday”; buy for an active workflow.

THE 30-DAY TRIAL PROTOCOL

1. Pick your five highest-value weekly tasks before you subscribe.
2. Run the same five tasks in each contender — identical prompts, real stakes.
3. Score outcomes, not vibes: time saved, quality gained, errors avoided.
4. Set a calendar reminder for day 25 to keep, switch, or cancel deliberately.

THE PRIVACY CHECK (BEFORE CLIENT DATA)

- Training: does the plan train on your content by default? Is opt-out available — and did you actually toggle it?
- Retention: how long are chats and files stored; can you delete or export them?
- Connectors: which accounts (email, drive, calendar) has the tool been granted, and what can it do there?
- Business tiers: typically exclude training by default — worth the seat price for regulated or client work.

The monthly value review

Subscriptions drift. Re-decide on evidence — not attachment to a brand or fear of switching.

KEEP PAYING IF...

- You use it weekly for real deliverables.
- You rely on paid-only capacity: models, limits, agents, or tools.
- It creates at least ~\$20/mo of time, quality, or opportunity value — one saved hour usually clears the bar.

DOWNGRADE OR CANCEL IF...

- You rarely open it, or stay comfortably inside free limits.
- It duplicates another subscription's role in your stack.
- You mostly ask questions free tools answer well — capacity you don't use is a donation, not an investment.

COPY / PASTE: "CHOOSE MY BEST AI PLAN" — RUN IT IN ANY ASSISTANT

Act as my independent, vendor-neutral AI subscription advisor. Using what you know about my work, business model, recurring tasks, privacy needs, and technical comfort: (1) identify my five highest-value weekly AI workflows; (2) compare staying free, one \$20 plan, two complementary \$20 plans, and one \$100+ plan across ChatGPT, Claude, Google Gemini, Perplexity, and Manus — only where each is genuinely relevant; (3) for each workflow, assign the best-fit tool and explain why, with estimated monthly cost, usage pressure, time saved, and overlap; (4) do not recommend premium capacity for work a free or low-cost model handles. End with: BEST FIT, SECOND CHOICE, DO NOT BUY, a 30-DAY TEST plan, KEEP/CANCEL METRICS, and the single plan I should start today.

SIXTY-SECOND GLOSSARY

Token — the billing unit; $\approx$ 0.75 words	Context window — working memory per chat	Inference — the model generating output
Rate limit — usage cap per hour/day/session	Agent — AI that executes multi-step tasks	API — pay-per-token access for builders

VERIFY LIVE BEFORE YOU BUY — PRICES CHANGE OFTEN

chatgpt.com/pricing · claude.com/pricing · gemini.google/subscriptions · perplexity.ai/pro · manus.im/pricing

All figures reflect published U.S. individual pricing verified July 18, 2026 from official pricing pages and current coverage. This brief is educational and vendor-neutral; it is not financial advice, and no platform sponsored or influenced it.