

QUICK GUIDE

10 Money-Saving Marketing Moves with AI

Quick-reference guide to cutting your bill 25–60%

✓ THREE TESTED THIS MONTH

Moves 1–3

1 Grok 4.7

Cheaper reasoning

Grok 4.7 handles reasoning work like copy variations and first-pass research at a fraction of Fable's price. Route those tasks to it and save premium models for the final pass.

2 Jev

High-volume work

Jev is built for sorting and labeling. Send it lead scoring, tagging, and sentiment checks: your highest-volume work, and the kind that rarely needs a premium model.

3 Cancel Expensive Tools

Lower stack cost

Per-seat tools like Asana and Cursor now have open-source replacements. Switching cuts a recurring cost that grows with every new hire.

☀ AUTOMATIONS & INFRASTRUCTURE

Moves 4–6

4 Kill Silent Automations

30–50% of automation budget

Scheduled automations keep running and billing long after anyone stops reading their output. Audit your scheduled jobs and pause the ones nobody uses.

5 Cache System Prompts

15–30% per workflow

Most workflows resend the same long instructions on every call. Caching loads that context once and reuses it, so you stop paying for the same tokens again.

6 Batch Similar Tasks

10–20% per batch

Running tasks one at a time repeats the setup cost on every call. Group similar work, like 10 product descriptions, into one run so the shared context is sent once.

MOVE 6 IN PRACTICE

☒ Shared context ☐ Task

One-by-one: 10 calls



Context sent 10 times



Batched: 1 call



Context sent once



MEASUREMENT & CONTROLS

Moves 7-10

7 Use Cheaper Models for Simple Work

5-10%

Summaries and other simple tasks do not need your most expensive model. Swap Fable for Opus on this work and keep Fable for strategy and deep research.

8 Measure Token Usage Per Workflow

20-40% when you optimize

You cannot cut spend you cannot see. Log token usage for each workflow to find which ones cost the most and where the waste is.

9 Trade Real-Time for Overnight

5-25%

Reports, bulk rewrites, and other non-urgent jobs rarely need an instant answer. Schedule them as overnight batch work, around 2am, instead of running them in real time.

10 Sandbox Expensive Features

10-15% cost control

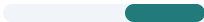
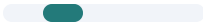
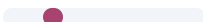
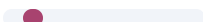
Image generation can run up costs fast when anyone can trigger it. Put it behind an approval step so each expensive request gets a sign-off first.

MODEL ROUTING TABLE

TASK TYPE	BEST MODEL	COST PER TASK	DO THIS
Classification/Tagging	 Jev	Not benchmarked	Lead scoring, sentiment, priority flags
Fast Reasoning	 Grok 4.7	~\$2.73/task	Copy variations, research, strategy first-pass
Great Daily Driver	 Claude Opus	~\$3/task	Copyediting, drafts, brand voice
Thinking & Planning	 Claude Fable 5.1	~\$7.63/task	Strategy docs, deep research, fact-checking
Code/Infrastructure	 Codex	Specialized	Landing pages, tracking, integrations

Costs are Artificial Analysis weighted benchmark-task averages; actual production cost varies by prompt length, reasoning effort, caching, and output.

WHERE THE SAVINGS STACK UP savings range per move, 0-50% scale. Moves 1-3 are not shown as percentages.

4 Kill silent automations		30-50%	6 Batch similar tasks		10-20%
8 Measure token usage		20-40%	10 Sandbox features		10-15%
5 Cache system prompts		15-30%	7 Cheaper models		5-10%
9 Overnight batch work		5-25%			