



FOR BANKING TEAMS · CREDIT EFFICIENCY

Get the most from every credit.

Credits are a shared resource. The question isn't whether to use AI — it's which activities are worth the spend, and which quietly drain the pool. This guide shows you where the line is, and how to keep bankers on the right side of it.

-
- ✓ High-value work that earns its cost — with real examples
 - ✓ The five-question efficiency framework
 - ✓ Habits for individuals, leads, and admins
 - ✓ Ten low-value patterns and the cheaper way to do each
 - ✓ A tool-to-task credit reference
 - ✓ Talking points for coaching your team

THE GOAL

Spend credits where they pay off

Every AI call — a chat, an agent run, a report, a synthesis — draws credits from your bank's balance. Getting real leverage from Hapax isn't about using it less; it's about knowing which work is worth the spend and which quietly drains the shared pool.

How Hapax credits work

Credits are the usage currency

Every AI task makes model calls; the cost is converted to credits.

Two pools feed every task

Each user's included monthly allowance, plus a shared purchased pool — usage draws from the allowance first, then the shared pool.

Bigger tasks cost more

Long prompts, long responses, large document sets, and repeated back-and-forth all increase credit use.

You have full visibility

Admins see exactly where credits go in Company Settings → Billing → Usage Details, broken down by feature, model, and user.

What actually drives cost

It isn't the number of chats. Three things move the needle:

Task complexity

A quick chat and a multi-step agent run are worlds apart in cost.

Context length

Long prompts, big document sets, and pasted text all add up fast.

Repetition

Every re-ask and every extra turn re-processes the same context.

The kicker: a single full agent run can cost 10–50× a simple chat. Credit cost is driven by task complexity, context length, and repetition — and matching the tool to the task is where most of the savings live.

TWO KINDS OF AI WORK · 01

Worth the credits

These produce outsized value, save hours of manual work, or reduce risk. The spend is justified because the alternative is expensive staff time, missed signals, or slower decisions.

ACTIVITY	WHY IT'S HIGH-VALUE	BANKING EXAMPLE
Executive QBR & board prep	Compresses hours of cross-tool research into one structured brief.	Pull core banking, CRM, loan pipeline, and branch performance data into one board or regulator-meeting brief.
Customer health & risk scoring	Surfaces churn risk, adoption gaps, and escalation signals early.	Generate weekly health briefs for top 50 commercial depositors or high-risk loan relationships.
Deal pipeline analysis & forecasting	Deal-stage insight without manual CRM exports.	Which commercial loans are stuck in underwriting, and what should relationship managers do this week?
Compliance & regulatory research	Cited answers that cut legal and research time.	What do FFIEC examination procedures require for BSA/AML training? Summarize Reg E dispute timing and notice rules.
Security alert investigation	Speeds incident response and shrinks the exposure window.	Investigate anomalous wire transfers, login patterns, or suspicious account activity.
Cross-tool data stitching	Unifies scattered information into one coherent picture.	Combine CRM, core banking, loan origination, and call notes for a 360° view of a commercial borrower.
Complex report generation	Automates work that would otherwise take analysts hours.	Build call report support, HMDA analysis, CECL/allowance, or branch profitability reports.
Strategic proposal & pitch creation	Accelerates revenue-generating deliverables.	Draft a tailored treasury management or commercial lending proposal for a local business or municipality.
Complaint pattern analysis	Identifies systemic issues and operational improvements.	Analyze Reg E, ADA, or branch-level complaint volume, categories, and resolution trends.
Automated follow-up & response drafting	Preserves context and tone while saving time.	Draft personalized follow-ups after a commercial loan meeting or a post-exam response to regulators.

TWO KINDS OF AI WORK · 02

Quietly draining the pool

These consume credits without commensurate value — usually from treating Hapax like a search engine, calculator, or content generator rather than a strategic reasoning partner. Each has a cheaper alternative.

PATTERN	WHY IT COSTS MORE	BETTER APPROACH
Repeatedly asking the same question	Each re-prompt re-processes context.	Better: Save the answer, create a process, or document it in a team runbook.
Vague, open-ended research queries	Broad prompts force long, unfocused responses.	Better: Narrow the scope: “Compare 3 vendors on pricing and security” beats “Tell me about vendor management.”
Regenerating similar outputs from scratch	Each generation re-consumes the same context.	Better: Save successful prompts as processes or documented runbooks.
Using AI for simple lookups	A single fact, balance, or record is cheaper in your core system.	Better: Use Hapax for synthesis; use core banking or CRM for raw lookups.
Long, meandering chat threads	Context grows with every turn, inflating token cost.	Better: Start a fresh chat for new topics; summarize and restart.
Running powerful agents for trivial tasks	A full agent run can cost 10–50× a simple chat.	Better: Match the tool to the task: chat for quick questions, agents for multi-step work.
Ignoring existing processes and playbooks	Re-creating standard documents wastes credits.	Better: Use saved processes, vault documents, and team runbooks.
Asking for output without acting on it	Analysis that sits unread is pure cost.	Better: Only run reports and briefs tied to a decision or action.
Unattended automations on stale triggers	Old tasks firing on irrelevant events burn credits silently.	Better: Audit Orchestrator tasks quarterly; pause or delete obsolete ones.
Copy-pasting huge documents into chat	Long inputs drive up input-token costs fast.	Better: Use document references or vault links instead of pasting full text.

BEFORE YOU START

Five questions before any task

The banker's credit-efficiency framework — a quick gut-check that keeps credits pointed at work that matters.

1 Would this take a human 30+ minutes to do well?

Yes → a good candidate for AI. No → consider doing it manually or with a simpler tool.

2 Is the output tied to a decision, action, or deliverable?

Yes → credits are likely well spent. No → pause and clarify the purpose first.

3 Could I reuse this prompt or output next time?

Yes → save it as a process or runbook to avoid future credit burn. No → keep it narrow and specific.

4 Am I using the right tool for the job?

Quick question → Chat. Recurring multi-step work → Process or Task. Deep research → Knowledge / Compliance agent. Structured data → Data Warehouse agent.

5 Is my prompt specific enough?

Vague prompts waste tokens. Add audience, format, length, and exclusions.

Practical habits to reduce waste

FOR INDIVIDUAL BANKERS

- Start fresh chats for new topics instead of dragging one thread across unrelated work.
- Reference documents, don't paste them — use vault links or attachments so Hapax reads efficiently.
- Iterate once, not ten times. Refine the prompt in a turn or two, then use the result.
- Save winning prompts as notes or runbooks; convert recurring ones to published processes.

FOR TEAM LEADS

- Review Usage Details weekly; look for spikes by feature, model, and user.
- Set per-user limits for lighter-AI roles; reserve the shared pool for high-value work.
- Standardize recurring work as published processes so teams don't reinvent prompts.
- Train users on when to use chat vs. agents vs. tasks.

FOR ADMINS

- Set monthly spend caps and auto-reload thresholds that match budget reality.
- Audit Orchestrator tasks quarterly; pause automations that no longer produce actionable output.
- Encourage process reuse so the org doesn't pay to generate the same analysis repeatedly.

COACH THE HABIT

How to talk about credit value

Simple ways to explain when AI is — and isn't — worth the credits.

HIGH-VALUE USE

“This board brief pulls together core banking, loan pipeline, and branch performance data that would take 2 hours to compile manually. The AI cost is small compared to the executive time saved.”

LOWER-VALUE USE

“Instead of asking Hapax to recalculate this each time, let's pull it from the core banking report or spreadsheet view so we don't burn credits on a 30-second lookup.”

PROCESS HABIT

“You run this same loan pipeline review every Monday. Let's turn it into a scheduled task so it runs once, cleanly, instead of being rebuilt in chat every week.”

Quick reference: match the tool to the task

TASK TYPE	BEST TOOL	RELATIVE CREDIT COST
Quick question or clarification	Hapax Assistant chat	Low
Recurring report or brief	Published Process + Task	Medium
Multi-source research with citations	Knowledge / Compliance agent	Medium-High
Structured data query	Data Warehouse agent	Low-Medium
Complex synthesis across many documents	Agent with document context	High
One-off creative draft (email, proposal)	Assistant or Email Drafter	Low-Medium

The bottom line

Credits are cheapest when Hapax does work a banker would otherwise do slowly, poorly, or not at all. They're most expensive on work that takes a banker seconds, or work that never leads to action. Treat credits like any other operational budget: spend deliberately, measure usage, and build reusable processes so the same investment keeps paying off.

Questions about your credit usage? Talk to your Hapax account team — [askhapax.ai](#)