

AI INFRASTRUCTURE ECONOMICS

A GrowCFO playbook produced in partnership with Cloud Capital

01

See the full cost

02

Forecast the drivers

03

Attribute and price

ABOUT THIS PLAYBOOK

Built for CFOs managing the economics of AI

GrowCFO produced this playbook in partnership with Cloud Capital. It accompanies the webinar based on [The Cost of Compute: AI Edition](#) and turns the report findings into exercises, questions and templates that finance teams can apply immediately.

The promise: by the end of this guide, a CFO should be able to locate hidden AI cost, identify the drivers causing forecast variances, test customer economics and agree the next actions with Product and Engineering.

THE CROSS-FUNCTIONAL TEAM



Finance

Owns margin and forecast



Product

Connects strategy to action



Engineering

Explains architecture and usage



Commercial

Connects value, product and price

Start here: five next-day actions

1

NEXT DAY

Find the full cost

Reconcile model, compute, data, storage, network and tooling costs.

2

NEXT DAY

Name the cost drivers

Link spend to users, workloads, products and customer activity.

3

NEXT DAY

Stress-test the forecast

Model base, high-adoption and extreme-usage scenarios.

4

NEXT DAY

Attribute one feature

Trace one AI feature from infrastructure spend to customer value.

5

NEXT DAY

Test one pricing decision

Decide whether current pricing recovers the cost at heavy usage.

Each action is expanded throughout the playbook so you can build an initial economic diagnostic as you read.

Exercise 1: find the AI cost you already have

Capture	Example evidence
Total AI cost	Model, compute, data, storage, network and tooling
Business driver	Users, requests, workload size, generation volume
Allocation level	Total, feature, segment, customer or transaction
Commercial result	Revenue, cost, gross profit and margin by usage band

ASK FINANCE AND ENGINEERING FOR ONE NUMBER

Ask both teams to estimate last month's total AI-related infrastructure cost. Compare the answers, the scope included and the evidence used. The disagreement reveals the visibility gap.

QUESTION 1

Where is AI cost hiding in your P&L?

Practical answer: treat AI cost as a full-stack infrastructure question, not merely a token bill.

QUESTION 1

Look beneath the model bill



Visible cost

Model APIs, token usage, dedicated AI vendors and directly billed inference.



Embedded cost

Compute, GPU capacity, data pipelines, storage, networking, egress and observability.



Operational cost

Experimentation, unused capacity, reliability work, human review and exception handling.

DO TOMORROW

Tomorrow morning action: reconcile one AI feature to every infrastructure cost line it drives. Mark anything that cannot be separated from general cloud spend.

QUESTION 1

Exercise 2: sort today's AI cost into layers

Layer	Example cost	Minimum evidence
Visible	Model API and token spend	Invoice plus usage report
Embedded	GPU, compute, storage and egress	Tags, workload map and unit rates
Operational	Observability and reliability	Named service owner and allocation key

TEAM TASK

Take five live AI costs and place them in one of these three layers.

QUESTION 1

The “before you forecast” test

- ✓ Does the forecast cover the full AI cost stack?
- ✓ Is usage linked to a measurable business driver?
- ✓ Are adoption, workload size and unit cost separate assumptions?
- ✓ Have base, upside and extreme-usage cases been modelled?
- ✓ Can Finance explain the forecast variance by component?
- ✓ Has Engineering challenged the architecture assumptions?

DO TOMORROW

Use this checklist before accepting the next AI infrastructure forecast.

5 MINUTES

Micro-exercise: take the last forecast and identify which assumption caused the largest variance.

QUESTION 1

Build a one-page AI cost-driver register

Field	Example entry
Feature	AI forecasting assistant
Owner	VP Product and FP&A lead
Usage driver	Active customers × weekly requests
Cost stack	Model, compute, data, storage, egress
Allocation key	Tagged workload plus request ID
Forecast	Base / high / extreme usage
Pricing Recovery	Fixed allocation
Margin status	Uncertain — heavy-user test required

DO TOMORROW

Create the register in a spreadsheet. Start with one product and improve it monthly.

Minimum fields

Feature · Owner · Usage driver · Cost stack ·
Allocation key · Forecast · Price · Margin status

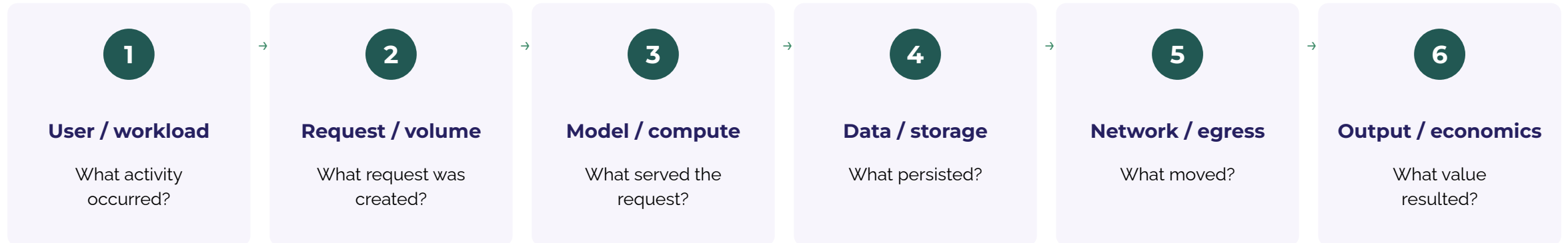
QUESTION 2

Why does AI cost keep breaking forecast?

Practical answer: forecast business activity and architecture together, then reconcile every variance.

QUESTION 2

Map the AI cost journey



DO TOMORROW

Tomorrow morning action: choose one AI feature and map this cost journey. Mark every unmeasured stage with a question mark.

COLLECT

Evidence to collect: usage telemetry, service tags, unit rates, storage growth, architecture map, customer activity and pricing terms.

QUESTION 2

Exercise 3: ask Engineering the question that matters

Send this to your engineering lead

"For our highest-growth AI feature, can you show the business activity that drives cost, the infrastructure services involved, the allocation tags available and the assumptions Finance should use to forecast the next quarter?"

Record the answer in the driver register

- Usage driver: Known / Assumed / Unknown
- Unit cost: Known / Estimated / Unknown
- Allocation tags: Complete / Partial / Missing
- Forecast owner: Named / Unclear
- Pricing recovery: Tested / Not tested

QUESTION 2

Ten questions Finance should ask Engineering

- 1 Which AI features and workloads create material cost?
- 2 Which business activity is the best forecast driver?
- 3 Can spend be traced to product and customer?
- 4 What sits beneath the visible model bill?
- 5 What level of capacity commitment does forecast usage support?
- 6 Which costs are fixed, stepped and truly variable?
- 7 How do architecture changes alter unit cost?
- 8 What usage or allocation telemetry is missing?
- 9 Where did the last forecast variance originate?
- 10 Which commercial decision needs better cost evidence?

DO TOMORROW

Pick the fastest-growing feature. Ask for its driver, full cost and customer economics.

DO NOT ACCEPT

"It scales with usage" without a defined driver, unit rate, owner and reconciliation.

QUESTION 2

Exercise 4: run a 30-minute forecast audit

Check	Question to answer
Scope	Does the forecast include the full cost stack?
Drivers	Are costs linked to measurable business activity?
Scenarios	Are adoption and unit-cost shocks modelled?
Attribution	Can actuals be reconciled by product or customer?
Architecture	Have planned engineering changes been reflected?
Ownership	Are Finance and Engineering owners named?

30 MIN

Output: one page showing green = evidenced, amber = assumed, red = unknown.

OWNER

Engineering supplies drivers; Finance owns forecast and margin implications.

QUESTION 3

How do attribution and pricing protect gross margin?

Practical answer: trace cost to the unit that creates value, then choose a pricing model that recovers it.

QUESTION 3

Use the Cost Attribution Ladder

1. Total spend	Cloud and AI infrastructure combined	P&L baseline
2. AI spend	AI separated from general cloud	Cost visibility
3. Product / feature	Cost attached to what customers use	Product economics
4. Segment	Cost grouped by customer type	Pricing design
5. Customer / unit	Cost traced to customer or value unit	Unit economics

DO TOMORROW

Tomorrow morning action: place your highest-cost AI feature on the ladder and define the next attainable level.

RULE OF THUMB

Each step enables a better commercial decision. Do not wait for perfect allocation before starting.

QUESTION 3

Exercise 5: classify the economics before choosing the price



Margin-positive

Price comfortably recovers attributable cost across normal and heavy usage.



Uncertain

Attribution is incomplete or margin depends on untested usage assumptions.



Margin-negative

Incremental revenue does not recover the full infrastructure cost.

DO TOMORROW

Tomorrow morning action: classify three AI features as margin positive, uncertain or negative. Where evidence is missing, name the measurement required.

QUESTION 3

Choose pricing for value and resilience

Customer value

What outcome is valuable enough to pay for?

Usage variability

How wide is the gap between typical and heavy users?

Attribution quality

Can cost be traced to the priced unit?

Adoption maturity

Is demand stable enough to price confidently?

Margin resilience

Does pricing survive volume and unit-cost shocks?

DO TOMORROW

Tomorrow morning action: choose one feature and compare bundled, tiered, add-on and usage-based pricing against these five tests.

QUESTION 3

Exercise 6: validate one AI unit before scaling it

- ✓ Can I trace full cost to the feature?
- ✓ Does usage reconcile to the invoice?
- ✓ Is cost measured per customer or value unit?
- ✓ Does price recover the cost at typical usage?
- ✓ Does price recover the cost at heavy usage?
- ✓ Would this stand up to board or investor challenge?

DO TOMORROW

Take one AI feature and calculate its contribution margin at typical, heavy and extreme usage.

CORE RULE

AI growth is not healthy until unit economics are checked.

QUESTION 3

Set the CFO's economic red lines

- No AI launch without a named cost and margin owner.
- No forecast built solely from token or API pricing.
- No bundled AI pricing without a heavy-usage stress test.
- No material AI spend without product or workload attribution.
- No long-term capacity commitment beyond the period or volume supported by the forecast..
- No margin claim that excludes material infrastructure cost.

DO TOMORROW

Tomorrow morning action: choose three red lines and agree them with Product and Engineering.

Turn the diagnostic into a 30-day rhythm

Timing	Focus	Practical output
Day 1	Establish the baseline	Reconcile last month's total AI infrastructure cost.
Day 2	Map one feature	Connect usage, architecture, cost and customer value.
Week 1	Build the driver register	Define owners, drivers, allocation keys and unknowns.
Week 2	Stress-test the forecast	Model base, high-adoption and extreme usage.
Week 3	Test pricing recovery	Compare typical and heavy-user economics.
Week 4	Agree interventions	Prioritise instrumentation, pricing and commitment actions.

Template: AI cost-driver register

Field	Prompt
Feature / workload	What AI capability creates the cost?
Owner	Who owns usage, cost and margin?
Business driver	What measurable activity predicts consumption?
Cost stack	Which model, compute, data, storage and network costs apply?
Allocation key	How is cost traced to feature, segment or customer?
Forecast assumptions	What changes under base, high and extreme usage?
Pricing recovery	How does the pricing model recover attributable cost?
Status	Evidenced, assumed, unknown or action required?

Template: AI economics readiness score

Score 1–5: visibility and attribution

- AI separated from cloud
- Full-stack cost captured
- Product / feature attribution
- Customer-level attribution
- Usage telemetry quality
- Owner accountability

Score 1–5: forecast and commercial response

- Driver-based forecasting
- Scenario modelling
- Variance reconciliation
- Pricing recovery
- Commitment confidence
- Margin reporting

Strong: scale confidently (8-10)

Developing: instrument now (5-7)

Exposed: intervene before scale (under 5)

Template: monthly AI economics review

1 Baseline

Capture full AI infrastructure spend.

2 Explain

Reconcile variance to usage and architecture.

3 Attribute

Assign cost to product, segment or customer.

4 Test

Stress-test forecast, price and margin.

5 Decide

Change architecture, pricing or commitment.

6 Repeat

Refresh drivers and improve instrumentation.

DO TOMORROW

Tomorrow morning action: schedule a monthly 45-minute review owned jointly by Finance and Engineering.