

## Executive Summary

AI introduceert een fundamenteel nieuw kostenparadigma.

Waar traditionele IT voorspelbare licenties kende, verschuift AI naar een dynamisch consumptiemodel gebaseerd op GPU-capaciteit, tokens, opslag en dataverkeer.

Dit model creëert potentieel:

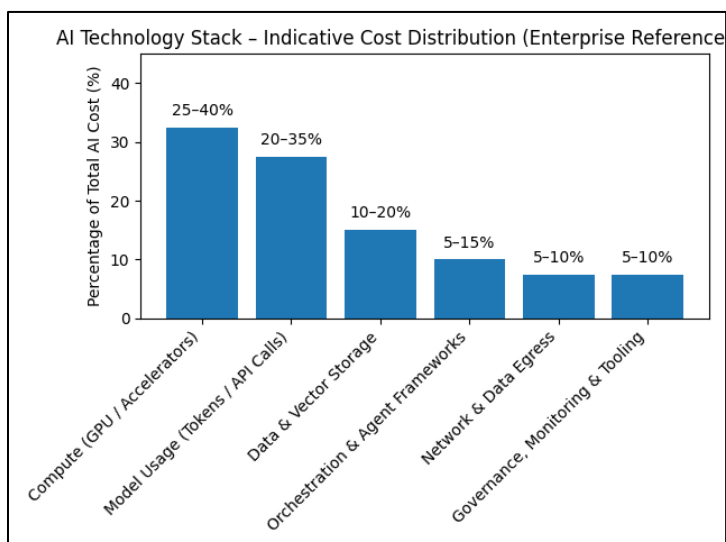
- Hoge variabele kosten
- Moeilijk voorspelbare uitgaven
- Exponentiële schaal bij succesvolle adoptie
- Contractuele risico's bij open-ended afspraken

Voor CFO's betekent dit: AI is geen IT-project, het is een financieel hoog risicoprofiel. AI FinOps is de benodigde discipline die AI-consumptie bestuurbaar en controleerbaar maakt.

## Introductie AI Technology Stack – Kostenverdeling (Indicatief Model)

Voor een goed begrip van AI FinOps laten we eerst eens goed kijken naar de gehele AI Technology Stack. Hoe is die opgebouwd? En welke componenten spelen een significante rol? En welke componenten lijken klein maar kunnen exponentieel opschalen?

Onderstaande verdeling in figuur 1 is gebaseerd op marktobservaties bij AI-implementaties op bedrijfsschaal. Werkelijke percentages kunnen verschillen per use case, maar dit model geeft een eerste realistisch referentiekader voor management.



**Figuur 1. Kostenverdeling AI Tech Stack.**

De eerste inzichten zijn nu duidelijk. Potentieel is 75% (!) van de kosten variabel en dat kan met uitrol, schalen en 'self-learning' AI Agents financieel behoorlijk uit de hand lopen.

Hieronder een gedetailleerde toelichting per laag in de Tech\_Stack.

Wat betekent het nou precies? En wat zijn de relatieve kosten per laag? Vast/variabele kosten? Waar zitten de echte commerciële risico's?

### **Level 1. Compute (25–40%)**

GPU-capaciteit die via partijen zoals Amazon Web Services, Microsoft Azure en Google Cloud wordt gebruikt. 'Compute' vormt vaak de grootste directe kostenpost.

### **Level 2. Model Usage (20–35%)**

Tokenverbruik bij leveranciers zoals OpenAI (ChatGPT), Anthropic en Google. Multi-agent systemen die veel onderling communiceren kunnen dit segment exponentieel vergroten (!).

### **Level 3. Netwerk & Egress (5–10%)**

Met name relevant bij multi-cloud en data-intensieve workloads. Network & Egress verwijst naar de kosten voor dataverkeer binnen en buiten cloud omgevingen.

Het bestaat uit twee componenten:

#### **Netwerkkosten**

- Dataverkeer binnen een cloud omgeving
- Dataverkeer tussen verschillende services
- Dataverkeer tussen availability zones of regio's

#### **Data Egress**

- Kosten voor data die een cloud omgeving verlaat

Cloud providers zoals Amazon Web Services, Microsoft Azure en Google Cloud rekenen vrijwel altijd kosten voor egress (data die de cloud omgeving verlaat, bijvoorbeeld in een Hybride cloud met 2 providers of data wordt bijvoorbeeld verplaatst naar een Data Lake 'On Premise').

Waarom is dit relevant? Egress-kosten zijn veelal niet direct 100% zichtbaar in AI-prijsmodellen en alleen zichtbaar in 'de kleine lettertjes'.

### **Level 4. Data & Vector Storage (10–20%)**

Vector storage is een gespecialiseerde manier om data op te slaan als numerieke representaties (vectoren), zodat AI-systemen betekenis en context kunnen begrijpen.

In plaats van tekst "letterlijk" op te slaan, wordt tekst eerst omgezet in een reeks getallen ('embeddings'). Die getallen representeren de betekenis van de tekst.

Kort samengevat, Vector storage is:

- De geheugenlaag van generatieve AI
- Essentieel voor bedrijfsdata-integratie
- Variabel en schaalbaar in kosten
- Indien goede governance ontbreekt een zeer structurele Opex-component

### Level 5. Orchestration & Agent Frameworks (5–15%)

Workflow engines, prompt management, agent runtimes.

### Level 6. Governance & Monitoring (5–10%)

De totale kosten (medewerkers, tooling) van monitoring, observability, compliance tooling.

## Wat is belangrijk en essentieel voor CEO, CFO en Hoofd Inkoop?

- 60–75% van AI-kosten is direct variabel.
- **Kosten schalen** niet lineair maar **exponentieel** bij agent-gedreven architecturen.
- De **grootste kostenpost** (compute + tokens) is **volledig consumptiegedreven**.
- De werkelijke exposure zit in **onbegrensde schaal** (!).

## 1. AI is geen softwarelicentie, het is ‘energieverbruik’

AI gedraagt zich financieel meer als energie dan als software.

Kenmerken:

- Pay-per-use
- Continue consumptie
- Directe koppeling tussen gebruik en kosten
- Geen natuurlijke bovengrens

Elke AI-interactie triggert:

- Compute
- Model inference
- Opslag
- Netwerkverkeer

Bij AI agent-based systemen vermenigvuldigt elke taak zichzelf via onderliggende model calls.

## 2. De Black Box van AI-consumptie

CFO's ontvangen maandelijks een cloud- of modelrekening. Maar vaak ontbreekt het inzicht in:

- Welke business unit en gebruikers de kosten veroorzaakt
- Welke process workflow inefficiënt is
- Welke prompts excessief token verbruik genereren
- Welke AI agents sub-agents aanroepen.

- En hoeveel AI Agents en Sub Agents zijn er eigenlijk in de bedrijfsorganisatie?
- En welke LLMs worden gebruikt? En welke vector dbases?
- Etc

Kortom, AI-consumptie in de gehele AI Tech Stack is technisch complex en financieel diffuus. Zonder goede governance ontstaat direct een **'black box'**.

### 3. Het contractuele risico: Open-einde exposure

Veel AI-contracten bevatten:

- Pay-per-token (vaak zonder enig plafond)
- Onbepaalde autoscaling
- Geen spend caps
- Onduidelijke definities van usage
- Volume staffels zonder een absolute limiet

Inkoop onderhandelt vaak over de prijs per eenheid en niet over de maximale exposure in de AI Tech stack. Dit brengt ongewenste risico's:

- Onverwachte budgetoverschrijdingen
- Onvoorspelbare Opex
- Margedruk bij digitale businessmodellen met hoge consumptie
- Maximale exposure bij ongecontroleerde uitrol van AI Agents

### 4. AI Agents: De financiële multiplier

AI-agents introduceren een nieuwe dynamiek die we niet op deze schaal eerder hebben meegemaakt:

- Eén klantvraag → meerdere model calls
- Eén workflow → 1 of meerdere AI agents
- Iteratieve besluitvorming → exponentiële token consumptie

Wat lineair lijkt, schaalt vaak exponentieel. Voor CFO's betekent dit Kosten per Transactie goed moeten worden gemodelleerd vóór de AI uitrol om werking van de financiële multiplier te voorkomen en tot een minimum te reduceren.

## 5. AI FinOps als strategische discipline

AI FinOps combineert:

- Financiële forecasting
- Contractstrategie
- Technische governance
- Continue optimalisatie

Het doel van AI FinOps is niet om kosten te minimaliseren, het doel is om exposure goed te beheersen: 100% 'In Control'.

## 6. Strategische aanbevelingen voor CFO & Procurement

### 6.1 Onderhandel op exposure, niet alleen op prijs.

Simpelweg een lager token-tarief onderhandelen zonder een plafond is géén risicoreductie.

Enkele contractuele best practices:

- Jaarlijkse spend caps
- Maandelijkse ceilings (bijvoorbeeld 'belbundels')
- Hard throttling bij overschrijding (een dienst of platform wordt het gebruik direct en strikt beperkt zodra een vooraf ingestelde limiet is bereikt, bijvoorbeeld het aantal API calls. Geen geleidelijke vertraging, maar een harde afkapping van capaciteit of toegang)
- Transparante usage-definities (zoals bij mobiel bellen 'belminuten')
- Inzage in token- en compute metrics (een transparant AI Dashboard)
- Exit-clausules bij prijswijzigingen
- Geen vrijbrief voor jaarlijkse indexatie en / of inflatiecorrectie

### 6.2 Maak AI-kosten voorspelbaar per business use case

Vraag niet: "Wat kost AI?". Maar stel de volgende vragen:

- Wat kost een AI-interactie (TCO) per gebruiker?
- Wat kost een agent (TCO) per workflow?
- Wat is de cost-per-decision?
- Wat is de overal impact op de winst en marge (positief/negatief)?

Kortom, maak AI-kosten integraal onderdeel van unit economics. Een integraal onderdeel van uw margeberekeningen voor uw producten/services.

### 6.3 Voer pre-implementation cost simulations uit

Aangezien AI kosten zeer variabel zijn (45% en potentieel 75%) is het essentieel dat AI modellen en bijbehorende kosten worden gesimuleerd alvorens AI agents worden uitgerold:

- Simuleer tokenverbruik
- Modelleer worst-case scenario's
- Analyseer exponentiële schaalpatronen
- Definieer break-even volumes

AI zonder vooraf een goede kostensimulatie ('What-If' scenario's) is financieel eigenlijk **zeer speculatief**.

### 6.4 Governance als expliciete ontwerpeis

Geen AI uitrol zonder goed Governance model. De financiële controle moet technisch worden ingebouwd.

De AI-architectuur moet bevatten:

- Budget guardrails
- Rate limits
- Depth limits in agent-calls
- Real-time cost dashboards
- Management Alerts bij afwijkingen

## 7. De CFO-agenda voor 2026 en verder

De organisaties die AI succesvol opschalen:

- Behandelen AI als strategische energievoorziening
- AI Governance als ontwerpeis
- Begrijpen hun kostenstructuur per laag in de AI Technology Stack
- Simuleren vooraf kostenscenario's en begrijpen de impact op marge / winst
- Hebben contractuele plafonds
- Monitoren cost-per-use case

AI is geen 'leuk innovatieproject' en / of pilot meer. Bij uitrol van AI Agents is het een structurele kostenpost met strategische impact.

## Conclusie

AI biedt concurrentievoordeel. Maar zonder een goed fundament, AI FinOps ontstaat:

- Onbegrensde Opex
- Contractuele afhankelijkheid (Strategische vendor lock-in)
- Margedruk
- Verminderde financiële voorspelbaarheid

Met behulp van AI FinOps ontstaat:

- Transparantie
- Contractuele beheersing
- Beschermde marge
- Strategische controle

Voor CFO's en Inkoop is de kernvraag niet: "Hoe snel kunnen we AI uitrollen?" maar met name "Hoe zorgen we dat AI schaalbaar is en bijdraagt een winstgevende groei, zonder onbeheersbare exposure?"

## AI FinOps Assessment

Met de opkomst van de AI is Inkoop van IT services geëvolueerd van voorspelbare, eenmalige kapitaaluitgaven en transparante operationele uitgaven naar variabele, op verbruik gebaseerde maandelijkse facturen. Het gevolg is dat AI engineers (en zelfs de AI Agents) dagelijks beslissingen nemen over uitgaven die de winstgevendheid van uw bedrijf razendsnel beïnvloeden, terwijl inkoop en de financiële teams moeite hebben om het tempo en de details van deze uitgaven bij te houden.

Hoe kan uw organisatie de AI kosten bijhouden en inzichtelijk maken? En winstgevend groeien? Bent u überhaupt 'In Control'? Of is er sprake van een 'black box'?

Met ons AI FinOps Assessment:

- Verbetert u het inzicht in AI TCO, zodat u goed geïnformeerde afwegingen kunt maken tussen snelheid, kwaliteit en kosten.
- Brengt u financiën, inkoop, technologie en bedrijfsleiderschap samen in een bedrijfsoverkoepelende samenwerking om innovatie, aanpassingsvermogen en adoptie te bevorderen op een gecontroleerde wijze.
- Stelt AI leiders in staat om de kosteneffectiviteit van de AI Stack te beheersen, de bedrijfswaarde van de AI te verhogen en uw concurrentievoordeel te versterken.

Edwin Knoop, February 2026. PS: All views are on our own.