De reelle omkostninger ved AI: forbrugsafregnede cloudtjenester eller en fast licens på jeres egne servere

Hvordan forbrugsafregnet AI faktureres, hvorfor regningen vokser i takt med succesen, hvilke omkostninger der aldrig står på fakturaen, og et beregningsskema til at sammenligne en forbrugsafregnet tjeneste med en fast licens på jeres egne servere.

10. november 2026



empisto.ai · lokal · privat · pålidelig

Indhold

01 Resumé

02 Sådan faktureres forbrugsafregnet AI

03 Derfor vokser regningen med succesen

04 Omkostningerne på fakturaen

05 Omkostningerne der ikke står på fakturaen

06 Hvad en fast licens koster

07 Hvornår hvilken model giver mening

08 Et beregningsskema til jeres egen sammenligning

09 Læs mere

01 Resumé

En cloud-AI-tjeneste ser billig ud pr. spørgsmål. En enkelt forespørgsel koster en brøkdel af en cent, og der er intet at installere. Problemerne opstår først når værktøjet virker. Regningen vokser med hver person der bruger det, hvert spørgsmål de stiller og hvert dokument der sendes med. Og nogle af de største omkostninger står slet ikke på fakturaen: det juridiske arbejde med at overlade data til en udbyder, de ændringer I ikke selv har valgt, og den værdi der går tabt når folk holder igen med et værktøj fordi hvert spørgsmål koster penge.

En fast licens på jeres egne servere har en anden profil. Omkostningerne kendes før I går i gang: licensen, hardwaren og opsætningen. Derefter koster mere brug ikke ekstra. Det er ikke altid det billigste valg, og denne whitepaper siger hvornår det ikke er det. Til gengæld får I en omkostning I kan budgettere med, og et værktøj medarbejderne kan bruge så meget som det hjælper dem.

Denne whitepaper forklarer hvordan forbrugsafregnet AI faktureres, hvor de synlige og skjulte omkostninger kommer fra, hvornår hver model giver mening og hvordan I sammenligner de to for jeres egen organisation.

FORBRUGSAFREGNET AI

Faktureres pr. token eller pr. brugerlicens med forbrugsgrænser. Regningen følger brugen.

HVAD STYRER REGNINGEN

Brugere, spørgsmål og de dokumenter og den historik der sendes med hvert spørgsmål.

EMPISTO-LICENS

Én fast årspris. Ubegrænset antal spørgsmål for alle med adgang.

HARDWARE

Jeres egen, ofte servere I allerede har. Ny hardware køber I direkte, uden avance.

OPSÆTNING OG SUPPORT

Prissat én gang, på forhånd.

CLOUDMODELLER

Valgfrie. Forbruget faktureres af udbyderen, på jeres egen konto.

02 Sådan faktureres forbrugsafregnet AI

De fleste AI-tjenester tager betaling pr. token, små bidder af ord. Et spørgsmål og alt det der sendes med, tæller som input-tokens, mens svaret tæller som output-tokens, der ofte er dyrere. Større og nyere modeller koster som regel mere pr. token end mindre modeller. Nogle tjenester sælger i stedet brugerlicenser til en fast pris pr. bruger, men de har som regel forbrugsgrænser, og bruger I tjenesten mere eller vil I have bedre modeller, rykker I op på et højere niveau.

Det der overrasker de fleste købere, er hvor meget der sendes med hvert spørgsmål. Når en assistent svarer ud fra virksomhedens dokumenter, indeholder hver forespørgsel langt mere end det spørgsmål nogen har skrevet.

Instruktioner
Hvordan assistenten skal opføre sig, sendt hver gang

+

Samtale
Alt hvad der er sagt i chatten indtil nu

+

Dokumenter
De uddrag der er fundet til at besvare spørgsmålet

+

Spørgsmål
Det personen skrev

=

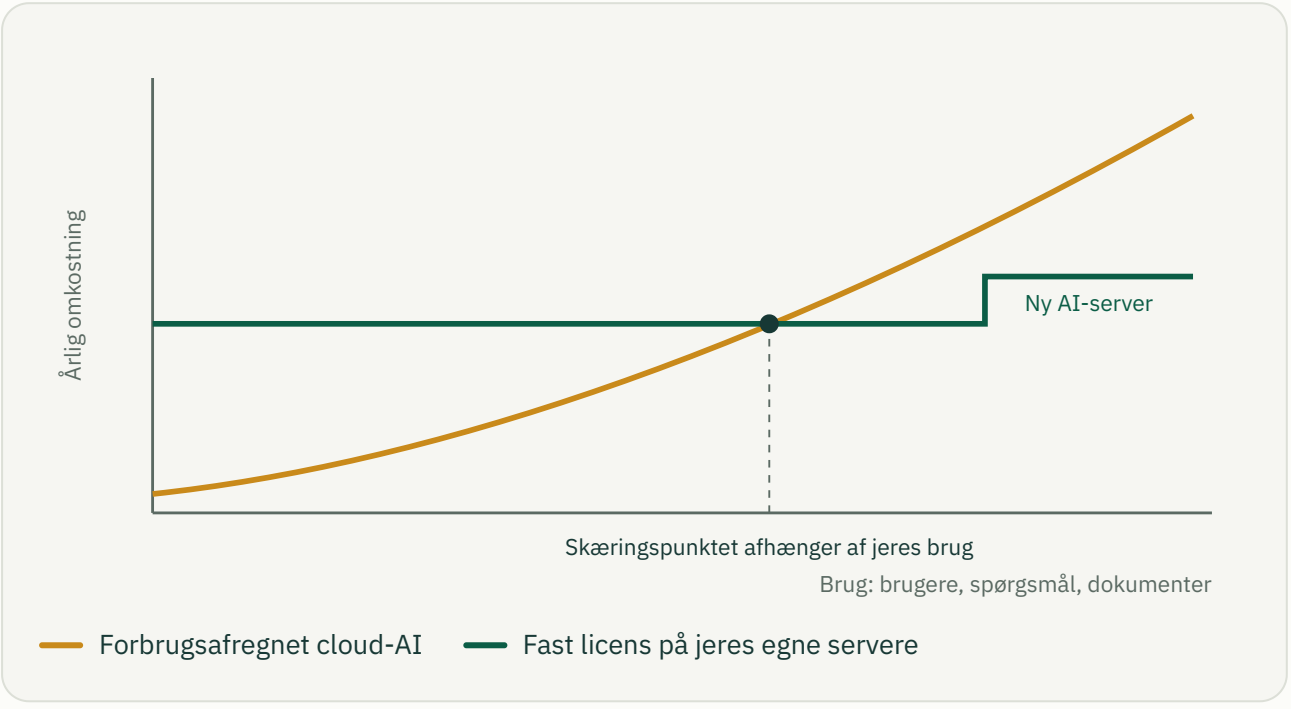
Input-tokens
Faktureres for hvert spørgsmål, plus svaret som output-tokens

Figur 1. Det ét spørgsmål sender til en forbrugsafregnet model. Ved spørgsmål om dokumenter fylder uddragene og samtalen som regel mange gange mere end selve spørgsmålet.

Et kort spørgsmål kan derfor trække tusindvis af tokens med sig, og i en lang samtale bliver det mere for hver ny besked fordi hele historikken sendes med igen. Gang det med antal spørgsmål pr. person, antal brugere og antal arbejdsdage, så står det klart hvordan regningen tager form.

03 Derfor vokser regningen med succesen

Målet med en AI-assistent er at medarbejderne bruger den hver dag til rigtigt arbejde. Med et taxameter gør hvert skridt mod det mål regningen større: flere teams, flere spørgsmål, længere samtaler, større dokumenter, bedre modeller. Regningen følger udbredelsen, og det er netop udbredelse I gerne vil fremme.



Figur 2. Formen på de to omkostninger, ikke deres størrelse. Taxameteret starter lavere og bliver ved med at stige, mens den faste licens starter højere og forbliver flad når brugen vokser.

Der er også en mindre synlig effekt. Når hvert spørgsmål har en pris, begynder organisationer at rationere: adgang for nogle få teams, loft over forbruget pr. person, den billigere model til det meste af arbejdet. Hvert af de greb sparer penge på fakturaen, men koster værdi i forretningen. Og medarbejdere der ikke frit kan bruge det officielle værktøj, har det med at gribe til offentlige AI-værktøjer i stedet og tage virksomhedens data med sig.

04 Omkostningerne på fakturaen

Omkostning	Forbrugsafregnet cloud-AI	Empisto på jeres egne servere
Brug	Pr. token eller pr. brugerlicens med grænser. Vokser med hvert spørgsmål	Ingen. Ubegrænset antal spørgsmål indgår i licensen
Bedre modeller	Højere pris pr. token eller et højere niveau	Vælg selv de åbne modeller I vil have, uden pris pr. spørgsmål
Dokumenter	Lagring og indeksering af dokumenter kan blive faktureret særskilt	Med i prisen, på jeres egne servere
Dataoverførsel	Det koster ofte at flytte data ud af en cloududbyders netværk	Ingen. Data bliver hvor de er
Licens	Ingen. I betaler løbende efter forbrug	Én fast årspris, ud fra jeres størrelse og de arbejdsområder I har brug for
Hardware	Intet på forhånd	Jeres egne servere, ofte nogle I allerede har. Nye købes direkte, uden avance
Opsætning og support	Som regel selvbetjening, og integrationen med jeres systemer må I selv stå for	Installation, onboarding, opdateringer og support, prissat én gang på forhånd

05 Omkostningerne der ikke står på fakturaen

Nogle af de største omkostninger ved en forbrugsafregnet tjeneste står aldrig på dens regning. De viser sig i andre budgetter, i tid som andre medarbejdere bruger, og i risici som nogen skal bære.

At give data fra sig. Når I sender spørgsmål og dokumenter til en udbyder, bliver udbyderen en del af jeres databehandling. Det betyder kontrakter, vurderinger af databeskyttelsen, kontrol af hvor data havner og hvem der ellers ser dem, og en ny gennemgang hver gang udbyderen ændrer sine vilkår. For fortrolige data eller personoplysninger er svaret nogle gange at de slet ikke må sendes, og så kan værktøjet ikke bruges dér hvor det ville gøre mest gavn. Vores whitepaper om GDPR og EU's AI-forordning går i dybden med det.

Ændringer I ikke selv har valgt. Hos en udbyder bliver modeller udfaset og erstattet når det passer udbyderen. Priser, grænser og vilkår ændrer sig. Hver ændring kan betyde at værktøjet skal testes igen, at måden det bruges på skal justeres, eller at aftalen skal genforhandles. Omkostningen er reel, også når fakturaen er uændret.

Et uforudsigeligt budget. En forbrugsafregnet regning skal forudsiges, og prognosen afhænger af udbredelsen, som ingen kender på forhånd. Økonomiafdelingen svarer igen med lofter og

godkendelser, og de bremser netop den udbredelse som værktøjet blev købt for at opnå.

Rationering og bivirkningerne. Grænser pr. person og billigere modeller holder regningen nede, men de holder også værdien nede. Hvor det officielle værktøj er rationeret, udfylder medarbejderne ofte hullet med offentlige AI-værktøjer, uden at IT og databeskyttelsesfolkene kan se det.

Prisen for at skifte. Dokumenter indekseret i en udbyders format, arbejdsgange bygget på dens grænseflade og prompts tilpasset dens modeller skal alle laves om hvis I skifter. Jo længere I bliver, desto dyrere er det at gå. Empisto kobles på gennem åbne standarder og holder jeres dokumenter og registreringer på jeres egne servere, så næste ændring handler om at skifte én del ud, ikke om en migrering.

06 Hvad en fast licens koster

On-premise AI er ikke gratis, og det ville være forkert at sammenligne den som om den var det. Med Empisto består omkostningerne af følgende, og I kender dem alle før I går i gang.

1. **Licensen.** Én fast årspris for jeres organisation, ud fra størrelsen og de arbejdsområder I har brug for, med ubegrænset antal spørgsmål. Den fornyes til den aftalte pris, og antallet af arbejdsområder kan ændres ved fornyelsen. En pilotlicens tæller med i den fulde licens.
2. **Hardwaren.** En server til platformen og en GPU-server til AI-modellerne. Mange organisationer starter på servere de allerede har. Giver ny hardware mening, dimensionerer vi den sammen med jer, og I køber den direkte, uden avance til os. I ejer den, og den gør tjeneste i årevis.
3. **Driften.** Strøm, plads og noget IT-tid til serverne, styresystemet og backup, som for enhver anden server I driver.
4. **Opsætning og support.** Installation, onboarding af jeres team, opdateringer og support, prissat én gang, på forhånd.
5. **Cloudmodeller, kun hvis I vil have dem.** Et arbejdsområde kan bruge en cloudmodel til egnede opgaver. Forbruget faktureres af den pågældende udbyder på jeres egen konto, så den udgift forbliver synlig og under jeres kontrol.

07 Hvornår hvilken model giver mening

Ingen af de to modeller er bedst i alle situationer. Det afgørende er hvor meget værktøjet vil blive brugt, hvilke data det kommer i berøring med, og hvor stor sikkerhed I har brug for i budgettet.

Situation	Passer som regel bedst
Nogle få brugere, lejlighedsvise spørgsmål, intet fortroligt	En forbrugsafregnet tjeneste. Lavt forbrug holder regningen lille, og der er ingen data at beskytte
En kort afprøvning af hvad AI kan	En forbrugsafregnet tjeneste eller en Empisto-pilot på jeres egne dokumenter
Mange medarbejdere der bruger AI hver dag på virksomhedens dokumenter	En fast licens. Jo mere den bruges, desto større bliver forskellen på de to omkostninger

Situation	Passer som regel bedst
Fortrolige data, personoplysninger eller reguleret materiale	En fast licens på jeres egne servere. Data forlader dem aldrig
Et budget der skal kendes på forhånd	En fast licens. Omkostningen ligger fast før året begynder
Frihed til at skifte modeller og værktøjer i takt med at AI udvikler sig	En fast licens bygget på åbne standarder, hvor en ny del afløser en gammel uden en migrering

08 Et beregningsskema til jeres egen sammenligning

En ærlig sammenligning laves med jeres egne tal. Skemaet nedenfor viser hvad der skal tælles med. Sammenlign de samlede beløb over tre år, for hardware gør tjeneste i flere år, mens et taxameter kører så længe I bruger tjenesten.

Linje	Hvad der skal tælles med	Hvor tallet kommer fra
Forbrugsafregnet tjeneste		
1. Brugere	Alle der skal bruge assistenten	Jeres planer for udrulningen
2. Spørgsmål	Spørgsmål pr. person pr. arbejdsdag og arbejdsdage pr. år	Forbruget under en afprøvning eller et forsigtigt skøn
3. Input pr. spørgsmål	Instruktioner, samtale, dokumentuddrag og selve spørgsmålet, i tokens	Udbyderens forbrugsrapporter under en afprøvning
4. Output pr. spørgsmål	Længden af et typisk svar, i tokens	De samme rapporter
5. Priser	Prisen pr. million input- og output-tokens for den model I vil bruge	Udbyderens prisliste
6. Årligt forbrug	Brugere × spørgsmål × arbejdsdage × (input × inputpris + output × outputpris), divideret med én million	Linje 1 til 5
7. Tillæg	Brugerlicenser, dokumentlagring, dataoverførsel, premiumfunktioner	Udbyderens prisliste
8. Skjulte omkostninger	Jura- og databeskyttelsesarbejde, tid til governance, omarbejde efter ændringer	Jeres jura-, IT- og complianceteams
Fast licens		
9. Licens	Den årlige licens	Jeres prisoverslag fra Empisto
10. Hardware	Eksisterende servere, eller prisen på nye fordelt over levetiden	Jeres prisoverslag og jeres leverandør

Linje	Hvad der skal tælles med	Hvor tallet kommer fra
11. Driftsomkostninger	Strøm, plads og IT-tid	Jeres IT-team
12. Opsætning og support	Prissat én gang	Jeres prisoverslag fra Empisto

Vi udfylder gerne siden for den faste licens sammen med jer som en del af prisoverslaget, og vi hjælper også gerne med at samle tallene til den forbrugsafregnede side.

09 Læs mere

Prissiden forklarer hvordan licensen fungerer. Arkitekturbeskrivelsen beskriver hvad I får og hvordan det kører på jeres servere, og whitepaperen om GDPR og AI-forordningen dækker databeskyttelsen. Vil I have et prisoverslag til jeres organisation, så start her.