

# Kompletterande material till din strategiska dataresa

Det här dokumentet samlar, förklarar och förtydligar centrala nyckelbegrepp och dataroller som nämns eller är relevanta för utbildningsserien '[Starta din strategiska dataresa](#)'.

# Innehållsförteckning

<b>Ordlista</b> .....	<b>4</b>
Rådata.....	4
Strukturerad vs. ostrukturerad data.....	4
Metadata.....	4
Grunddata (master data).....	4
Datalivscykeln.....	5
Datakvalitet – de sex dimensionerna.....	5
Data lineage (dataspårbarhet).....	5
Analytiska metoder.....	5
Generativ AI.....	6
Syntetisk data.....	6
API (application programming interface).....	6
Databas, data warehouse och data lake.....	6
Batch-överföring och realtidsdata.....	7
Silos (stuprör).....	7
Data governance (datastyrning).....	7
KPI (key performance indicator).....	7
Dashboard.....	8
Hallucinationer.....	8
Digital tvilling.....	8
Datamognadstrappan.....	8
Databehovs-pyramiden   Aiken-pyramiden.....	9
MCP (Model Context Protocol).....	9
Webhook.....	9
MLOps (Machine Learning Operations).....	10
Bias.....	10
Datakatalog.....	10
Data literacy (dataläskunnighet).....	10
<b>Rollbeskrivning</b> .....	<b>12</b>
Dataarkitekt   Data Architect.....	12
Dataingenjör   Data Engineer.....	12
Dataägare   Data owner.....	12
Datastyrningsspecialist   Data Governance Specialist.....	13
Dataförvaltare   Data Steward.....	13

Dataanalytiker   Data Analyst.....	13
Data Scientist.....	14
ML-ingenjör   Machine Learning (ML) Engineer.....	14
Generativ AI-ingenjör   Generative AI Engineer.....	14

# Ordlista

Ordlista kategoriserat över *grundbegrepp, datastyrning, processer, datakvalitet, analys, AI-verktyg, AI-risker, infrastruktur, organisation och uppföljning.*

## Rådata

**Data i sin ursprungliga, obearbetade form – siffror, text, bilder eller ljud precis som de samlats in.**

Rådata är sällan direkt användbar för beslut. Den innehåller ofta fel, luckor och inkonsekvenser. Tänk på det som råvaror innan de förädlats till en produkt. Värdet uppstår inte i rådata – utan i det arbete som görs med den.

GRUNDBEGREPP

## Strukturerad vs. ostrukturerad data

**Strukturerad data är ordnad i tabeller med rader och kolumner. Ostrukturerad data saknar fast format.**

Exempel på strukturerad data: bokföring, lagersaldon, kundregister. Ostrukturerad: e-post, bilder, mötestranskriptioner, sociala medier. Uppskattningsvis 80 % av en organisations data är ostrukturerad men används sällan. Generativ AI är särskilt bra på att analysera ostrukturerad data.

GRUNDBEGREPP

## Metadata

**Data om data – information som beskriver när, var och hur en uppgift skapades.**

Metadata fungerar som en innehållsförteckning som gör det möjligt att navigera i stora datamängder. Utan metadata vet du vad du har, men inte vad det betyder eller var det kommer ifrån. Det är avgörande för att undvika säkerhetsläckor och skapa ordning i datahanteringen.

GRUNDBEGREPP

## Grunddata (master data)

**Verksamhetskritisk information som organisationen måste kunna lita på – kundregister, produkter, platser och finansiella strukturer.**

Det är beständig data som inte förändras snabbt, men som ändå behöver underhållas aktivt. Om grunddata är felaktig sprider sig felet till allt som bygger på den: fakturor, leveranser och analyser. Det kräver aktivt ledarskap och tydliga regler för hur den används och uppdateras.

#### DATASTYRNING

## Datalivscykeln

**De sex stegen som data genomgår: planera och samla in → lagra och hantera → rensa och bearbeta → analysera och visualisera → dela → arkivera eller förstöra.**

Det osynliga arbetet under ytan utgör ungefär 80 % av allt dataarbete men budgeteras sällan för. Utan förståelse för denna cykel underskattar ledare konsekvent resurser och tid för dataprojekt. Det är i denna process som rådata förädlas till tillförlitligt beslutsunderlag.

#### PROCESSER

## Datakvalitet – de sex dimensionerna

**Datakvalitet mäts längs sex dimensioner: korrekthet (accuracy), fullständighet (completeness), konsekvens (consistency), unikheter (uniqueness), aktualitet (timeliness) och validitet (validity).**

Varje dimension identifierar specifika kostnader och risker. Det gör att man kan gå från vaga påståenden om 'dålig data' till konkreta prioriteringar. Forskning visar att bristfällig data är en av de största dolda kostnaderna i organisationer idag.

#### DATAKVALITET

## Data lineage (dataspårbarhet)

**Dokumentationen av datans hela resa – från ursprunget där den skapades, genom varje systembyte och beräkning, till den rapport som används i styrelserummet.**

Det gör det möjligt att svara på frågan 'var kommer den här siffran ifrån?' när något verkar fel. Utan spårbarhet tvingas medarbetare agera detektiver och förlita sig på gissningar. Med fungerande spårning kan man även förutse konsekvenserna av en systemändring innan den genomförs.

#### DATASTYRNING

## Analytiska metoder

**Fyra metoder som matchar olika typer av beslutsfrågor: deskriptiv, diagnostisk, prediktiv och preskriptiv analys.**

Deskriptiv: 'Vad hände?' (rapporter, dashboards). Diagnostisk: 'Varför hände det?' (samband, rotorsaker). Prediktiv: 'Vad kommer sannolikt att hända?' (prognoser). Preskriptiv: 'Vad bör vi göra?' (optimering, AI-rekommendationer). Att välja fel metod leder organisationen åt fel håll. Varje nivå introducerar en ny grad av osäkerhet – att förstå och kommunicera den osäkerheten är lika viktigt som resultatet i sig.

**ANALYS**

## Generativ AI

**En typ av AI-modell tränad primärt på språk, som kan skriva text, sammanfatta rapporter, tolka dokument och föreslå lösningar utan programmering.**

Det är en assistent som gör analysen tillgänglig – inte en analysnivå i sig. Kvaliteten på svaren är helt beroende av datakvaliteten som matas in. Felaktig eller inkonsekvent data ökar risken för hallucinationer – att AI:n ger övertygande men felaktiga svar.

**AI-VERKTYG**

## Syntetisk data

**Artificiellt genererade dataset som simulerar verkliga mönster, utan att exponera faktisk känslig information.**

Används när historisk data saknas, till exempel vid lansering av nya produkter eller tjänster. AI kan tränas och strategier testas på syntetisk data innan resurser investeras i verkligheten. Förutsättningen är att simuleringen bygger på korrekt kontext och relevant information.

**AI-VERKTYG**

## API (application programming interface)

**Ett standardiserat gränssnitt som gör att olika system kan utbyta data automatiskt.**

Det möjliggör realtidsanalys – data hämtas precis när den behövs. Utan API:er riskerar man att köpa system som låser in datan och omöjliggör automatisering. Som ledare bör man ställa krav på API-stöd redan vid val av leverantör.

**INFRASTRUKTUR**

## Databas, data warehouse och data lake

**Tre olika sätt att lagra data, anpassade för olika ändamål och skala.**

En databas lagrar data för specifika ändamål – operationella transaktioner eller analys. Ett data warehouse kombinerar data från flera källor för strukturerad analys. En data lake lagrar

stora mängder rådata i ursprungligt format för framtida användning. Val av lösning styrs av vad datan ska användas till och budgetens storlek.

#### INFRASTRUKTUR

## Batch-överföring och realtidsdata

**Två sätt att flytta data: i samlat paket vid ett bestämt tillfälle (batch) eller direkt när en händelse sker (realtid).**

Batch-överföring är kostnadseffektivt när omedelbar reaktion inte krävs. Realtidsdata skickas via API:er eller webhooks och krävs när beslut måste fattas omedelbart. Att välja fel metod innebär antingen onödiga kostnader eller för långsam information.

#### INFRASTRUKTUR

## Silos (stuprör)

**Isolerade system eller avdelningar som inte delar information med varandra.**

Resultatet är att sälj, ekonomi och kundtjänst kan ha tre olika och motstridiga bilder av samma kund. Det skapar dubbelarbete, förvirring och bristande tillit till data. Silos är en av de tre grundorsakerna till bristfällig datakvalitet.

#### ORGANISATION

## Data governance (datastyrning)

**Ett system av policys, processer och roller som säkerställer att data används korrekt, lagligt och konsekvent.**

Det handlar om att svara på frågan: 'Vem ansvarar för att den här informationen stämmer?' Det är inte ett IT-projekt – det är ett ledningsansvar. Utan datastyrning är datakvalitet 'allas ansvar', vilket i praktiken innebär ingens ansvar.

#### DATASTYRNING

## KPI (key performance indicator)

**Ett konkret och mätbart nyckeltal som visar om organisationen rör sig mot ett specifikt mål.**

Det är skillnaden mellan vaga ambitioner ('förbättra kundnöjdheten') och skarpa frågor ('öka kundnöjdheten med 15 % för riskanvändare'). KPI:er gör det möjligt att avgöra om ett dataprojekt faktiskt levererat värde. De bör sättas innan ett projekt startar – inte efteråt.

#### UPPFÖLJNING

## Dashboard

**En visuell sammanställning av de viktigaste nyckeltalen i realtid eller nära realtid.**

Den används för operativ övervakning och kan visa nuläge, trender och avvikelser med tydliga signaler. En dashboard är bara lika tillförlitlig som den data som matar den. Syftet är att rätt person ska kunna agera snabbt utan att behöva leta i flera system.

UPPFÖLJNING

## Hallucinationer

**När en AI-modell genererar svar som är övertygande formulerade men faktiskt felaktiga eller påhittade.**

Det sker när modellen matats med felaktiga, ofullständiga eller inkonsekventa data och saknar förmåga att 'veta att den inte vet.' Till skillnad från traditionell analys finns ingen människa som filter i automatiserade AI-processer. Risken sammanfattas som: 'Garbage in, disaster out.'

AI-RISKER

## Digital tvilling

**En virtuell kopia av ett fysiskt objekt, en process eller ett system som uppdateras med realtidsdata.**

Används inom preskriptiv analys för att simulera scenarier och testa beslut utan att påverka verkligheten. Det är ett kraftfullt verktyg inom preskriptiv analys för att optimera utfall. Förutsättningen är hög datakvalitet – annars ger simuleringarna missvisande resultat.

ANALYS

## Datamognadstrappan

**Ett ramverk med fem steg som visar hur väl en organisation använder data för beslut – från manuell och ostrukturerad hantering till datadrivna affärsmodeller.**

Stegen går från Nybörjare (data är svår att hitta och söka) via Utforskare, Utövare och Professionell – där AI varnar proaktivt om avvikelser – till Ledare, där data används innovativt som en del av själva affärsmodellen. Trappan används för att förstå var organisationen befinner sig idag och vad nästa steg konkret innebär. Det vanligaste misstaget är att försöka hoppa direkt till toppen utan att ha säkrat grunden.

PROCESSER

## Databehovs-pyramiden | Aiken-pyramiden

**Ett ledarverktyg som visar i vilken ordning en organisation måste bygga sin dataförmåga – från grundläggande ordning och säkerhet till innovation och AI.**

Pyramidens fyra nivåer är: Ordning (säker lagring och GDPR-efterlevnad), Tillit (ren och verifierad data som beslutsfattare vågar använda), Effektivitet (data flödar sömlöst mellan system och rätt person hittar rätt information i rätt tid) och Innovation och AI (data används som konkurrensfördel och möjliggör nya affärsmöjligheter). Verktuget används för att prioritera rätt investeringar i rätt ordning. Att hoppa direkt till AI utan att ha säkrat tillit och ordning förstärker befintliga misstag i stället för att lösa dem.

### ORGANISATION

## MCP (Model Context Protocol)

**En öppen standard som fungerar som en universell tolk och gör det möjligt för en generativ AI-assistent att kommunicera direkt med flera olika datakällor och verktyg – utan att varje koppling behöver byggas separat.**

MCP ersätter behovet av dyra och tidskrävande specialanpassningar varje gång ett nytt system ska anslutas till en AI-assistent. Det innebär att en organisation kan koppla ihop sin AI med existerande system som CRM, ekonomisystem eller databaser via ett gemensamt gränssnitt. Som ledare är det relevant att ställa frågan om leverantörens system stöder MCP innan avtal skrivs – annars riskerar man att låsa sig till lösningar som inte kan kommunicera med framtidens AI-verktyg.

### INFRASTRUKTUR

## Webhook

**En automatisk notifiering som skickas från ett system till ett annat i realtid när en specifik händelse inträffar – utan att mottagarsystemet behöver fråga om ny data.**

Till skillnad från ett API, där systemet aktivt hämtar data vid behov, fungerar en webhook som ett larm: systemet hör av sig självt när något händer. Ett exempel är att en analys startar automatiskt när en ny kundorder registreras, utan manuell aktivering. Webhooks är kostnadseffektiva för realtidsflöden där händelsen – inte tidpunkten – ska styra när data skickas vidare.

### INFRASTRUKTUR

## MLOps (Machine Learning Operations)

**Ett metodramverk och en uppsättning verktyg för att driftsätta, övervaka och underhålla maskininlärningsmodeller i produktionsmiljö – på ett kontrollerat och repeterbart sätt.**

MLOps är för AI-modeller vad löpande underhåll är för maskiner: utan det försämras modellernas prestanda över tid när verkligheten förändras men modellen inte uppdateras. Ramverket hanterar versionshantering av modeller, automatiserad testning, resursschemaläggning och övervakning av modellkvalitet i drift. Det är ML-ingenjörrens primära arbetsområde och en förutsättning för att maskininläring ska fungera pålitligt i skala – inte bara i en testmiljö.

ARTIFICIELL INTELLIGENS

## Bias

**En systematisk bias i en AI-modells slutsatser som uppstår när träningsdatan inte representerar verkligheten rättvist – och som leder till felaktiga eller orättvisa beslut i drift.**

Bias uppstår när historisk data speglar mänskliga fördomar, strukturella ojämlikheter eller blinda fläckar i datainsamlingen – och modellen lär sig att förstärka dessa mönster i stället för att korrigera dem. Till skillnad från en hallucination, där AI:n hittar på ett svar, är bias mer subtilt: svaret kan verka rimligt men systematiskt missgynna vissa grupper eller situationer. Det är en av de viktigaste riskerna att adressera redan i fas för datainsamling och modelldesign, inte i efterhand.

AI-RISKER

## Datakatalog

**Ett sökbart register över organisationens datamängder som beskriver vad som finns, var det lagras, vem som äger det och hur det får användas.**

Datakatalogen är datalivscykeln interna innehållsförteckning – den gör det möjligt för medarbetare att snabbt hitta rätt data utan att behöva fråga runt i organisationen. Den innehåller typiskt metadata om varje datamängd: ursprung, uppdateringsfrekvens, ansvarig dataägare och eventuella användningsbegränsningar som GDPR-krav. Utan en datakatalog förblir stora delar av organisationens data osynlig och oanvänd, trots att den tekniskt sett finns tillgänglig.

INFRASTRUKTUR

## Data literacy (dataläskunnighet)

**Förmågan att läsa, förstå, ifrågasätta och kommunicera med data på ett meningsfullt sätt – oavsett teknisk bakgrund.**

Data literacy handlar inte om att kunna programmera eller bygga modeller, utan om att förstå vad en siffra faktiskt mäter, när den är tillförlitlig och när den inte är det – och att våga ställa de frågorna högt. En organisation med hög data literacy har medarbetare som kan tolka en dashboard, förstå osäkerheten i en prognos och skilja på en trend och en slumpmässig variation. Det är den kulturella förutsättningen för att datadrivna beslut ska bli verklighet – utan den riskerar även korrekta analyser att missförstås, ignoreras eller övertolkas.

ORGANISATION

# Rollbeskrivning

Investera i rätt roll vid rätt tidpunkt för att optimera organisationens tid och budget. Vad kan du förvänta dig av olika dataroller? Nedan beskrivs olika dataroller inom data, datastyrning, analys och AI, och deras kompetenser.

## Dataarkitekt | Data Architect

**Ritar upp den övergripande strukturen för hur data ska organiseras, lagras och flöda genom organisationens hela tekniska landskap.**

Dataarkitekten arbetar i planeringsfasen av datalivscykeln och fattar beslut om datamodeller, lagringslösningar och integrationsmönster som påverkar alla efterföljande steg. Det är en strategisk roll som avgör om den tekniska infrastrukturen håller när verksamheten växer eller ändrar riktning. Felaktiga arkitekturbeslut är kostsamma att rätta till i efterhand och kan låsa organisationen i ineffektiva lösningar i flera år.

DATA

## Dataingenjör | Data Engineer

**Bygger och underhåller de tekniska system och datapipelines som transporterar data från källsystem till lagring och vidare för bearbetning.**

Dataingenjören arbetar i datalivscykeln tidiga faser – samla in, lagra och hantera samt rensa och bearbeta – och säkerställer att data flödar tillförlitligt och effektivt genom organisationens infrastruktur. Utan en dataingenjör saknar analytiker och AI-system tillgång till den data de behöver. Rollen kräver kompetens inom databashantering, molntjänster och programmeringsspråk som Python eller SQL.

DATA

## Dataägare | Data owner

**En ledare med mandat och budget som ansvarar ur ett verksamhetsperspektiv för att en specifik datamängd är korrekt och tillförlitlig och hanteras enligt organisationens regler – inte en teknisk roll utan en ledarskapsfunktion.**

Dataägaren är vanligtvis en verksamhetsansvarig chef, som är aktiv genom hela datalivscykeln men har särskilt ansvar för att data hanteras och arkiveras korrekt samt att tillgång till data styrs rätt. Rollen kräver djup förståelse för affärsprocesserna kopplade till datamängden. Dataägaren ställer krav på underlag, identifierar avvikelser och ser till att fel åtgärdas. När konflikter uppstår kring datakvalitet/användning, säkerställer dataägaren att

data följer regulatoriska krav som GDPR. Utan tydliga dataägare uppstår ansvarsluckor där ingen äger frågan om datakvalitet och relefterlevnad.

#### DATASTYRNING

## **Datastyrningsspecialist | Data Governance Specialist**

**Skapar och förvaltar de regler, riktlinjer och processer som styr hur organisationens data används, skyddas och kvalitetssäkras.**

Datastyrningsspecialisten arbetar tvärs över hela datalivscykeln och säkerställer att varje steg – från insamling till arkivering – sker i enlighet med lagkrav som GDPR, interna policyer och branschstandarder. Rollen bygger de strukturer som gör det möjligt för dataägare, ingenjörer och analytiker att arbeta med gemensamma definitioner och kvalitetskrav. Utan datastyrning riskerar organisationen inkonsekvent data, regelöverträdelser och beslut baserade på otillförlitligt underlag.

#### DATASTYRNING

## **Dataförvaltare | Data Steward**

**Utför det operativa, dagliga arbetet med att säkerställa att data håller den kvalitet och följer de regler som organisationen har beslutat – bryggan mellan datastyrningens regler och verklighetens data.**

Dataförvaltaren arbetar primärt i datalivscykeln's faser lagra och hantera samt rensa och bearbeta, och är den roll som i praktiken kontrollerar, korrigerar och dokumenterar avvikelser i datakvalitet. Där dataägaren har det formella ansvaret och datastyrningsspecialisten sätter ramverket, är det dataförvaltaren som ser till att reglerna faktiskt efterlevs i det dagliga arbetet. Utan dataförvaltare riskerar organisationen att ha välformulerade policyer som aldrig omsätts i praktiken.

#### DATASTYRNING

## **Dataanalytiker | Data Analyst**

**Omvandlar bearbetad data till konkreta insikter och beslutsunderlag genom analys, statistik och visualisering.**

Dataanalytikern arbetar primärt i datalivscykeln's senare faser – analysera, visualisera och dela – och är den roll som gör data begriplig för beslutsfattare. Rollen kräver kompetens inom verktyg som SQL, Excel och Power BI samt förmågan att förstå verksamhetens frågor och översätta dem till analytiska undersökningar. Det är dataanalytikern som identifierar trender, avvikelser och mönster som annars förbli dolda i stora datamängder.

#### ANALYS

## Data Scientist

**Kombinerar statistik, maskininlärning och affärsförståelse för att bygga prediktiva modeller som svarar på komplexa frågor som traditionell analys inte kan svara på.**

En Data Scientist arbetar i datalivscykels mittfaser – rensa och bearbeta samt analysera – med tyngdpunkt på att utveckla modeller som kan förutsäga utfall eller klassificera information. Rollen kräver kompetens inom statistik, programmeringsspråk som Python och förståelse för maskininlärningsalgoritmer. En datavetare utan tillgång till ren och strukturerad data – som dataingenjören levererar – kan inte producera tillförlitliga modeller.

DATA ANALYS ARTIFICIELL INTELLIGENS

## ML-ingenjör | Machine Learning (ML) Engineer

**Tar färdigutvecklade maskininlärningsmodeller och driftsätter dem i produktionsmiljö så att de fungerar tillförlitligt i verkliga system och i stor skala.**

ML-ingenjören arbetar i datalivscykels senare faser – från bearbetad data till driftsättning och delning av modellresultat – och är bryggan mellan Data Scientist experimentmiljö och organisationens faktiska IT-system. Rollen kräver kompetens inom mjukvaruutveckling, molninfrastruktur och MLOps – de metoder och verktyg som används för att övervaka och underhålla modeller i drift. Utan ML-ingenjören stannar maskininlärning på prototypstadiet och når aldrig verksamhetens slutanvändare.

ARTIFICIELL INTELLIGENS

## Generativ AI-ingenjör | Generative AI Engineer

**Utvecklar och integrerar lösningar baserade på generativa AI-modeller – som stora språkmodeller (LLM) – i organisationens produkter, processer och arbetsflöden.**

Den generativa AI-ingenjören arbetar i datalivscykels driftsättnings- och delningsfas och ansvarar för tekniker som prompt engineering, retrieval-augmented generation (RAG) och finjustering av modeller på organisationens egna data. Rollen kräver förståelse för hur stora språkmodeller fungerar, deras begränsningar och de risker som följer med genererat innehåll – som felaktigheter och bias. Det är en relativt ny roll och kompetensområdet utvecklas snabbt, vilket ställer krav på kontinuerligt lärande och kritisk bedömning av vad tekniken faktiskt kan leverera.

ARTIFICIELL INTELLIGENS