

# DATA FÖRST – AI SEN

Ett frukostseminarium om att ha sina data redo





TILLSAMMANS FRIGÖR VI KRAFTEN I INFORMATION

| Vi utvecklar verksamheter,  
system och kompetens



VI HJÄLPER DIG I ARBETET MED ATT UTVECKLA VERKSAMHETEN

# IRM erbjuder expertstöd genom verksamhetsutvecklingens alla faser

- Informationsarkitektur
- Verksamhetsarkitektur
- Processutveckling
- Förändringsledning
- UX- och UI-design
- Systemutveckling
- Lösningsarkitektur
- + Kurser och utbildningar



# Utbildningar och kurser

## Certifieringar

- Certifierad verksamhetarkitekt
- Certifierad informationsarkitekt
- Certifierad verksamhetsutvecklare

## Kurser

- Informationsmodellering
- Processutveckling
- Vintergatan – skapa din verksamhetskarta



### Certifierad verksamhetsutvecklare

Certifierad verksamhetsutvecklare är en utbildning som ger dig förmågan att driva framgångsrika utvecklings- och förändringsinitiativ utan att tappa fokus på...



### Certifierad verksamhetsarkitekt

Certifierad verksamhetsarkitekt - Sveriges största utbildning av verksamhetsarkitekter. Vår utbildning Certifierad verksamhetsarkitekt ger dig kompetensen att skapa ett tillstånd...



### Certifierad informationsarkitekt

Gör informationen tillgänglig och användbar för de behov verksamheten har. Certifierad informationsarkitekt ger dig verktyg för att beskriva och analysera verksamhetens...



im.se/kurser



### Processutveckling

Processutveckling! Lär dig ett och utveckla processer agilt utifrån en önskad översikt.



### Vintergatan - skapa din verksamhetskarta

Skapa en Vintergatan - skapa din verksamhetskarta. Inom området tydliggör ber...

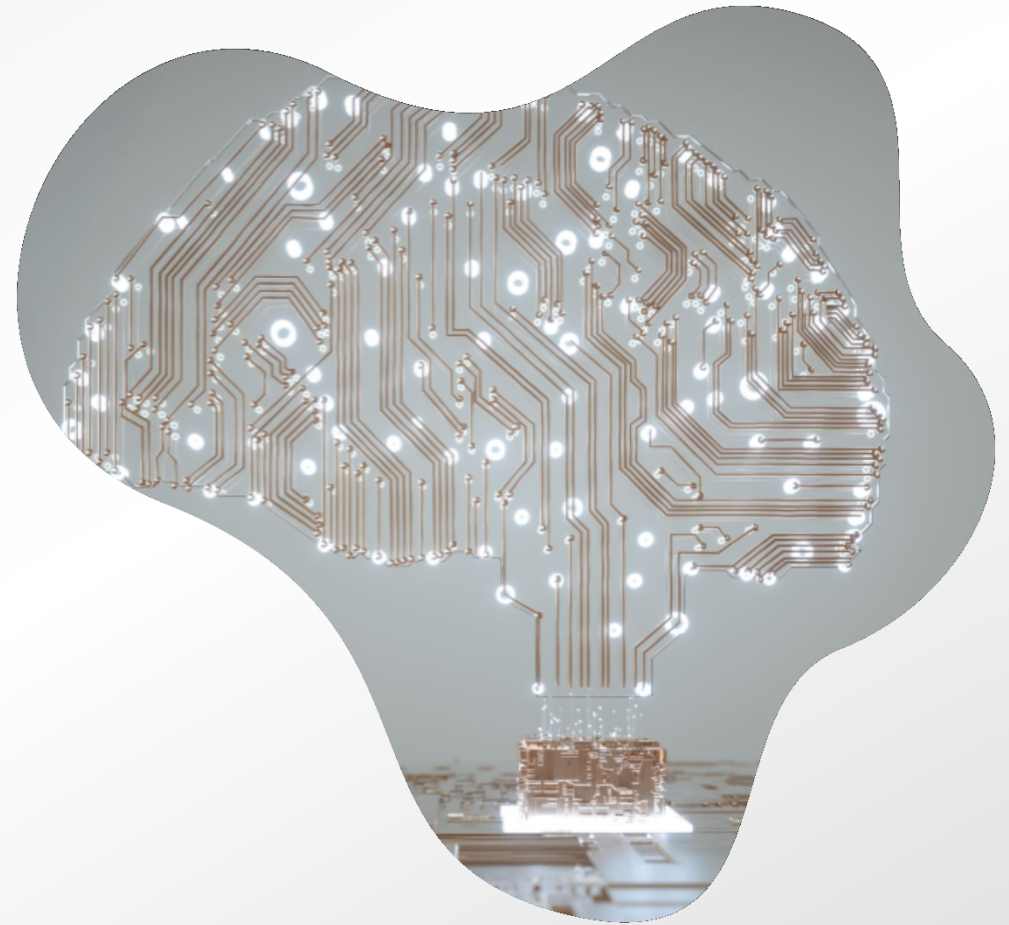


Vi kommer inte prata  
om vad  
**AI kan göra,**  
utan om vad  
**AI behöver** för att  
göra det den gör

# Data är för AI vad erfarenhet är för människor

Vi människor lär oss av det vi ser, hör och känner. Vi reagerar känslomässigt på vad som händer runt omkring oss.

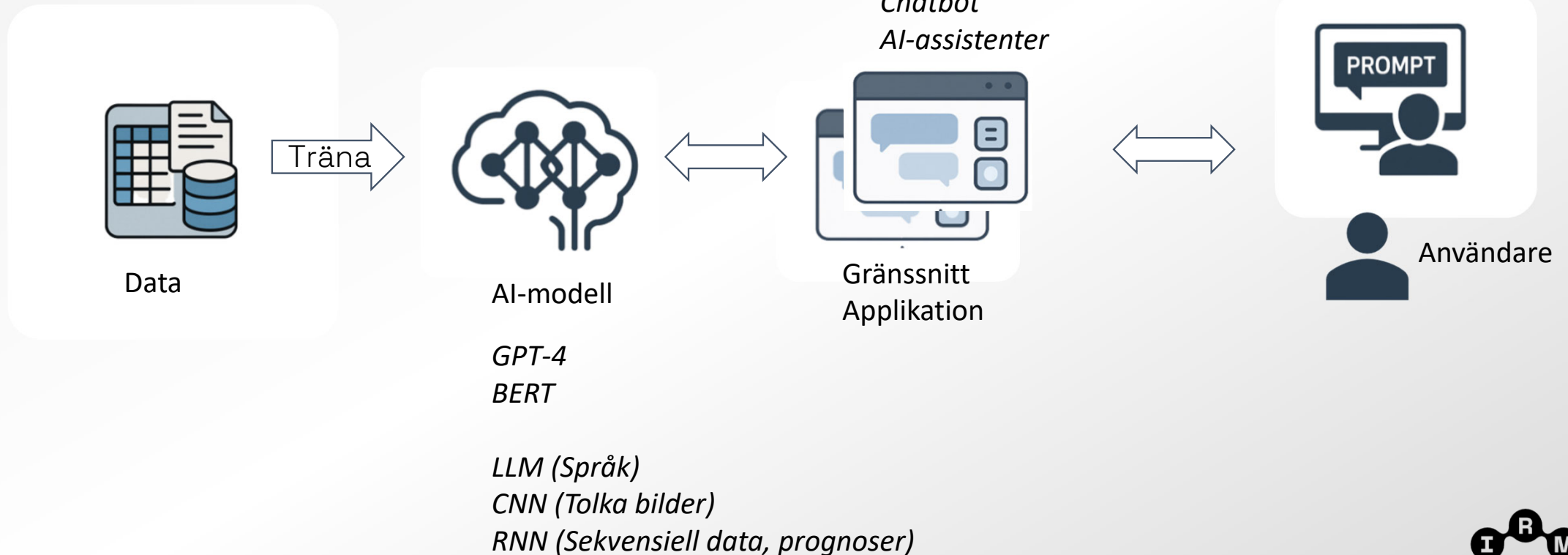
AI:n har inget medvetande, den förstår inte. Den räknar, sorterar och gissar vad som passar bäst utifrån tidigare exempel.



# AI-EKOSYSTEM – ÖVERSIKT

Räcker inte med *mycket* data, det måste också vara:

- **relevant** - annars lär sig modellen fel saker
- **strukturerad och kvalitetssäkrad** - annars fattas fel beslut
- **representativ** - annars riskerar vi snedvridna resultat och bias.



---

# Data management i en AI-satsning

---

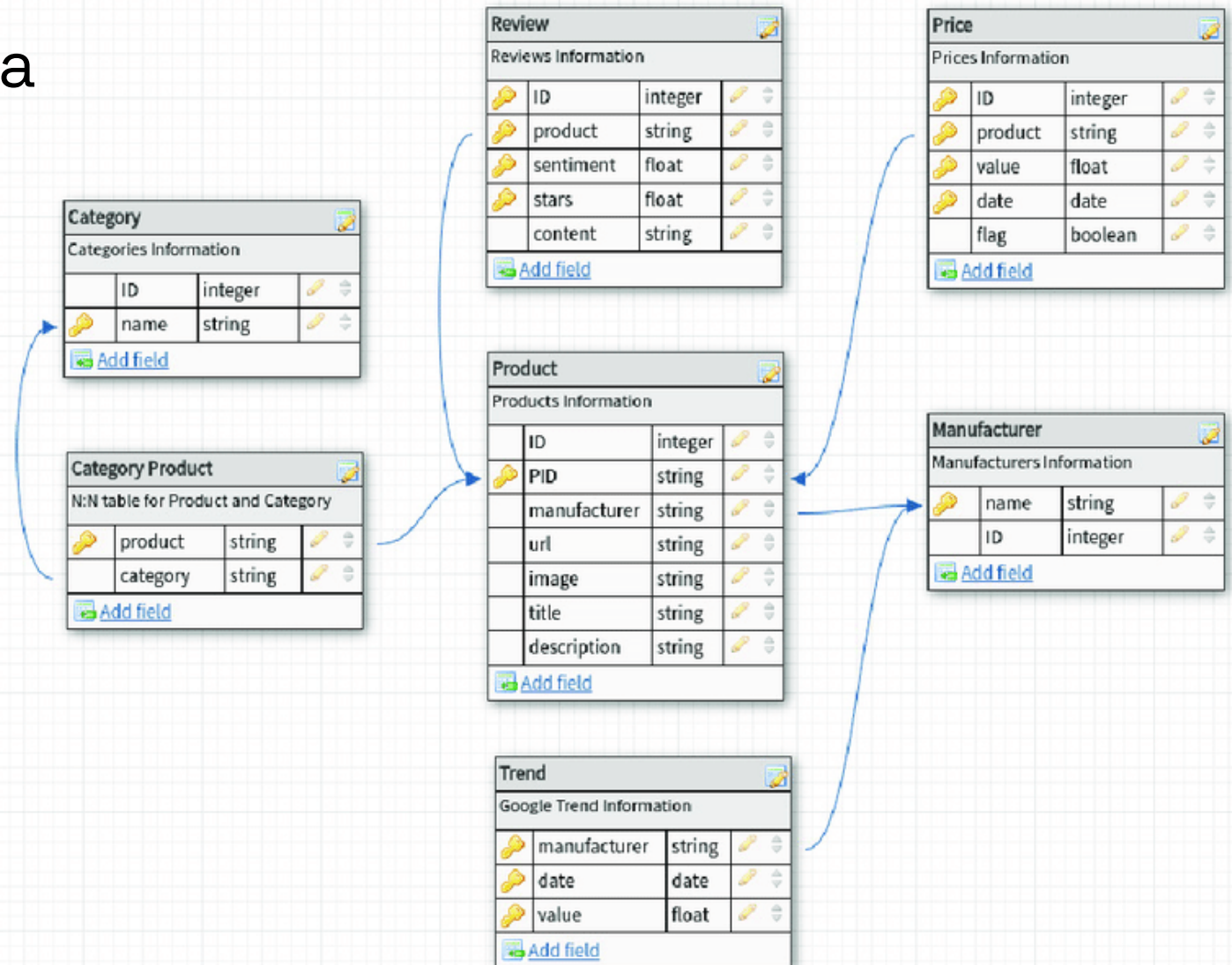
Datakategorier:

# Strukturerade data

Data strukturerade i tabeller, poster eller fält.

Exempel på tekniker/tillämpningar:

- Maskininlärning (Machine learning, ML).



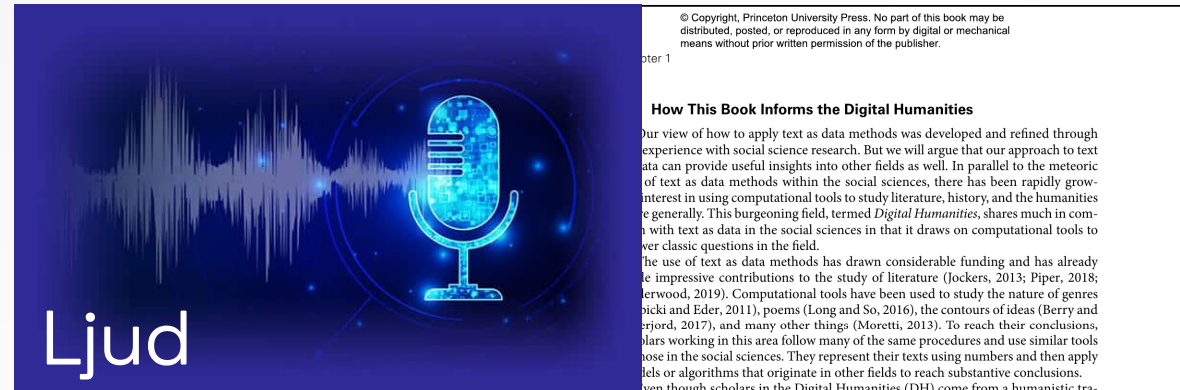
Datakategorier:

# Ostrukturerade data

Data som inte är strukturerade i tabeller, poster eller fält, till exempel text, bild, ljud eller video.

Exempel på tekniker/tillämpningar:

- Naturlig språkbehandling (Natural language processing, NLP)
- Datorseende (Computer vision)



Datakategorier:

## Semistrukturerade data

Kombination av strukturerade och ostrukturerade data, organiserat på någon nivå.

Exempel på tekniker/tillämpningar:

- Webbskrapning (Web scraping)
- Överföring av data

### Example

```
{  identifier: "fire",  
  label: "coordinates",  
  items: [  
    { fire: "1",  
      coordinates: "43.459,-71.881",
```

Identifier

```
<?xml version="1.0" encoding="UTF-8"?>  
- <EmployeeData>  
  - <employee id="34594">  
    <firstName>Heather</firstName>  
    <lastName>Banks</lastName>  
    <hireDate>1/19/1998</hireDate>  
    <deptCode>BB001</deptCode>  
    <salary>72000</salary>  
  </employee>  
  - <employee id="34593">  
    <firstName>Tina</firstName>  
    <lastName>Young</lastName>  
    <hireDate>4/1/2010</hireDate>  
    <deptCode>BB001</deptCode>  
    <salary>65000</salary>  
  </employee>  
</EmployeeData>
```

XML

```
"43.459",  
"-71.881",  
"/7/2009",  
",  
"T",  
70
```

Values

JSON

Typer av data:

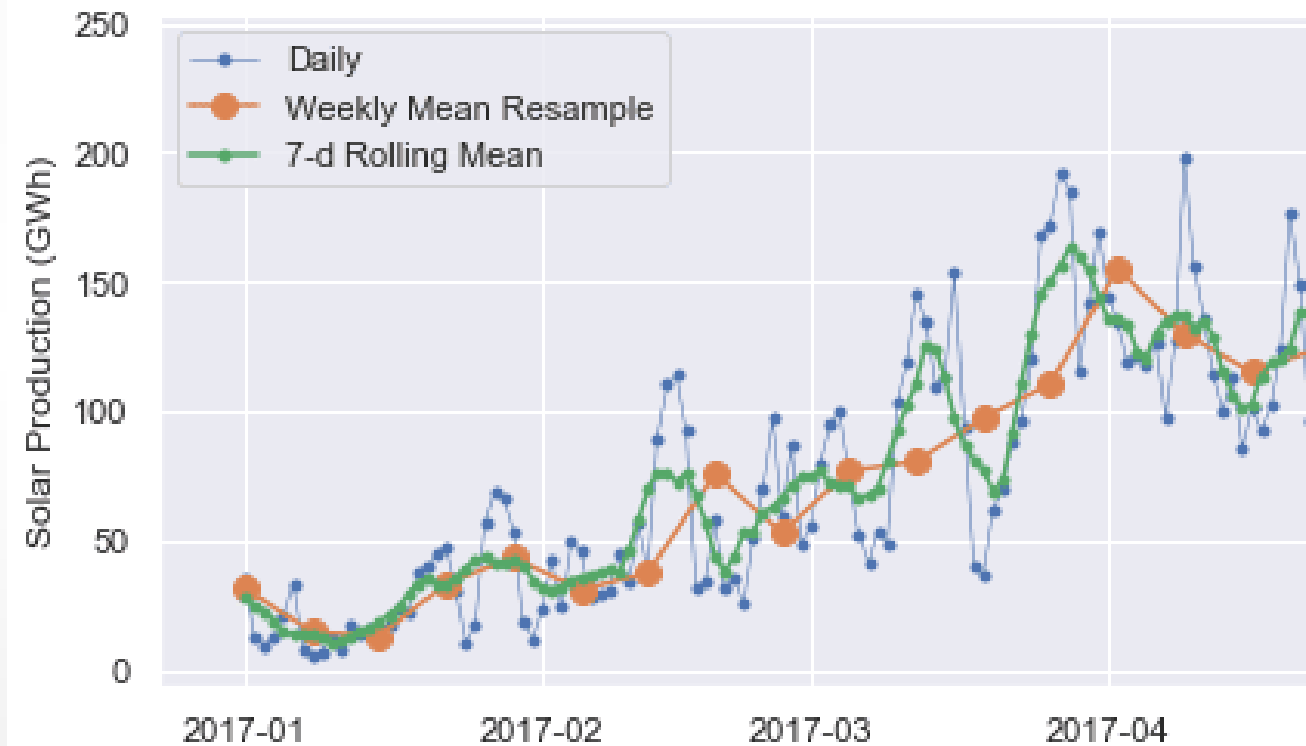
## Tidsseriedata

Registrering av observationer  
med regelbundna tidsintervall.

Exempel på tekniker/tillämpningar:

- Finans
- Sakernas internet  
(Internet of things, IoT)
- Hälsovård
- Tillverkning

För att upptäcka avvikelser och  
trender och prognostisera.



Typer av data:

# Textdata

Alla former av skrivna texter, till exempel artiklar, dokument, e-post.

Exempel på tekniker/tillämpningar:

- Naturlig språkbehandling (Natural language processing, NLP) för att analysera känslolägen
- Klassificering och sammanfattning av texter
- Begreppsigenkänning
- Översättning
- Chatbots

© Copyright, Princeton University Press. No part of this book may be distributed, posted, or reproduced in any form by digital or mechanical means without prior written permission of the publisher.

8 Chapter 1

## 1.2 How This Book Informs the Digital Humanities

Our view of how to apply text as data methods was developed and refined through our experience with social science research. But we will argue that our approach to text as data can provide useful insights into other fields as well. In parallel to the meteoric rise of text as data methods within the social sciences, there has been rapidly growing interest in using computational tools to study literature, history, and the humanities more generally. This burgeoning field, termed *Digital Humanities*, shares much in common with text as data in the social sciences in that it draws on computational tools to answer classic questions in the field.

The use of text as data methods has drawn considerable funding and has already made impressive contributions to the study of literature (Jockers, 2013; Piper, 2018; Underwood, 2019). Computational tools have been used to study the nature of genres (Rybicki and Eder, 2011), poems (Long and So, 2016), the contours of ideas (Berry and Fagerjord, 2017), and many other things (Moretti, 2013). To reach their conclusions, scholars working in this area follow many of the same procedures and use similar tools to those in the social sciences. They represent their texts using numbers and then apply models or algorithms that originate in other fields to reach substantive conclusions.

Even though scholars in the Digital Humanities (DH) come from a humanistic tradition, we will show how the goals of their analysis fit well within the framework of our book. And as a result, our argument about how to use text as data methods to make valid inferences will cover many of the applications of computational tools in the humanistic fields. A major difference between DH and the social sciences is that digital humanists are often interested in inferences about the particular text that is being studied, rather than the text as an indicator of some other, larger process. As a result, digital humanities have thus far tended to focus on the discovery and measurement steps of the research process, while devoting less attention to making causal inferences or predictions. Digital humanists use their large corpora to make new and important discoveries about organizations in their texts. They then use tools to measure the prevalence of those quantities, to describe how the prevalence of the characteristics has changed over time, or to measure how well defined a category is over time.

As with any field that rises so suddenly, there has been considerable dissent about the prospect of the digital humanities. Some of this dissent lies well outside of the scope of our book and focuses on the political and epistemological consequences of opening up the humanities to computational tools. Instead we will engage with other critiques of digital humanities that stipulate to the “rules” laid out in computational papers. These critics argue that the digital humanities is not capable of achieving the inferential goals it lays out and therefore the analysis is doomed from the start. A recent and prominent objection comes from Da (2019), who summarizes her own argument as,

In a nutshell the problem with computational literary analysis as it stands is that what is robust is obvious (in the empirical sense) and what is not obvious is not robust, a situation not easily overcome given the nature of literary data and the nature of statistical inquiry. (Da, 2019, p. 45)

Da (2019)’s critique goes to the heart of how results are evaluated and relies heavily on procedures and best practices imported from computer science (as does, it is worth noting, much of the work she is critiquing). As we have argued above, directly

For general queries, contact [webmaster@press.princeton.edu](mailto:webmaster@press.princeton.edu)



Typer av data:

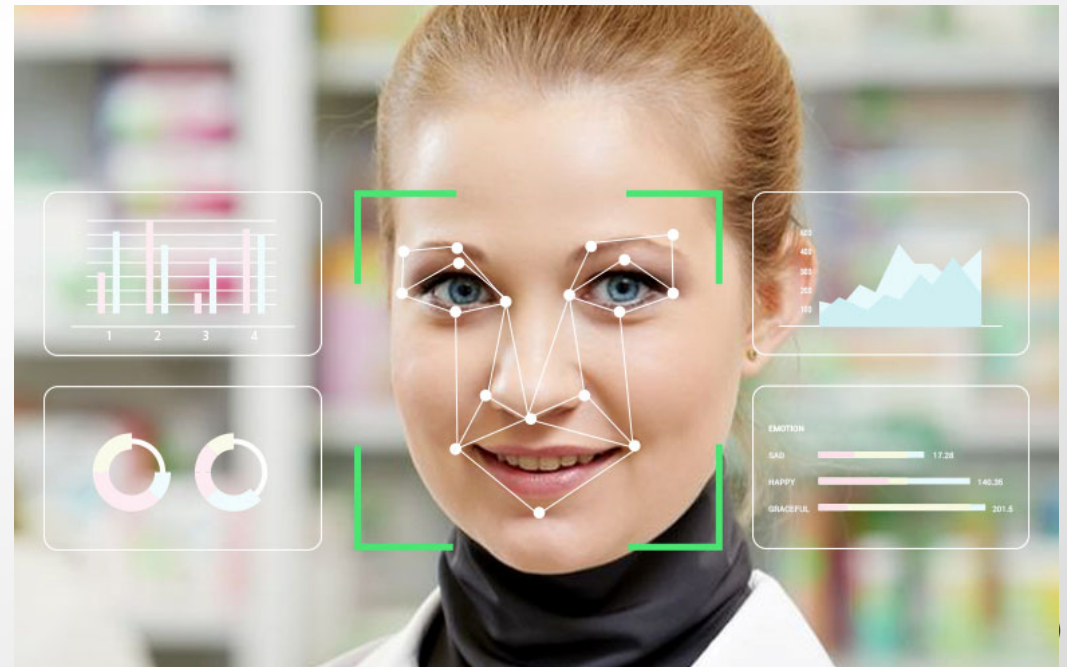
# Bilddata

Visuell information som fångas av kameror eller bildbehandlingsteknik, till exempel foton, medicinska bilder, satellitbilder.

Exempel på tekniker/tillämpningar:

Uppgifter inom datorseende (computer vision) för att:

- Identifiera objekt
- Klassificera bilder
- Ansiktsigenkänning
- Medicinsk analys
- Styrning av självkörande fordon
- Analys av satellitbilder



Typer av data:

## Ljuddata

Ljudinformation som fångas av mikrofoner eller ljudsensorer, till exempel inspelat tal, musik och omgivningsljud.

Exempel på tekniker/tillämpningar:

- Taligenkänning (speech recognition)
- Identifiering av talare
- Analys av känslolägen
- Analys av musik
- Röstassistans
- Avlyssning



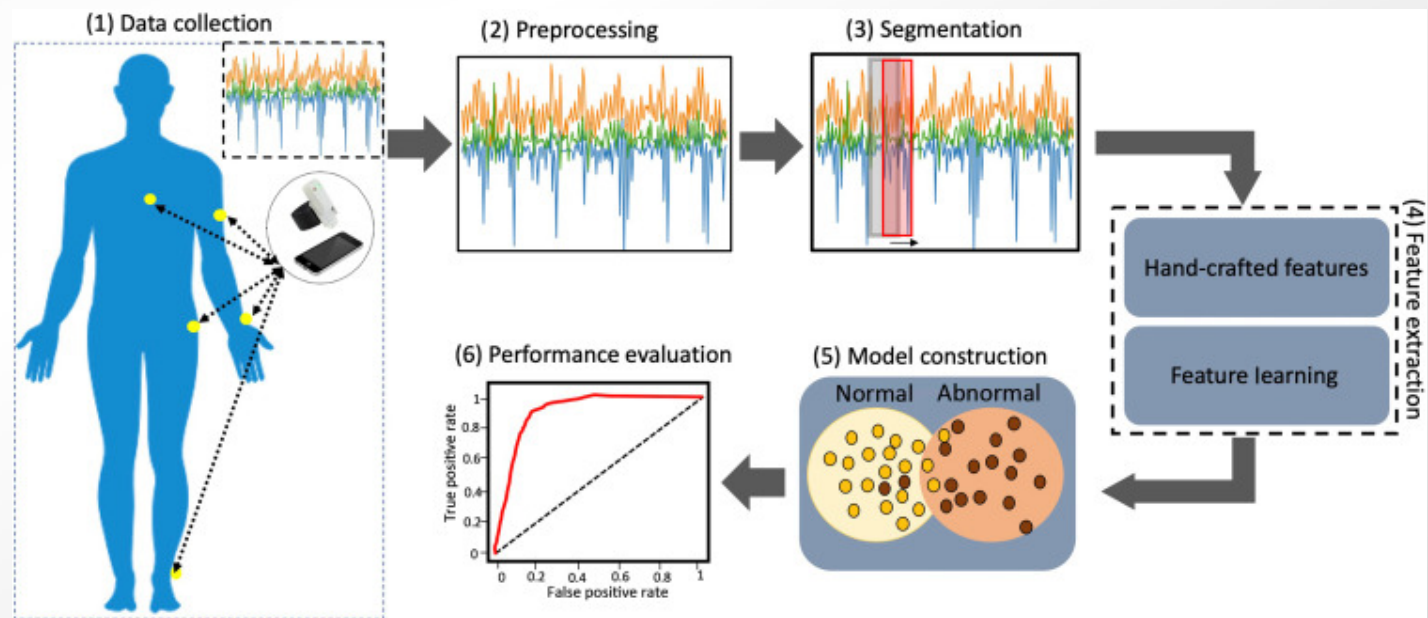
Typer av data:

# Sensordata

Data från fysiska givare, till exempel temperaturgivare, accelerometrar, gyroskop.

Exempel på tekniker/tillämpningar:

- Sakernas internet  
(Internet of things, IoT)
- Bärbara sensorer  
(wearables)
- Industriautomation
- Smarta byggnader
- Hälsoövervakning
- Miljötillsyn



Typer av datakällor:

## Publika datakällor

Ger data som är fritt tillgängliga för forskning, analys och undervisning.

Exempel på innehåll/tekniker/tillämpningar:

- Maskininlärning (Machine learning, ML)
- Benchmarking av algoritmer
- Prediktiva modeller
- Trendanalyser



Typer av datakällor:

## Interna källor

För data som verksamheten själv genererar eller samlar in genom sin drift, som transaktioner eller interaktioner med kunder, medarbetare eller partners.

Exempel på innehåll/tekniker/  
tillämpningar:

- Kunddata
- Försäljningssiffror
- Finansiella uppgifter
- Driftsdata



Typer av datakällor:

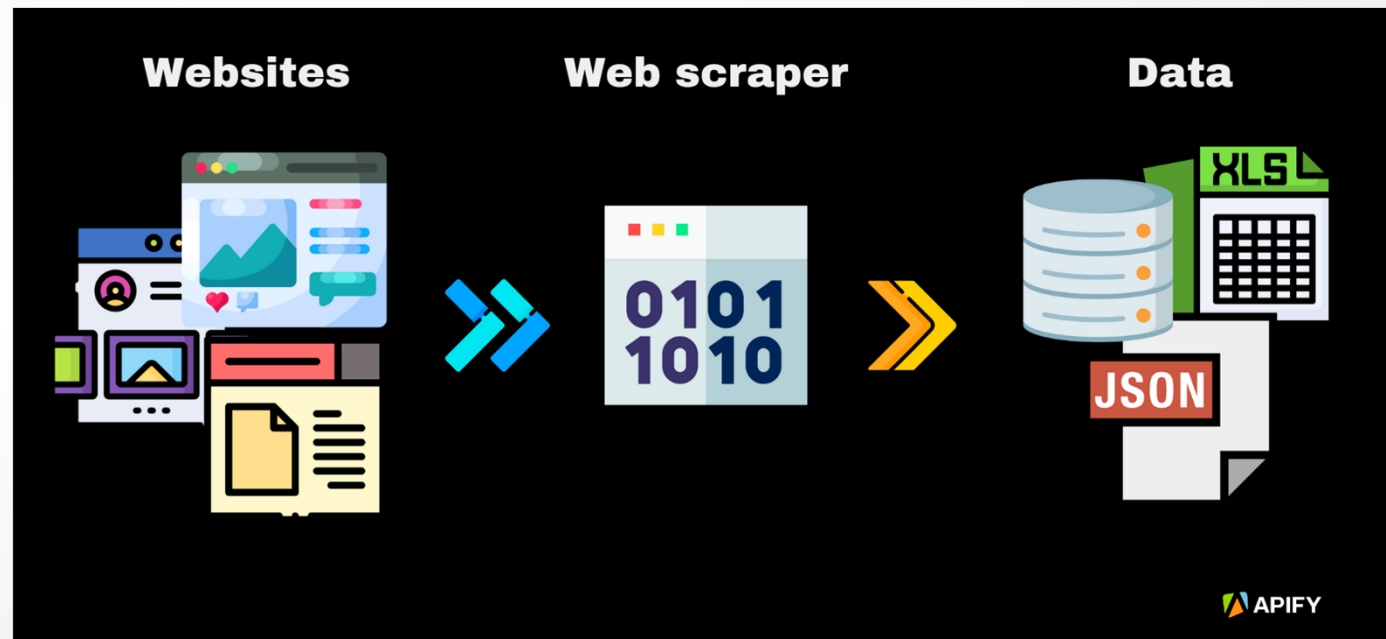
## Webbskrapning

Data från webbplatser som fångats automatiskt. Kan vara text, bilder, länkar, metadata och ibland även strukturerade data.

Exempel på innehåll/tekniker/  
tillämpningar:

- Produktinformation
- Nyheter
- Jobbannonser
- Innehåll i sociala media

För analys och bevakning.



Typer av datakällor:

## API:er och online-tjänster

Ger enklare åtkomst till färdigbearbetat data från många slags bakomliggande datakällor som sociala media (exempel: X/Twitter API) eller finansiella informationstjänster (exempel: Bloomberg API).

Exempel på innehåll/tekniker/tillämpningar:

- Diverse tillämpningar där man kan dra nytta av färdigbearbetade data av skilda slag.

BURL US Equity

Export

Inflection

KPI Correlation

Trend Analysis

Alternative Data Metrics Summary - Burlington Stores Inc

Data up to 2023-07-05

About the Data

|                           | Trailing YoY Growth % |        |       | Trailing PoP Growth % |        |       |
|---------------------------|-----------------------|--------|-------|-----------------------|--------|-------|
| Alt Data Metrics          | 91 Day                | 28 Day | 7 Day | 91 Day                | 28 Day | 7 Day |
| Bloomberg Second Measure  |                       |        |       |                       |        |       |
| Observed Sales            | 2.57                  | 5.34   | 7.64  | 7.27                  | -8.20  | 0.06  |
| Observed Transactions     | 5.88                  | 9.14   | 10.92 | 6.22                  | -5.71  | -0.34 |
| Observed Customers        | 3.22                  | 7.08   | 10.33 | 4.11                  | -5.29  | 0.67  |
| Average Transaction Value | -3.12                 | -3.49  | -2.96 | 0.99                  | -2.64  | 0.39  |
| Transactions per Customer | 2.50                  | 1.60   | 0.92  | 2.50                  | -0.78  | -0.90 |
| Sales per Customer        | -0.63                 | -1.63  | -2.44 | 3.04                  | -3.08  | -0.61 |
| Placer.ai                 |                       |        |       |                       |        |       |
| Estimated Visits          | 6.61                  | 12.21  | 15.28 | 10.70                 | -4.68  | -1.99 |

Bloomberg Second Measure:

U.S. consumer transaction data  
Panel: Millions of consumers  
Geo: Mid-west > coasts  
Card Type: Debit > credit

Placer.ai:

U.S. consumer foot-traffic data  
Panel: 25 million mobile devices  
Stores: Outdoor > indoor  
Operating Systems: iOS > Android

Observed Sales YoY Growth %

Comp Source

Analyst Curated (BI)

Growth

Year-over-Year

Period

Weekly

3M

6M

1Y

2Y

3Y

5Y

Max

2023

| Week Ending                   | 16-Apr | 23-Apr | 30-Apr | 07-May | 14-May | 21-May | 28-May | 04-Jun | 11-Jun | 18-Jun | 25-Jun | 02-Jul |
|-------------------------------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|
| Analyst Curated (BI) - Median | 5.16   | -0.85  | -5.41  | -6.85  | 13.49  | 2.40   | 0.65   | 0.15   | 5.47   | 2.28   | 4.40   | 7.13   |
| Burlington Stores Inc         | 3.49   | -5.40  | -8.22  | -0.85  | 19.17  | 6.96   | 7.22   | 0.86   | 10.93  | 2.90   | 1.84   | 7.15   |
| TJX Cos Inc/The               | 11.05  | 1.05   | -0.79  | -5.89  | 17.31  | 4.65   | 4.71   | 7.25   | 6.55   | 4.99   | 9.10   | 9.35   |
| Ross Stores Inc               | 6.83   | 1.53   | -2.60  | -7.82  | 9.66   | 0.15   | -3.40  | -0.57  | 4.40   | 1.66   | 6.96   | 7.11   |
| Citi Trends Inc               | -23.45 | -2.76  | -13.30 | -23.68 | -14.27 | -26.09 | -8.81  | -11.11 | 0.82   | -15.47 | -6.27  | -15.99 |

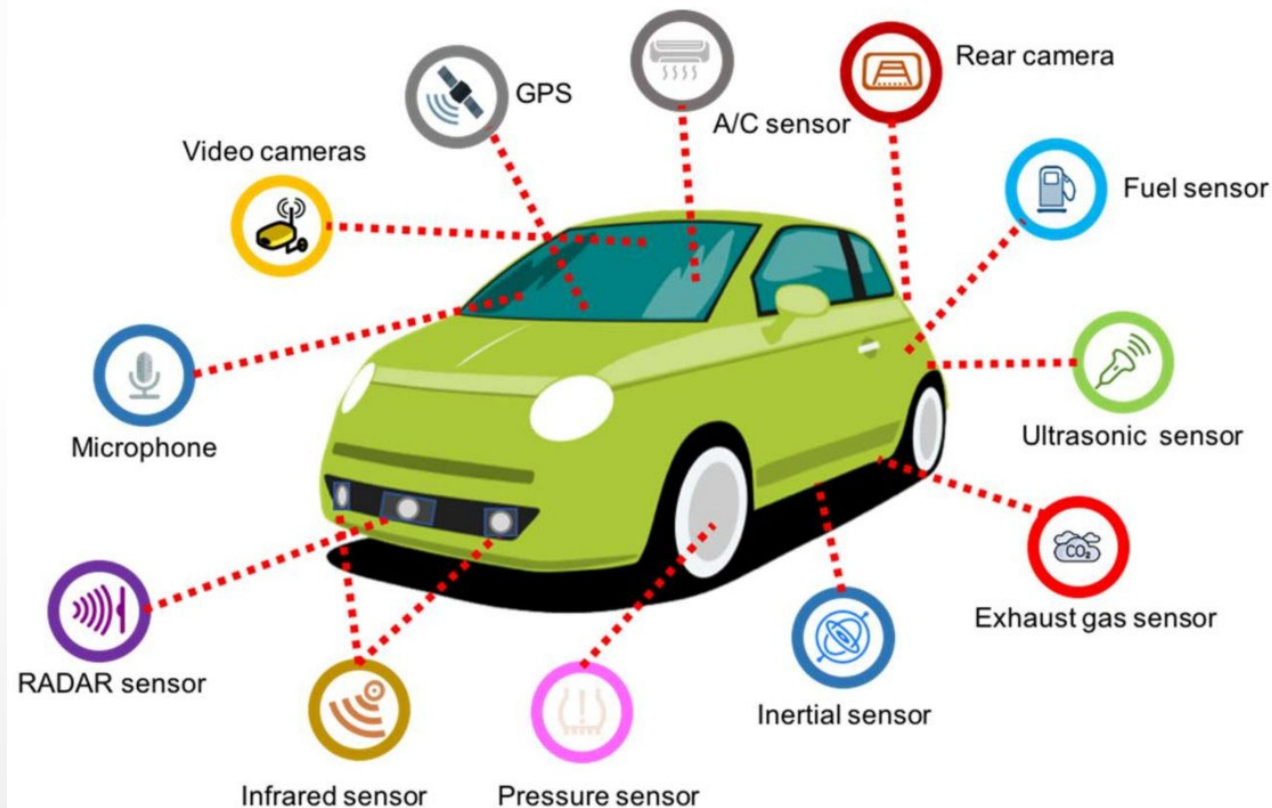
Typer av datakällor:

# Sensorer

Samlar data som temperatur, ljud, fuktighet, rörelser, ljud med mera från fysiska objekt och miljöer.

Exempel på innehåll/tekniker/  
tillämpningar:

- Smarta hem, byggnader och miljöer
- Industriautomation
- Hälsomonitorering
- Miljöbevakning



Typer av datakällor:

# Crowdsourcing

En större grupp människor som engagerats att utföra någon uppgift eller samla data, ofta via någon app eller online-plattform.

Exempel på innehåll/tekniker/tillämpningar:

- Klassificering av data
- Moderering av innehåll
- Analys av känslolägen i innehåll
- Återkoppling från användare



Typer av datakällor:

# Transaktionssystem

Register med transaktioner eller andra interaktioner mellan parter.

Exempel på innehåll/tekniker/  
tillämpningar:

- Köp
- Bokningar
- mm

Denna typ av data är en kärna då det gäller kundanalyser, rekommendationssystem, monitorering för upptäckt av bedrägerier och personifierad marknadsföring.



Typer av datakällor:

# Multimedia

Innehåller kombinationer av digitala mediatyper som bilder, video, ljud och animeringar.

Exempel på innehåll/tekniker/  
tillämpningar:

- Innehållsrekommendationer
- Övervakning
- Bildigenkänning
- Förstärkt verklighet  
(augmented reality, AR)
- Underhållning



Typer av datakällor:

## Sociala media och webbforum

Ger användargenererat innehåll.

Exempel på innehåll/tekniker/  
tillämpningar

Analys av känslolägen, eller trender, för:

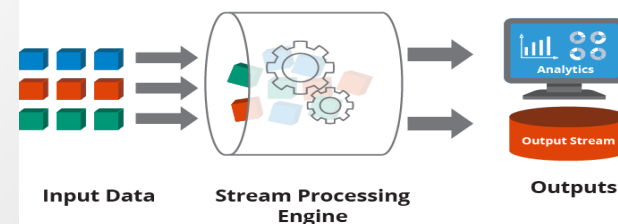
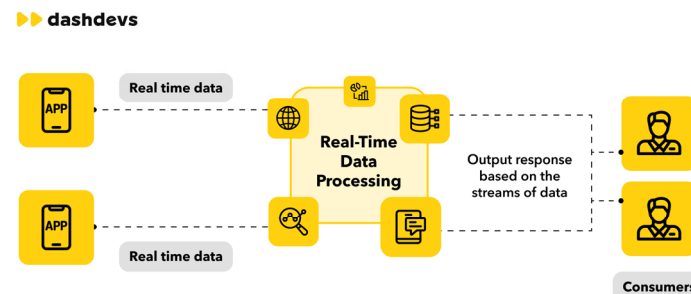
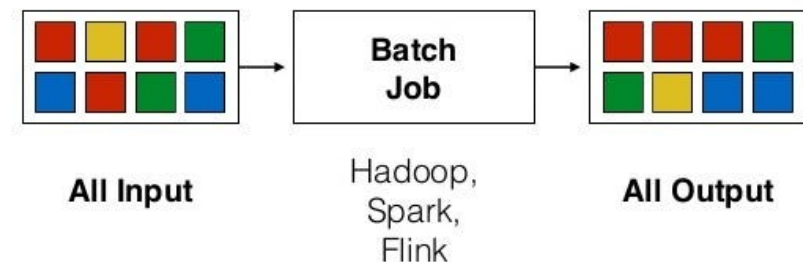
- Varumärkesbevakning
- Krishantering
- Forskning om sociala nätverk



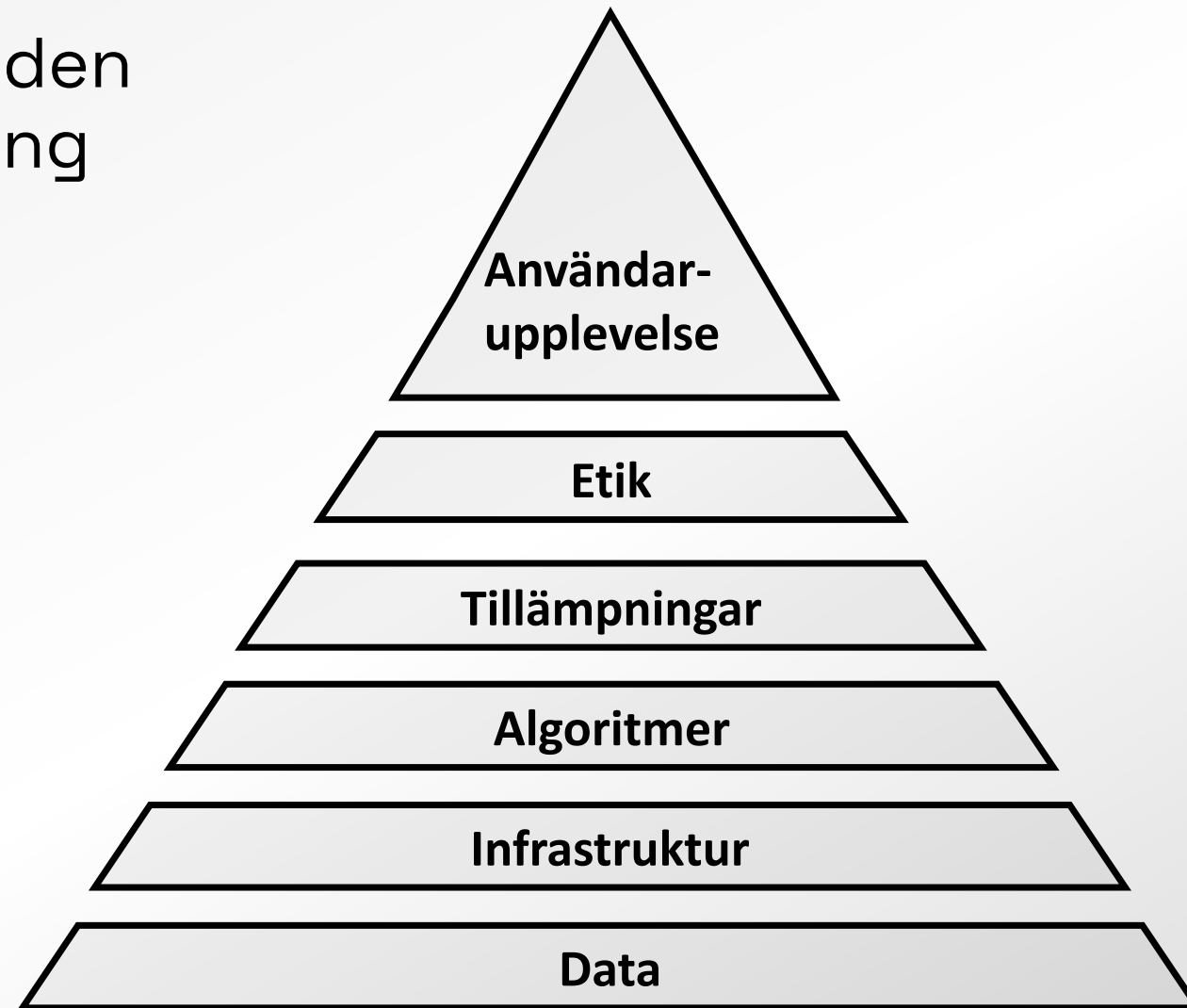
# Leveransmodeller

Hur data levereras och hanteras med avseende på tid

- **Batchbearbetning** (Batch processing)  
Bearbetar volymer av data vid ett valt tillfälle
- **Realtidsbearbetning** (Real-time processing)  
Reagerar på händelser när de händer, på sekunder eller millisekunder.
- **Strömning** (Stream processing)  
Bearbetar kontinuerligt med output lika snabb som input



# Behovspyramiden för AI-utveckling





AI needs Data

Data needs Governance

---

# Data management

---

Informationsförvaltning

**Definition:**

Utveckling, utförande och övervakande av planer, policies, program och arbetssätt för att leverera, styra, och förbättra **värdet av data och information genom hela livscykeln.**

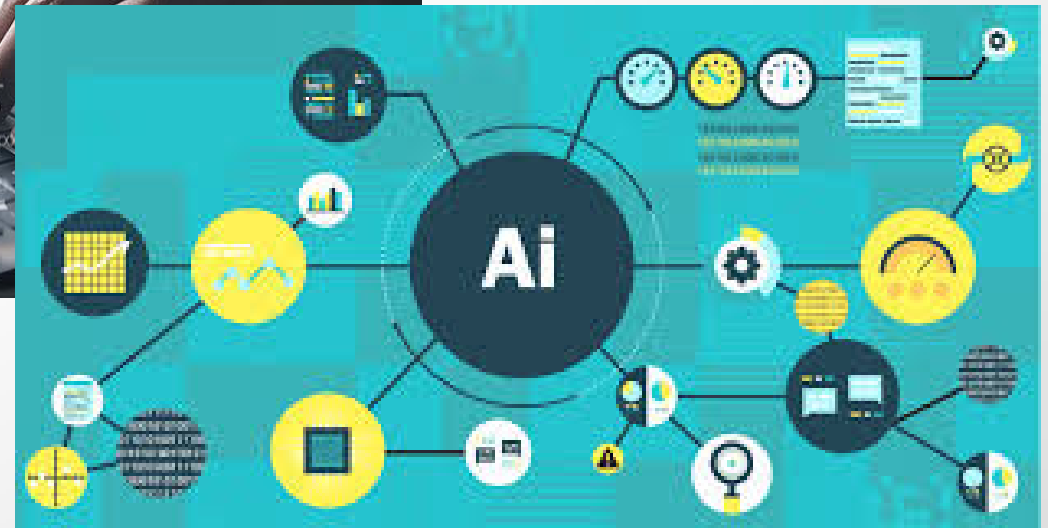
# Data management i en AI-satsning:



Utforskarfas



Utvecklingsfas



Produktionsfas

# Utforskarfas

## Steg 1: **Identifiera och värdera kandiderande datakällor**

Undersök datamängder,  
bedöm relevans för det vi  
vill åstadkomma.

## Steg 2: **Bygg leveranskedjor**

Till exempel: API:er eller  
filöverföringar

## Steg 3: **Utforska möjligheter**

med de data vi nu har  
tillgängliga



# Utforskarfas

## Steg 1: **Identifiera och värdera kandiderande datakällor**

Undersök datamängder,  
bedöm relevans för det vi  
vill åstadkomma.

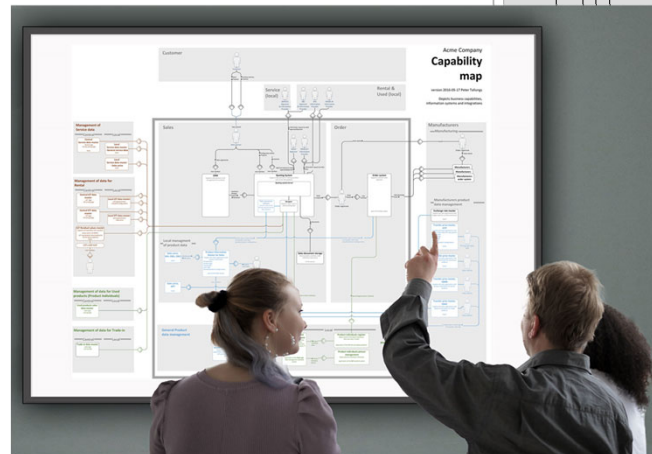
## Steg 2: **Bygg leveranskedjor**

Till exempel: API:er eller  
filöverföringar

## Steg 3: **Utforska möjligheter**

med de data vi nu har  
tillgängliga

- Informationsmodellering
- Datautvärdering
- Kartläggning av dataflöden



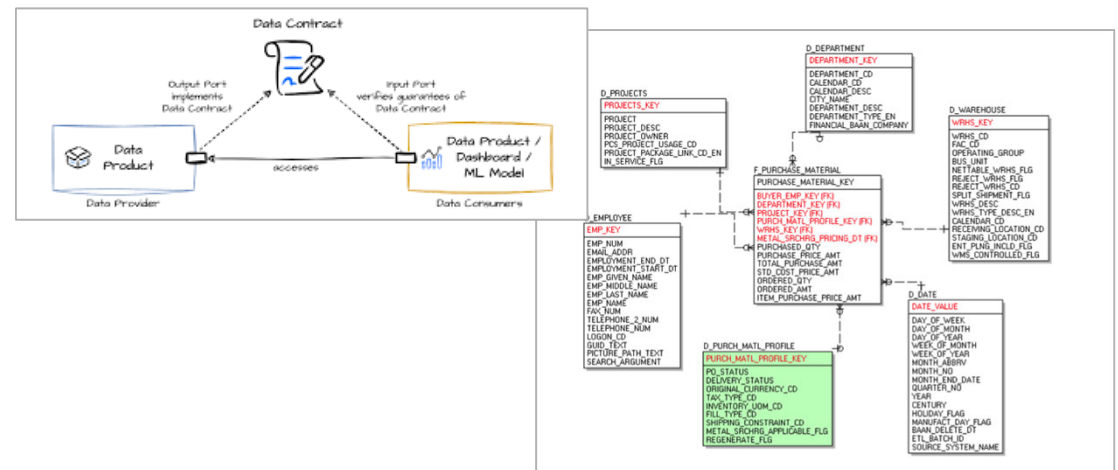
# Utforskarfas

**Steg 1: Identifiera och värdera kandidaterande datakällor**  
Undersök datamängder, bedöm relevans för det vi vill åstadkomma.

**Steg 2: Bygg leveranskedjor**  
Till exempel: API:er eller filöverföringar

**Steg 3: Utforska möjligheter**  
med de data vi nu har tillgängliga

- Utformning av datatjänster med kontrakt
- Design av datastrukturer
- Utformning av datalagring (ERD, Data Vault, dimensionsmodeller, etc)



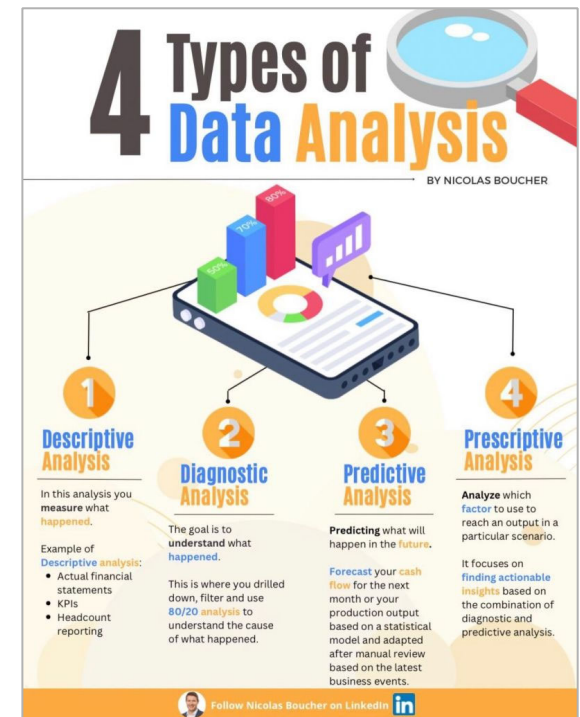
# Utforskarfas

**Steg 1: Identifiera och värdera  
kandiderande datakällor**  
Undersök datamängder,  
bedöm relevans för det vi  
vill åstadkomma.

**Steg 2: Bygg leveranskedjor**  
Till exempel: API:er eller  
filöverföringar

**Steg 3: Utforska möjligheter**  
med de data vi nu har  
tillgängliga

- Olika typer av dataanalys med olika analysverktyg



Data management i en AI-satsning

## Utvecklingsfas

### Steg 4: **Kurera data**

Bygg mekanismer för att tvätta, filtrera, sortera, etikettera, aggregera, strukturera.

### Steg 5: **Träna, utveckla, testa AI-modellen**

### Steg 6: **Bygg AI-system**

### Steg 7: **Produktionssättning**



Data management i en AI-satsning

# Utvecklingsfas

## Steg 4: Kurera data

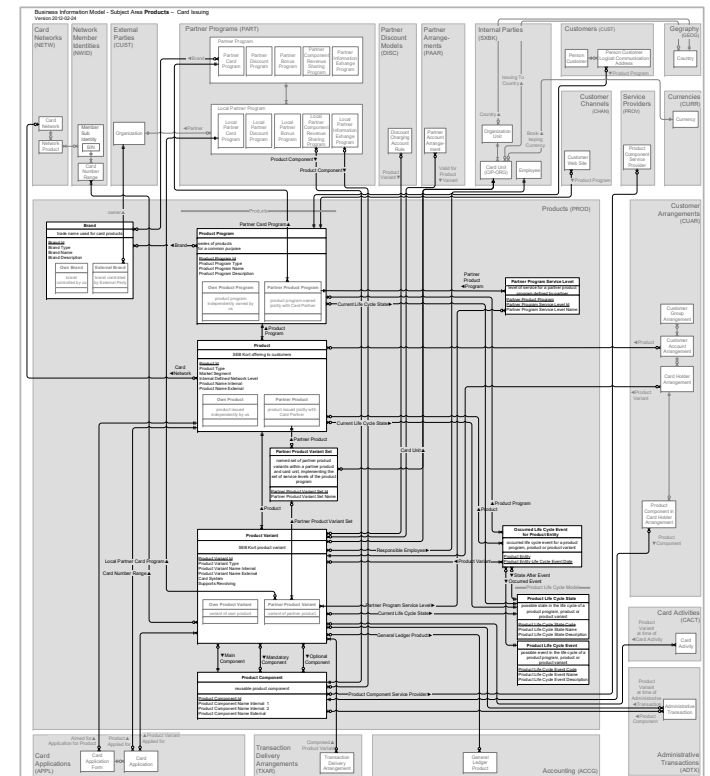
Bygg mekanismer för att tvätta, filtrera, sortera, etikettera, aggregera, strukturera.

## Steg 5: Träna, utveckla, testa AI-modellen

## Steg 6: Bygg AI-system

## Steg 7: Produktionssättning

- Informationsmodellering
- Kartläggning av data



# Utvecklingsfas

Bygg mekanismer för att tvätta, filtrera, sortera, etikettera, aggregera, strukturera.

## Steg 6: Bygg AI-system

## Steg 7: Produktionssättning

- [illegible]

Data management i en AI-satsning

# Utvecklingsfas

## Steg 4: Kurera data

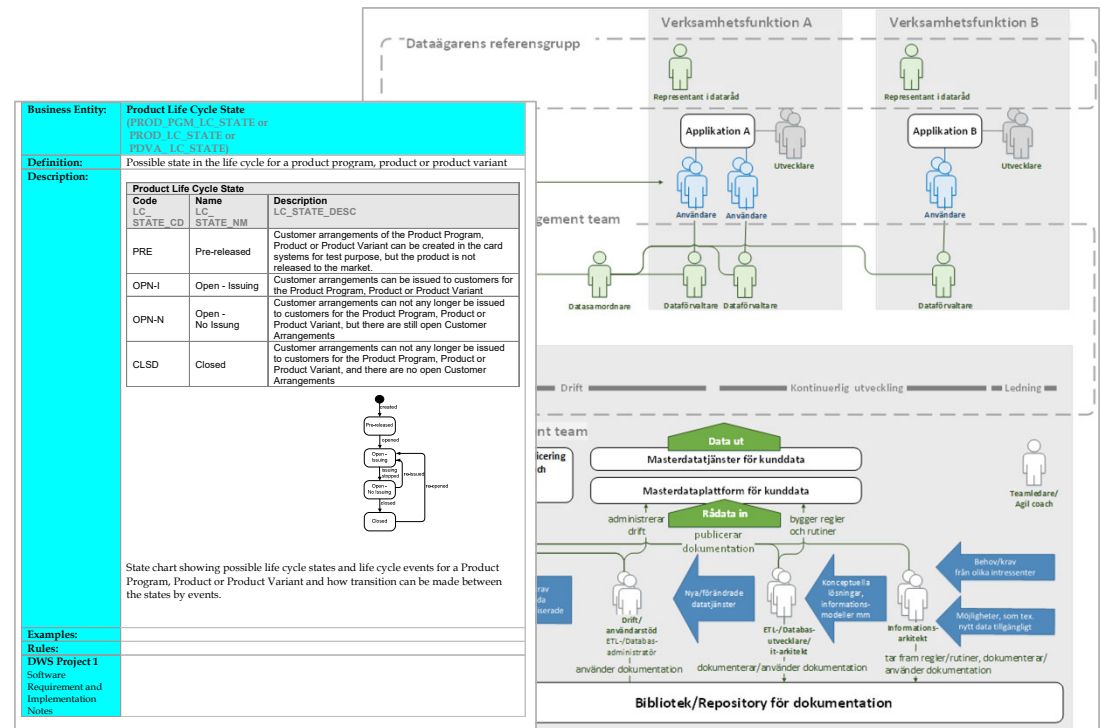
Bygg mekanismer för att tvätta, filtrera, sortera, etikettera, aggregera, strukturera.

## Steg 5: Träna, utveckla, testa AI-modellen

## Steg 6: Bygg AI-system

## Steg 7: Produktionssätt

- Olika former av dokumentation
- Utformning av organisation och ansvar



# Utvecklingsfas

## Steg 4: Kurera data

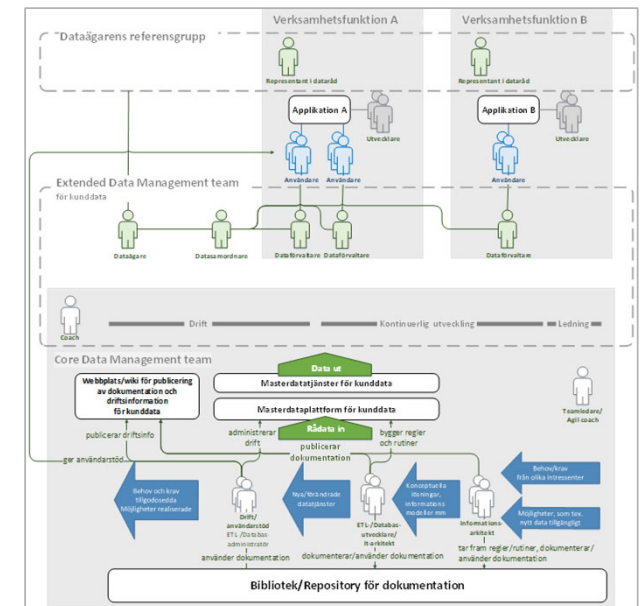
Bygg mekanismer för att tvätta, filtrera, sortera, etikettera, aggregera, strukturera.

## Steg 5: Träna, utveckla, testa AI-modellen

## Steg 6: Bygg

## Steg 7: Produktionssättning

- Dokumentera och utbilda för användning, drift och förvaltning
- Uppdatera beskrivningar för verksamhets- och it-arkitektur



# Produktionsfas

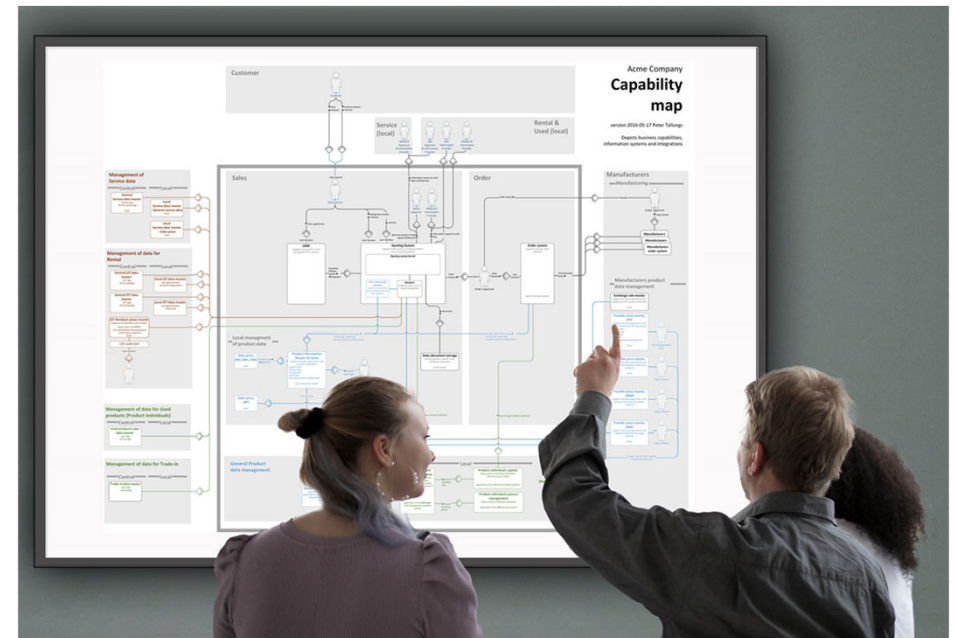
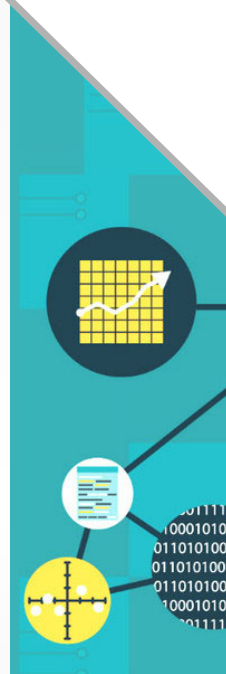
[illegible]

Data management i en AI-satsning:

# Produktionsfas

## Steg 8: Använd och förvalta AI-lösning

- Fånga upp och formulera vidare behov och möjligheter.



# Frågor och funderingar



# Tack!



**Peter Tallungs**

[peter.tallungs@irm.se](mailto:peter.tallungs@irm.se)

linkedin.com/in/  
peter-tallungs-b245011



**Annelie Wiberg**

[annelie.wiberg@irm.se](mailto:annelie.wiberg@irm.se)

linkedin.com/in/  
anneliewiberg/



- Läs mer om oss på [irm.se](http://irm.se)
- Prenumerera på IRM:s nyhetsbrev
- Följ oss på linkedin

