



PROCESS MINING CAMP 2018

A New Skill Set for Process and Data Scientists

prof.dr.ir. Wil van der Aalst
RWTH Aachen University
W: vdaalst.com T: @wvdaalst

Not easy to "sell" process mining



"Nothing is forever" 1704 - 2018





Chair of Process
and Data Science

RWTHAACHEN
UNIVERSITY



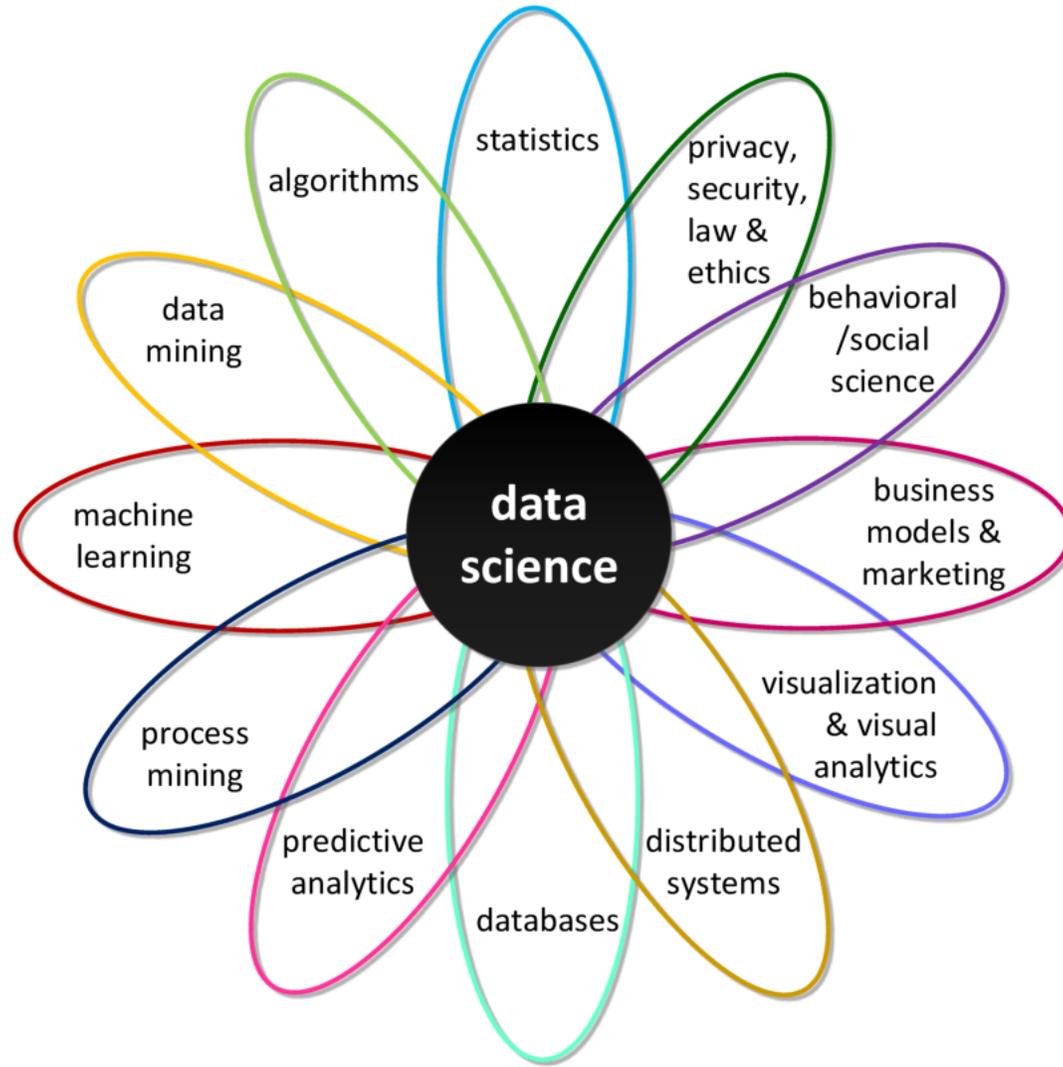
What is data science and why is it attracting so much attention?

Data Science, Big Data, AI, ML, etc. fueled by Moore's law



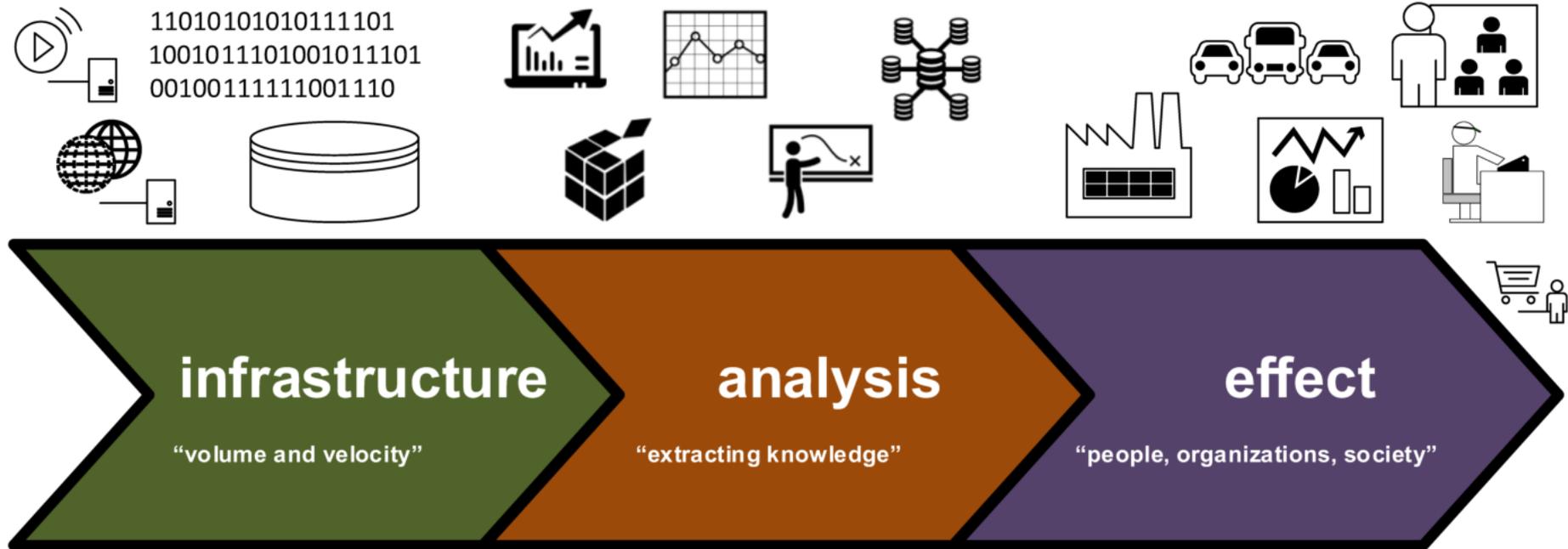
Smartphones, social media, smart homes, cloud computing, autonomous cars, etc.
Suddenly the bath is overflowing ...

Data Science Skills: A diagram that I created in 2012



DSC/e

The data science pipeline



- instrumentation
- big data infrastructures and distributed systems
- databases and data management
- programming
- security
- ...

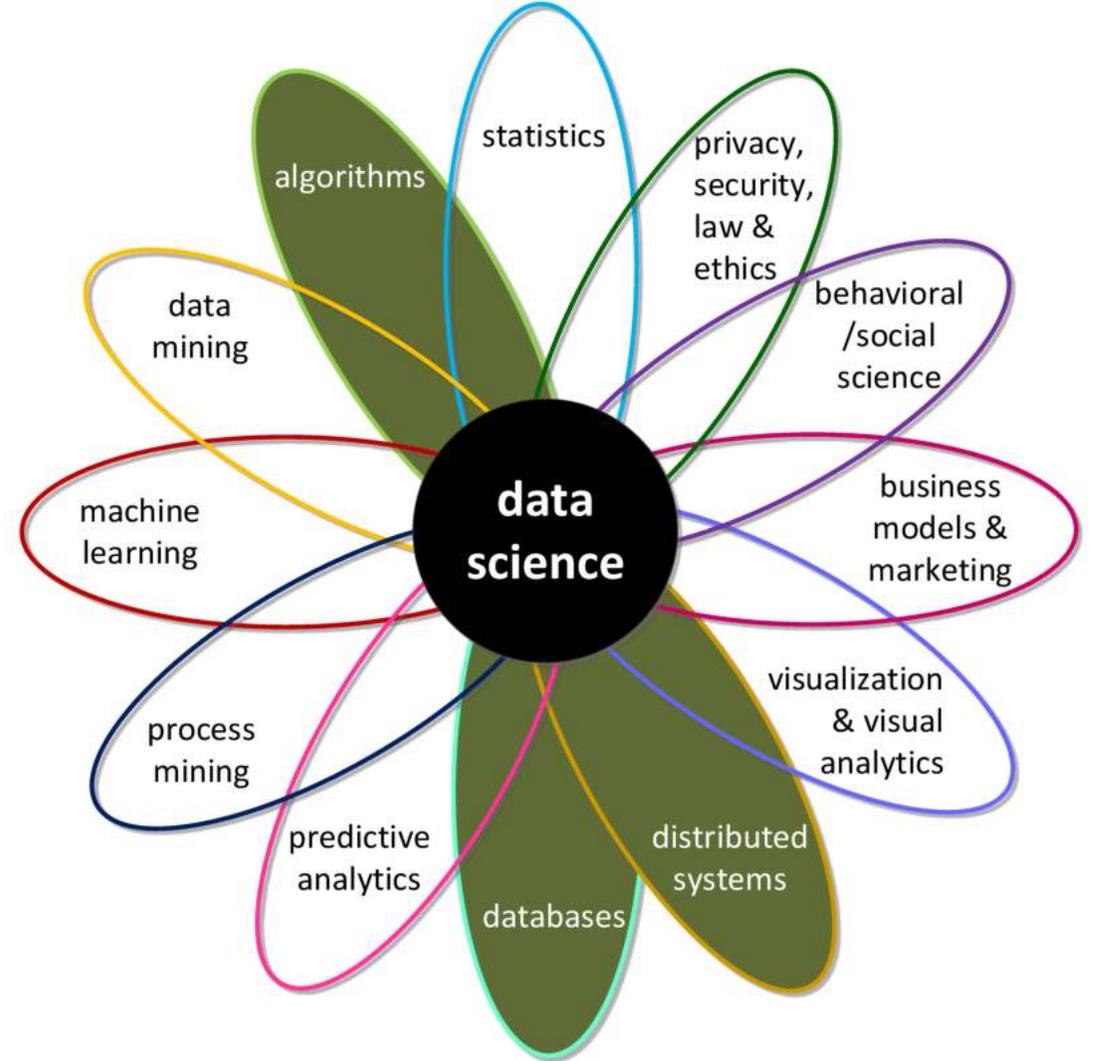
- statistics
- data/process mining
- machine learning/artificial intelligence
- operations research
- algorithms
- visualization
- ...

- ethics & privacy
- IT law
- human technology interaction
- operations management
- business models
- entrepreneurship
- ...

Infrastructure



- | | | |
|--|--|---|
| <ul style="list-style-type: none"> ○ instrumentation ○ big data infrastructures and distributed systems ○ databases and data management ○ programming ○ security ○ ... | <ul style="list-style-type: none"> ○ statistics ○ data/process mining ○ machine learning/artificial intelligence ○ operations research ○ algorithms ○ visualization ○ ... | <ul style="list-style-type: none"> ○ ethics & privacy ○ IT law ○ human technology interaction ○ operations management ○ business models ○ entrepreneurship ○ ... |
|--|--|---|



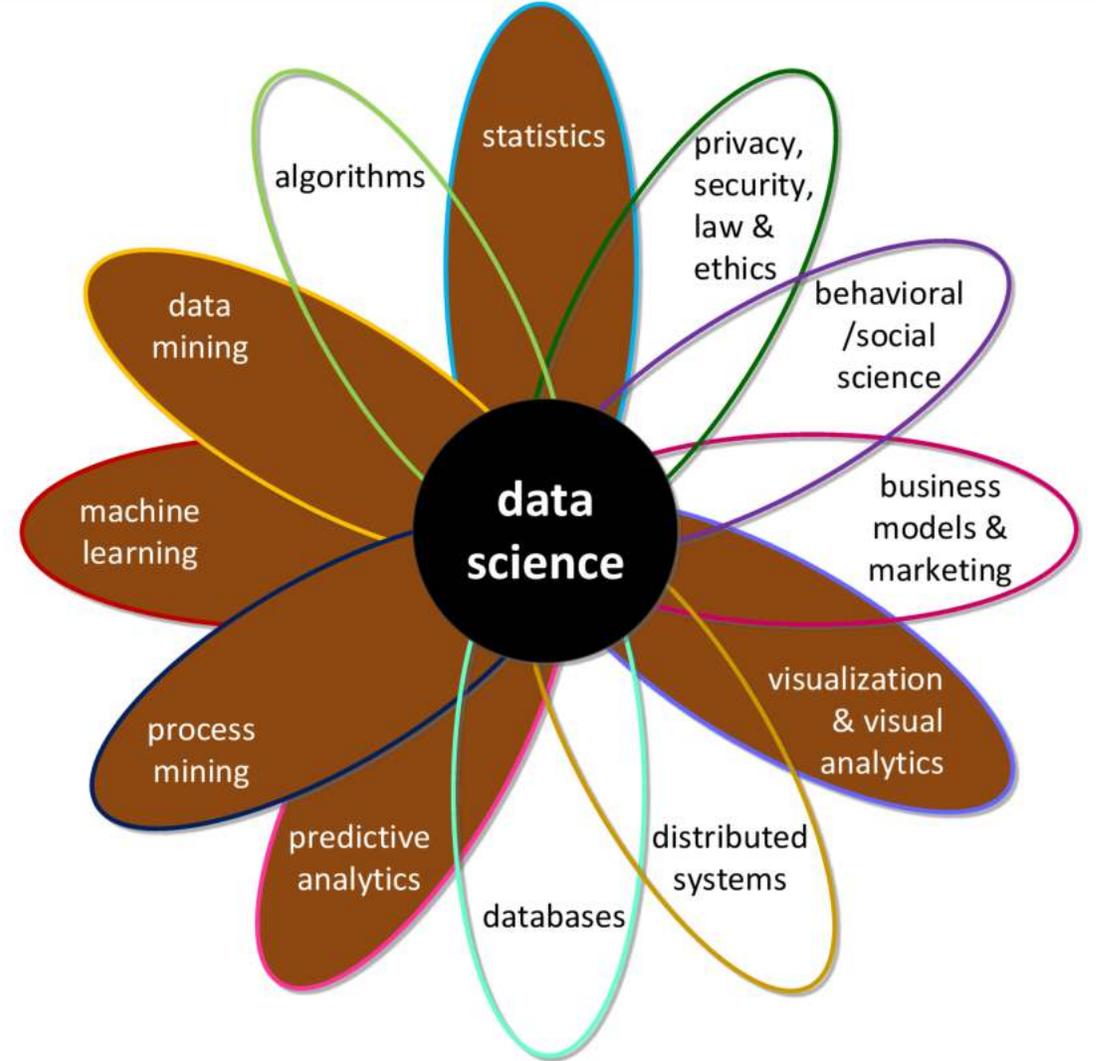
Analysis



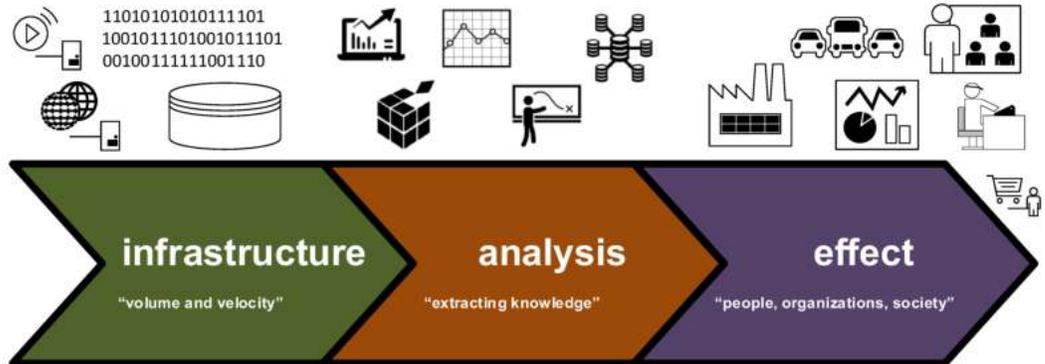
- instrumentation
- big data infrastructures and distributed systems
- databases and data management
- programming
- security
- ...

- statistics
- data/process mining
- machine learning/artificial intelligence
- operations research
- algorithms
- visualization
- ...

- ethics & privacy
- IT law
- human technology interaction
- operations management
- business models
- entrepreneurship
- ...



Effect

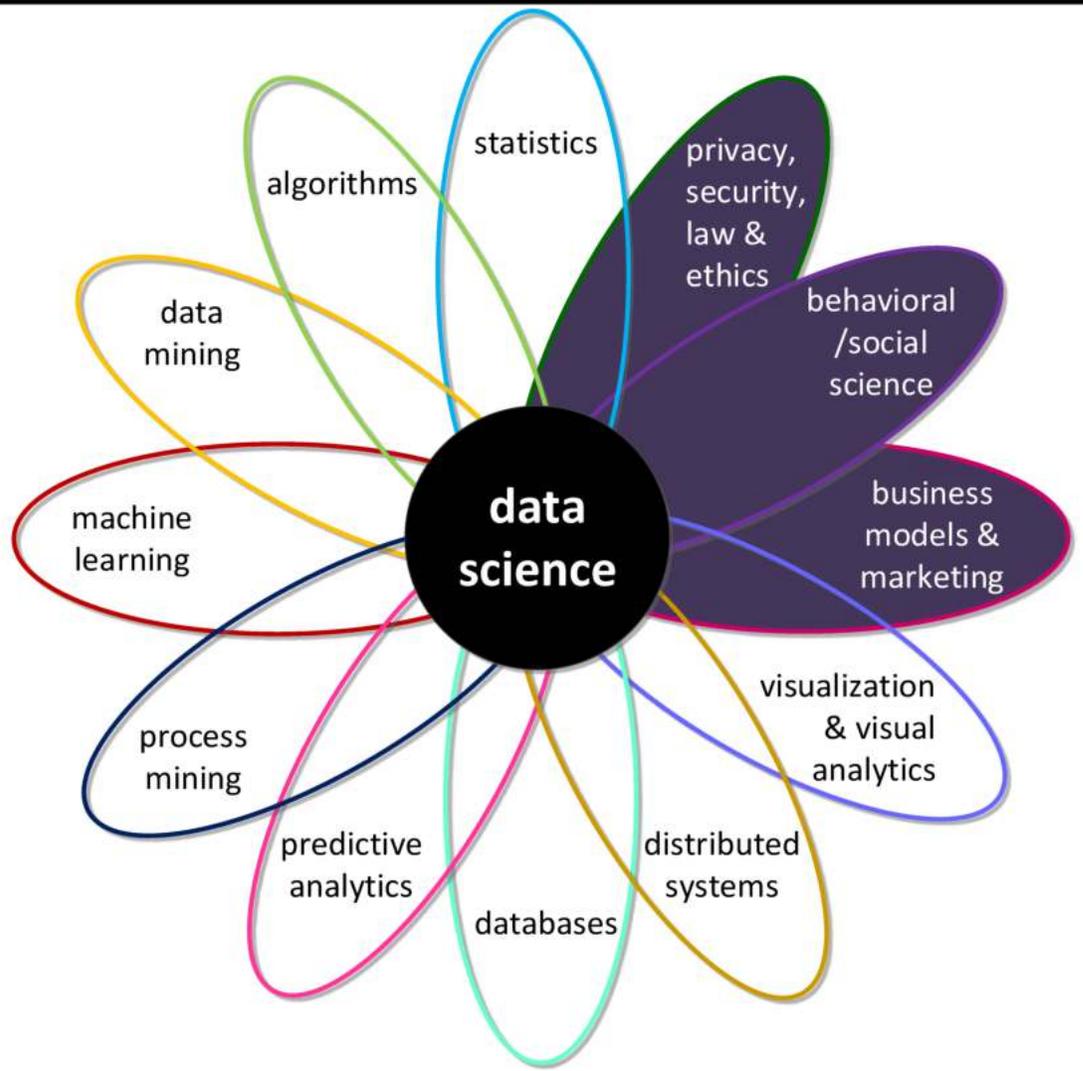


- instrumentation
 - big data infrastructures and distributed systems
 - databases and data management
 - programming
 - security
 - ...
- statistics
 - data/process mining
 - machine learning/artificial intelligence
 - operations research
 - algorithms
 - visualization
 - ...
- ethics & privacy
 - IT law
 - human technology interaction
 - operations management
 - business models
 - entrepreneurship
 - ...

RESPONSIBLE DATA SCIENCE



www.responsibledatascience.org



How do others define the skills of a data scientist?



- CAP Theorem
- Bayes Theorem
- SQL with pig
- Chukwa
- ggplot2
- principal component analysis
- ...

This is not a trivial job and it seems that many soft-skills are missing.

Picture by Swami Chandrasekaran (IBM)

MODERN DATA SCIENTIST

Data Scientist, the sexiest job of the 21st century, requires a mixture of multidisciplinary skills ranging from an intersection of mathematics, statistics, computer science, communication and business. Finding a data scientist is hard. Finding people who understand who a data scientist is, is equally hard. So here is a little cheat sheet on who the modern data scientist really is.

MATH & STATISTICS

- ☆ Machine learning
- ☆ Statistical modeling
- ☆ Experiment design
- ☆ Bayesian inference
- ☆ Supervised learning: decision trees, random forests, logistic regression
- ☆ Unsupervised learning: clustering, dimensionality reduction
- ☆ Optimization: gradient descent and variants



DOMAIN KNOWLEDGE & SOFT SKILLS

- ☆ Passionate about the business
- ☆ Curious about data
- ☆ Influence without authority
- ☆ Hacker mindset
- ☆ Problem solver
- ☆ Strategic, proactive, creative, innovative and collaborative

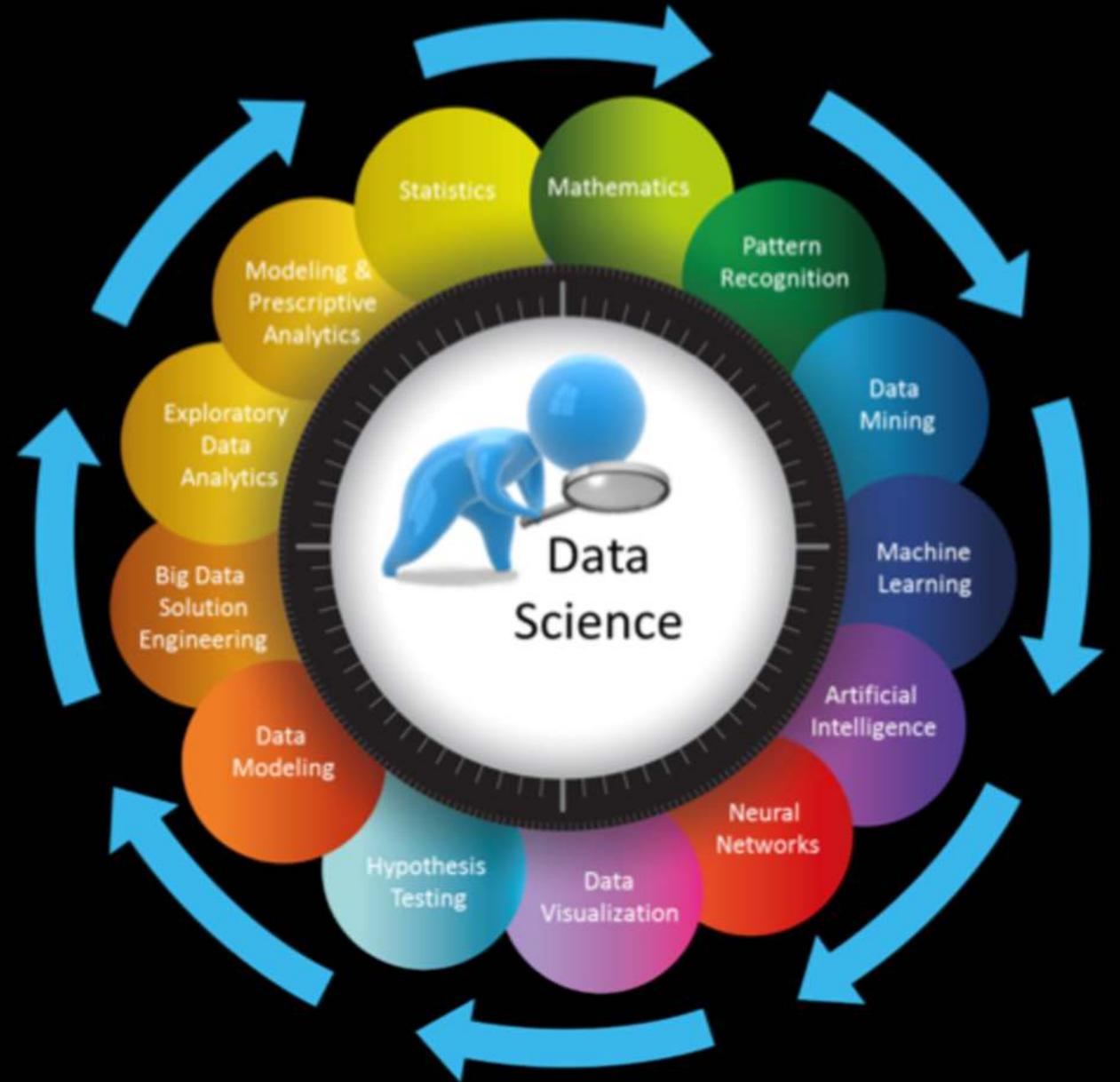
Picture by Marketing Distillery

PROGRAMMING & DATABASE

- ☆ Computer science fundamentals
- ☆ Scripting language e.g. Python
- ☆ Statistical computing packages, e.g., R
- ☆ Databases: SQL and NoSQL
- ☆ Relational algebra
- ☆ Parallel databases and parallel query processing
- ☆ MapReduce concepts
- ☆ Hadoop and Hive/Pig
- ☆ Custom reducers
- ☆ Experience with xaaS like AWS

COMMUNICATION & VISUALIZATION

- ☆ Able to engage with senior management
- ☆ Story telling skills
- ☆ Translate data-driven insights into decisions and actions
- ☆ Visual art design
- ☆ R packages like ggplot or lattice
- ☆ Knowledge of any of visualization tools e.g. Flare, D3.js, Tableau



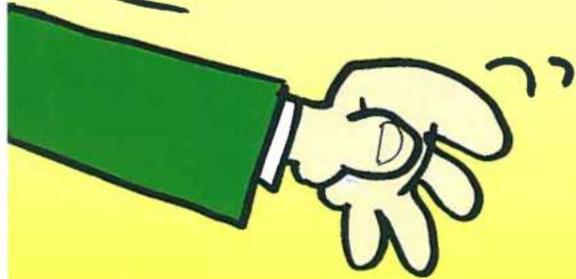
Picture by Emily Curtis

I saw her first!

I want to be a data-scientist ...

Come to us!

DSC/e 2014



Come here please!

She is mine!



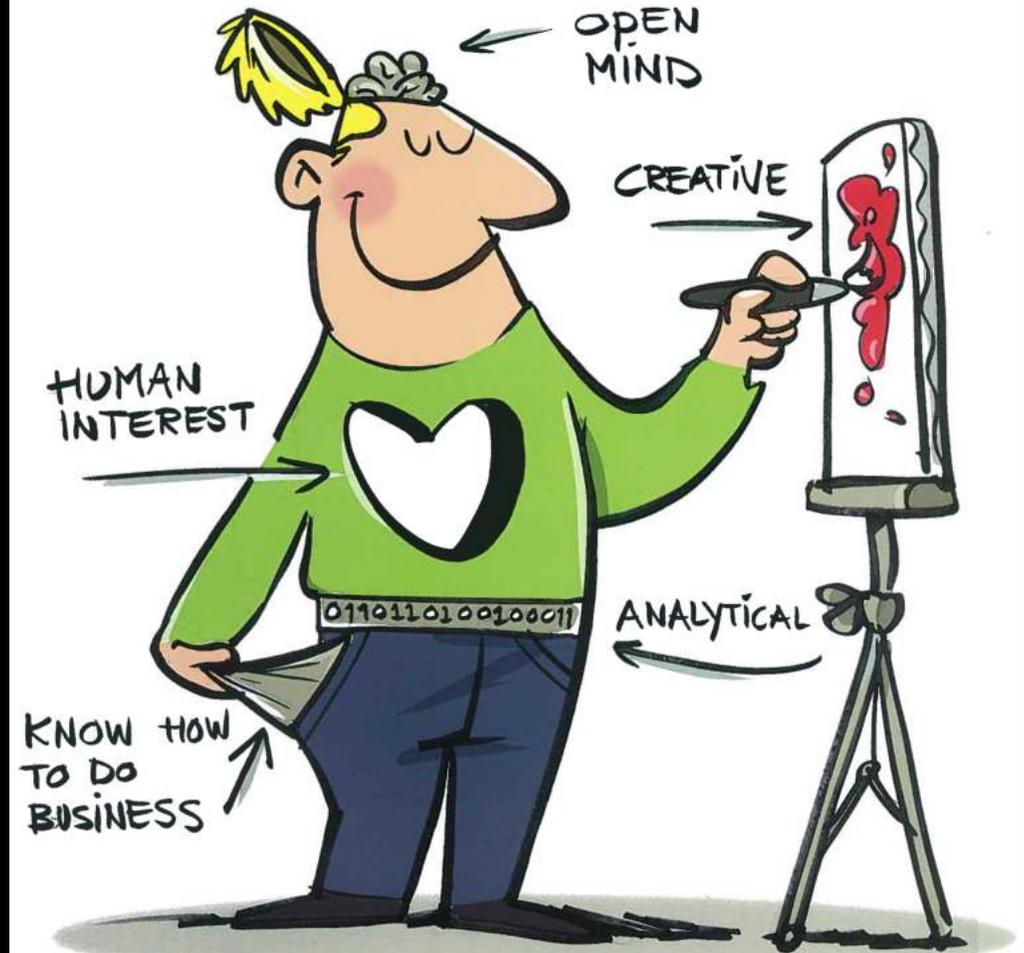
THE PERFECT DATA SCIENTIST



©Marion van de Wiel 2014

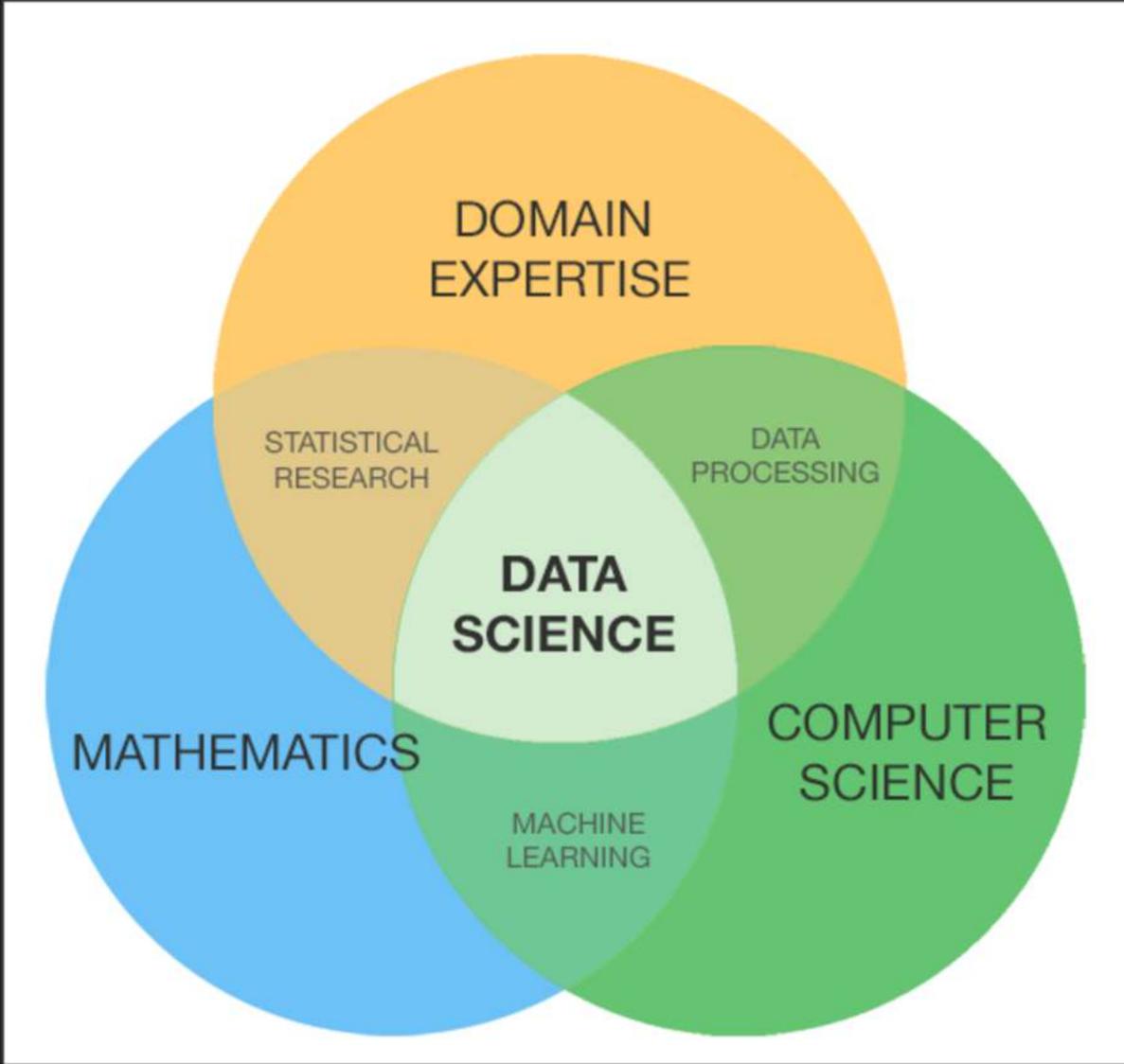
DSC/e 2014

THE PERFECT DATA SCIENTIST

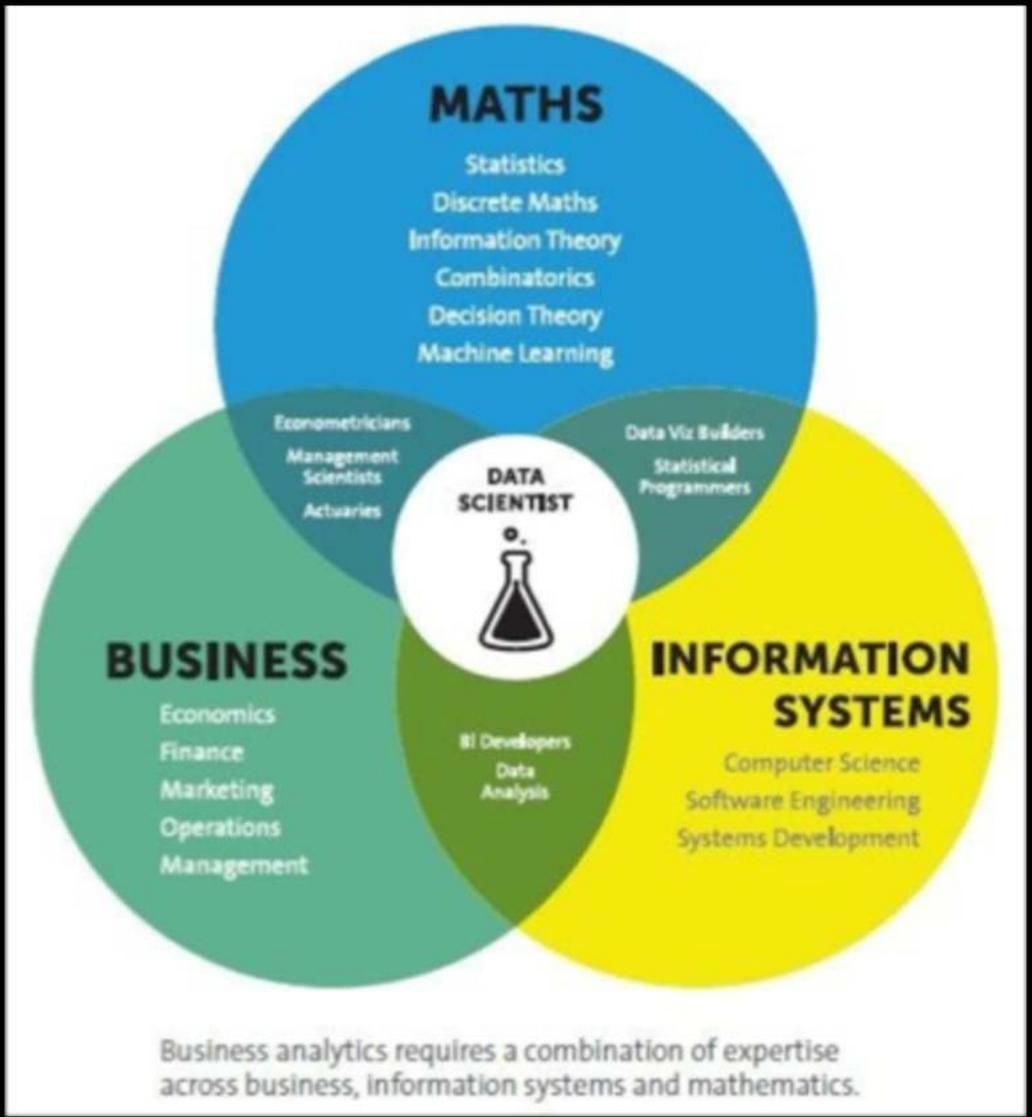


©Marion van de Wiel 2014

DSC/e 2014

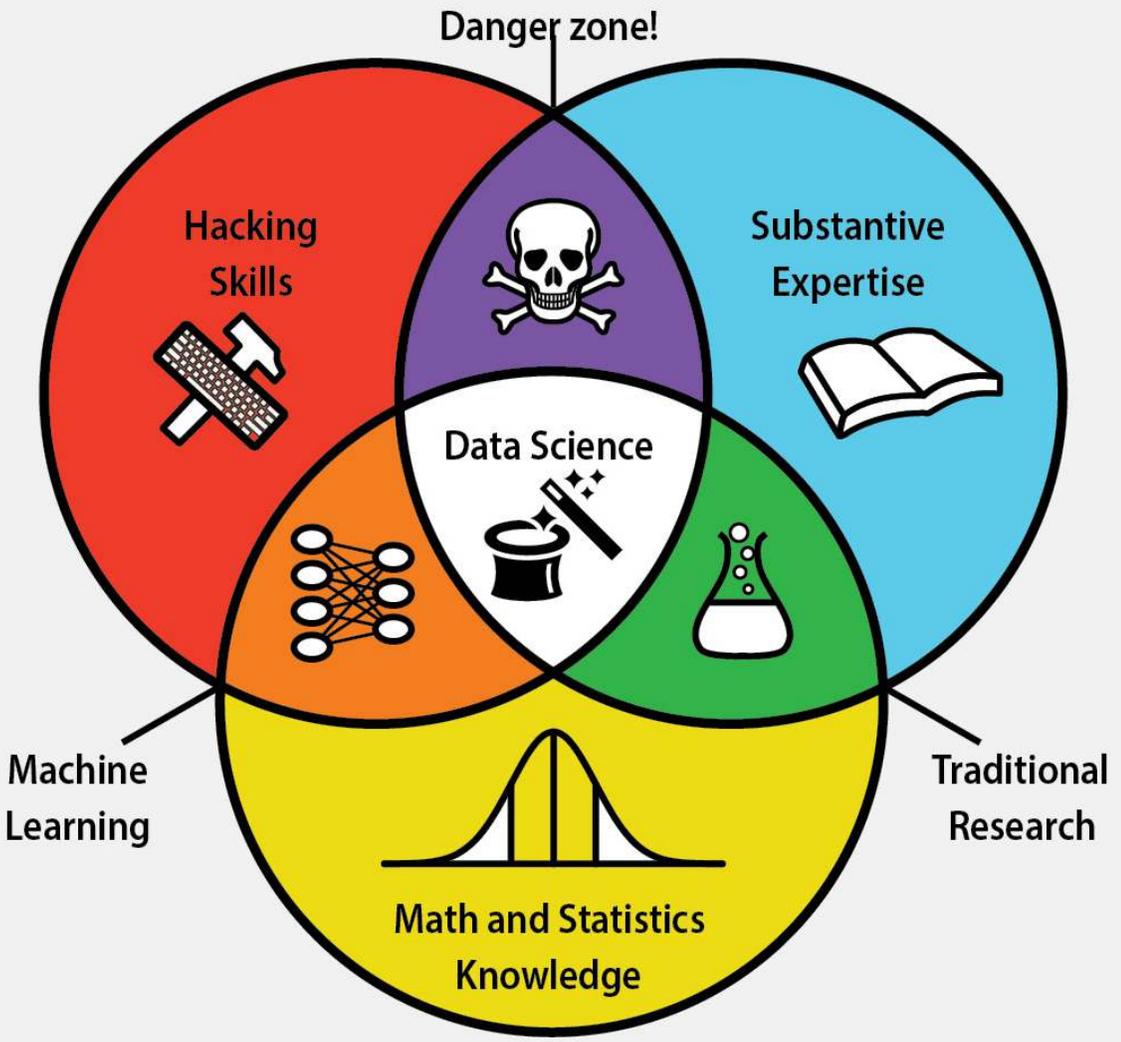


Picture by Shelly Palmer



Picture by Marbella International University

DATA SCIENCE SKILLSET



Data science, due to its interdisciplinary nature, requires an intersection of abilities: **hacking skills, math and statistics knowledge**, and **substantive expertise** in a field of science.



Hacking skills are necessary for working with massive amounts of electronic data that must be acquired, cleaned, and manipulated.



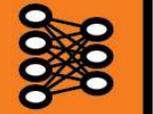
Math and statistics knowledge allows a data scientist to choose appropriate methods and tools in order to extract insight from data.



Substantive expertise in a scientific field is crucial for generating motivating questions and hypotheses and interpreting results.



Traditional research lies at the intersection of knowledge of math and statistics with substantive expertise in a scientific field.



Machine learning stems from combining hacking skills with math and statistics knowledge, but does not require scientific motivation.



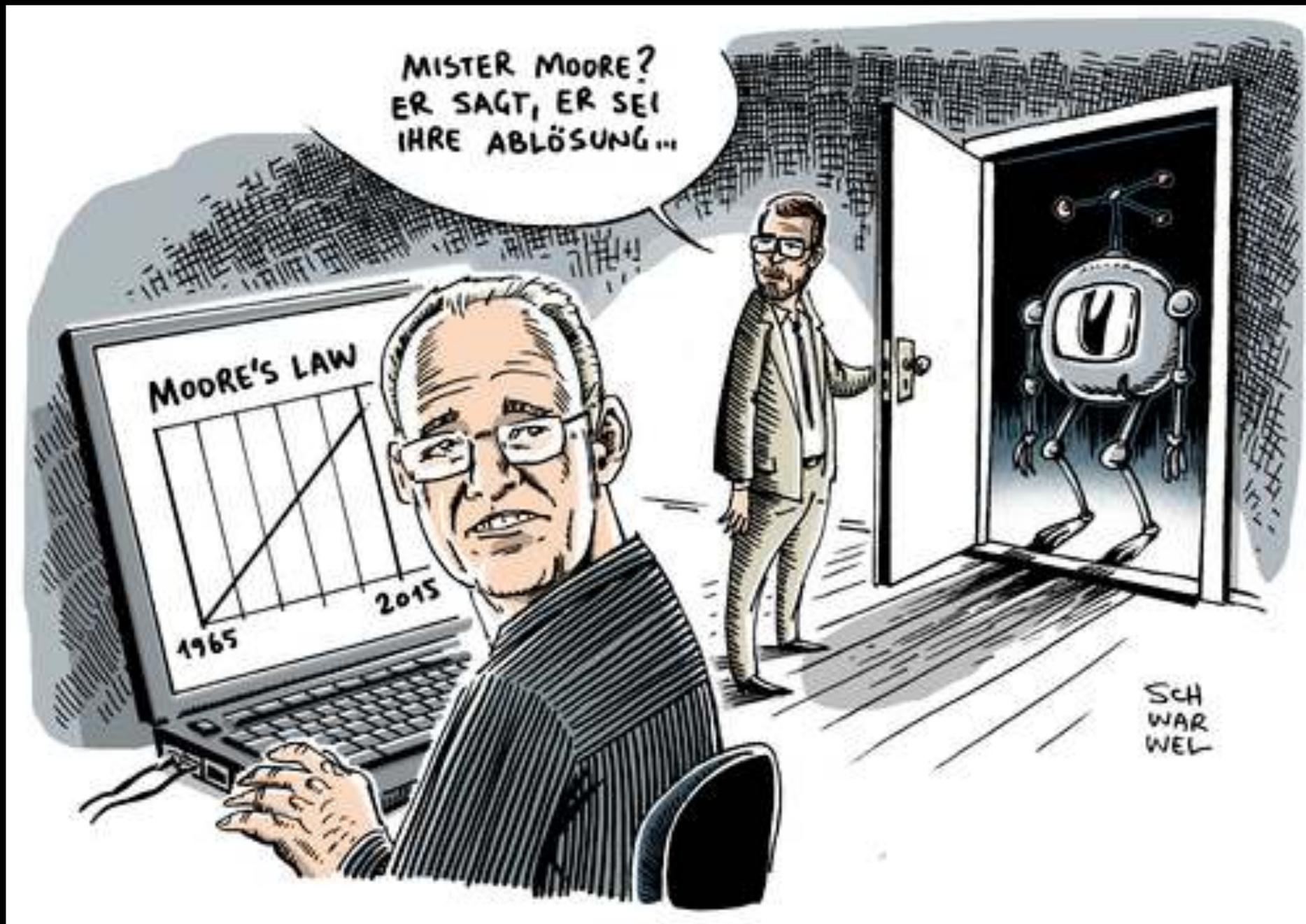
Danger zone! Hacking skills combined with substantive scientific expertise without rigorous methods can beget incorrect analyses.

Picture by Natalia Bilenko, Drew Conway, et al.

Poor Data Scientist



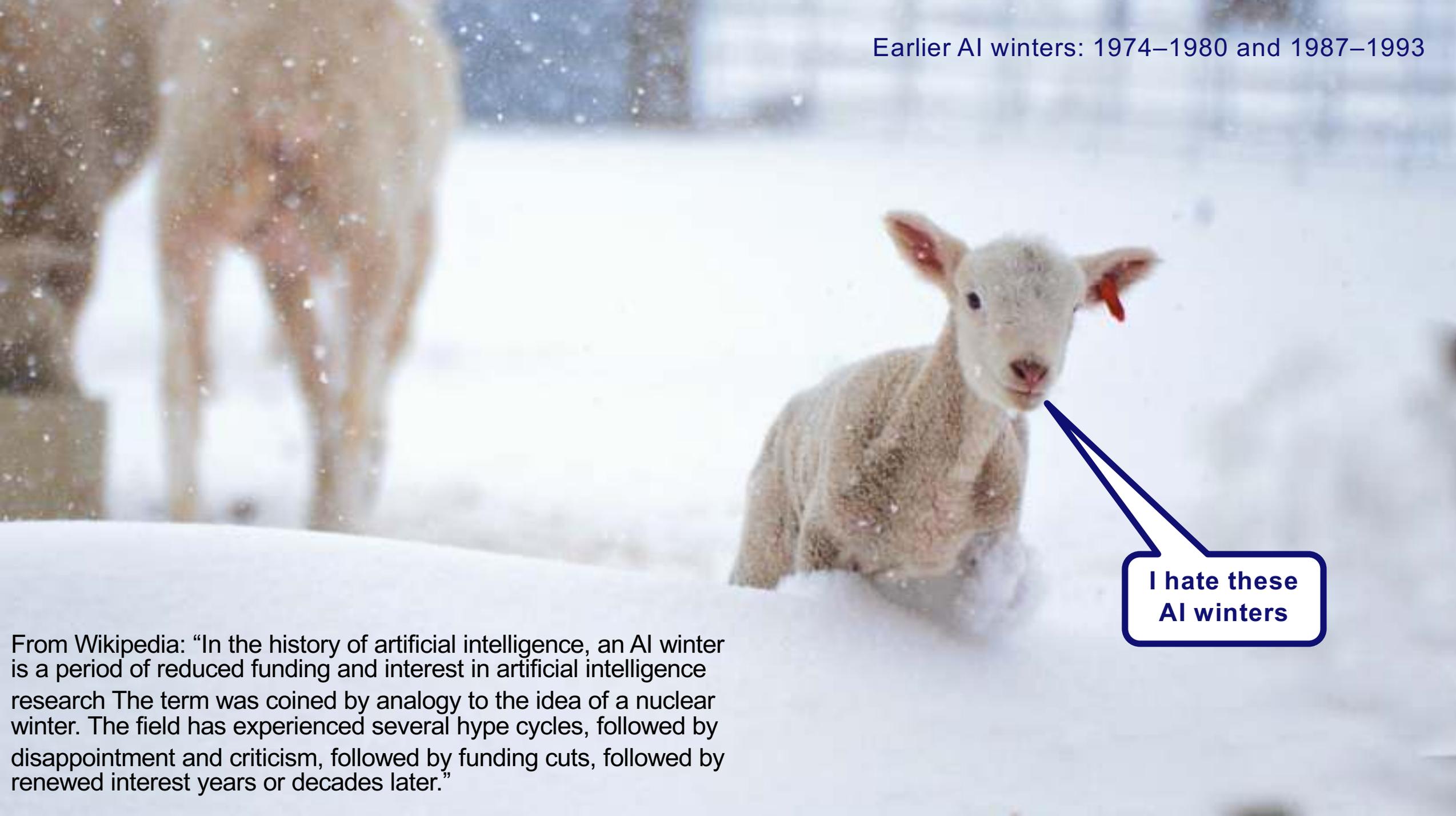
**(s)he was supposed to
have the “most sexy job”
of the century**



**Winter is
coming:
Inflated AI
expectations**

Picture by Thomas
Meitsch/ Schwarwel

Earlier AI winters: 1974–1980 and 1987–1993



**I hate these
AI winters**

From Wikipedia: “In the history of artificial intelligence, an AI winter is a period of reduced funding and interest in artificial intelligence research. The term was coined by analogy to the idea of a nuclear winter. The field has experienced several hype cycles, followed by disappointment and criticism, followed by funding cuts, followed by renewed interest years or decades later.”

Artificial Intelligence in the news



How Data **Intelligence** and Analytics are changing the Indian business ...

YourStory.com - vor 3 Stunden

However, **surprisingly** the demand for data scientists is way higher than their ...

Business Analytics comprises of the technologies, **applications**, skills, and ... quantitative analysis, predictive analytics, and **artificial intelligence**, to name a few.



How the Gulf region can embrace an AI-enabled future

ITP.net - vor 7 Stunden

It's no longer news that **Artificial Intelligence** (AI) will be a driving force ... for **Artificial Intelligence**, which covers development and **application** of AI in ... to centre their business models on AI, it's no **surprise** that the region wants ...



Artificial intelligence not even close to humanlike thought

gulfnews.com - 26.05.2018

The field of **artificial intelligence** doesn't lack for ambition. ... Yes, some **remarkable applications** have been built from it, including Google ...



Artificial intelligence: it may be our future

The Guardian - 15.05.2018

Artificial intelligence (AI) is a spectre that has been specifically **used**, but the **strange** beginning of



How AI Can Help Addicts Stay Sober

Daily Beast - 09.05.2018

Artificial intelligence is the future. ... That's changing, though, and in a new space—mental health. Before ... Basically, it listens in on conversations—but the **application's** not so interested in what's said as to how it's said.



PHOTO ILLUSTRATION BY THE DAILY BEAST



MACHINE MIND



How A.I. Can Help Addicts Stay Sober



Relapses are the biggest obstacle for addicts trying to get sober. A.I. can help with that.

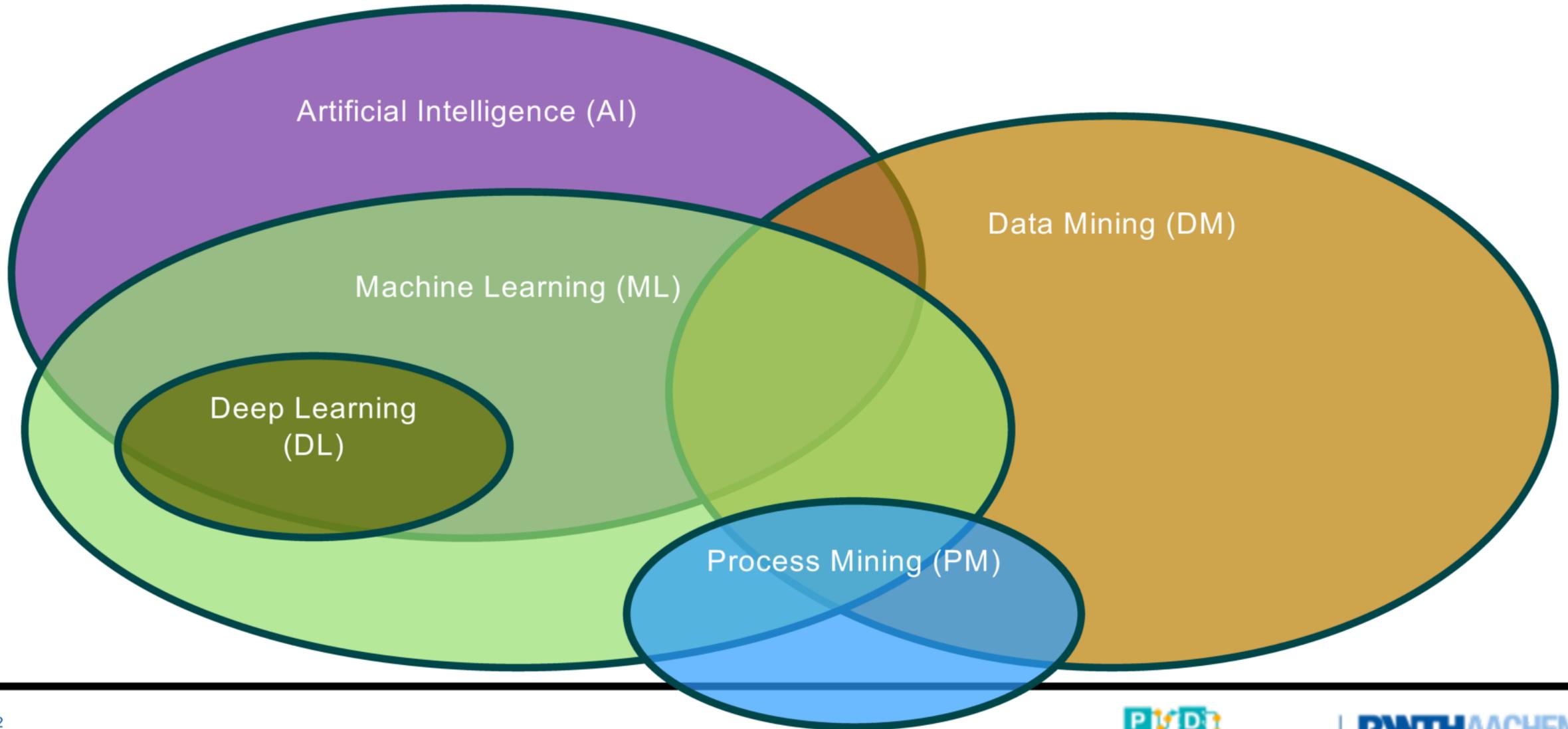


STARRE VARTAN 05.09.18 11:26 AM ET

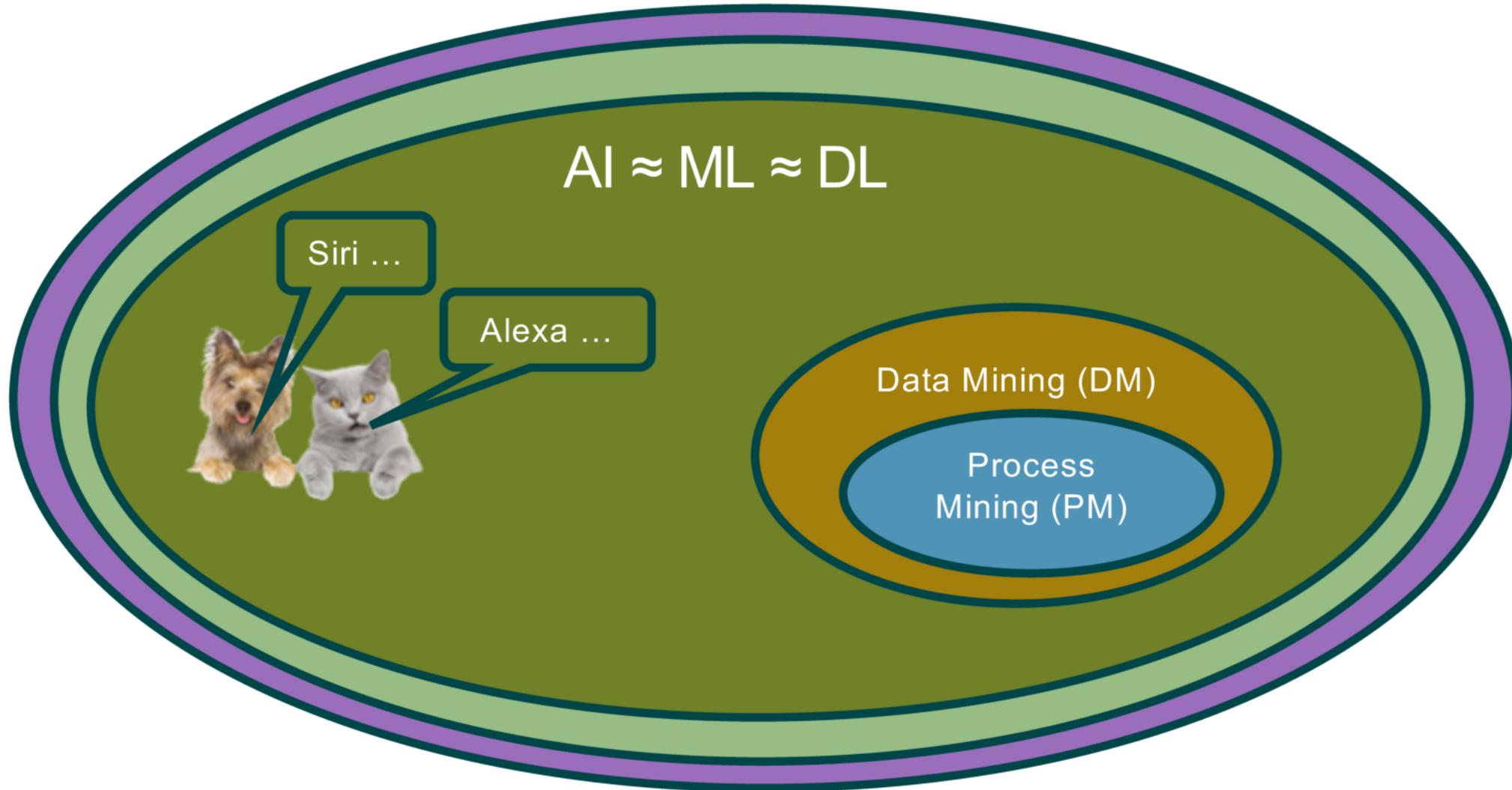
Artificial intelligence is the future. It can predict when we'll die, help us buy stuff we don't need, diagnose cancer, and imperfectly predict songs we'd like. It's not perfect, and sometimes it can be banal.

“Artificial intelligence is the future. It can predict when we'll die, help us buy stuff we don't need, diagnose cancer, and imperfectly predict songs we'd like.” ... “When it comes to mental health care, AI can do something similar—see what's going to happen before it does, so we can treat it.”

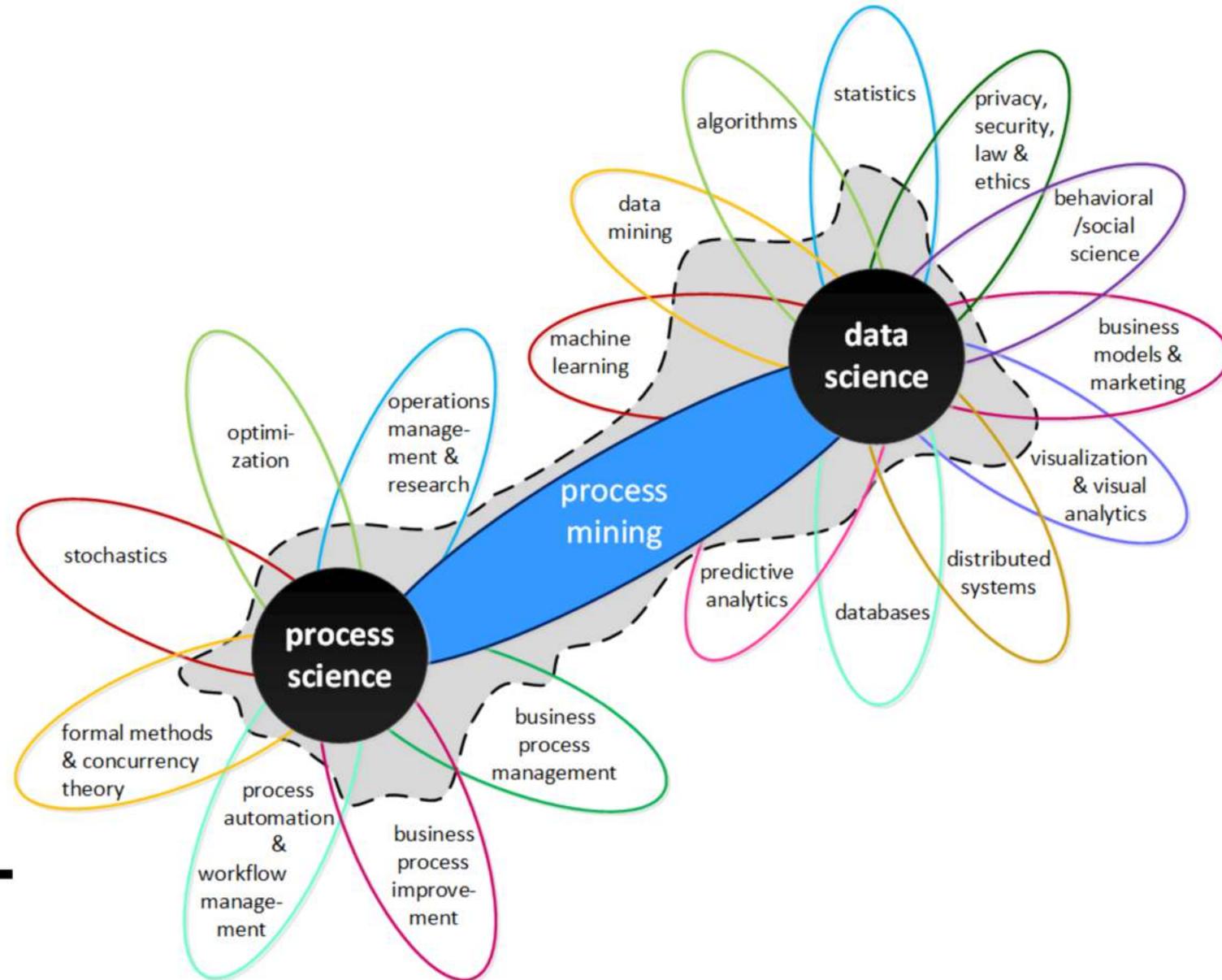
My view of the data science landscape (take with grain of salt)

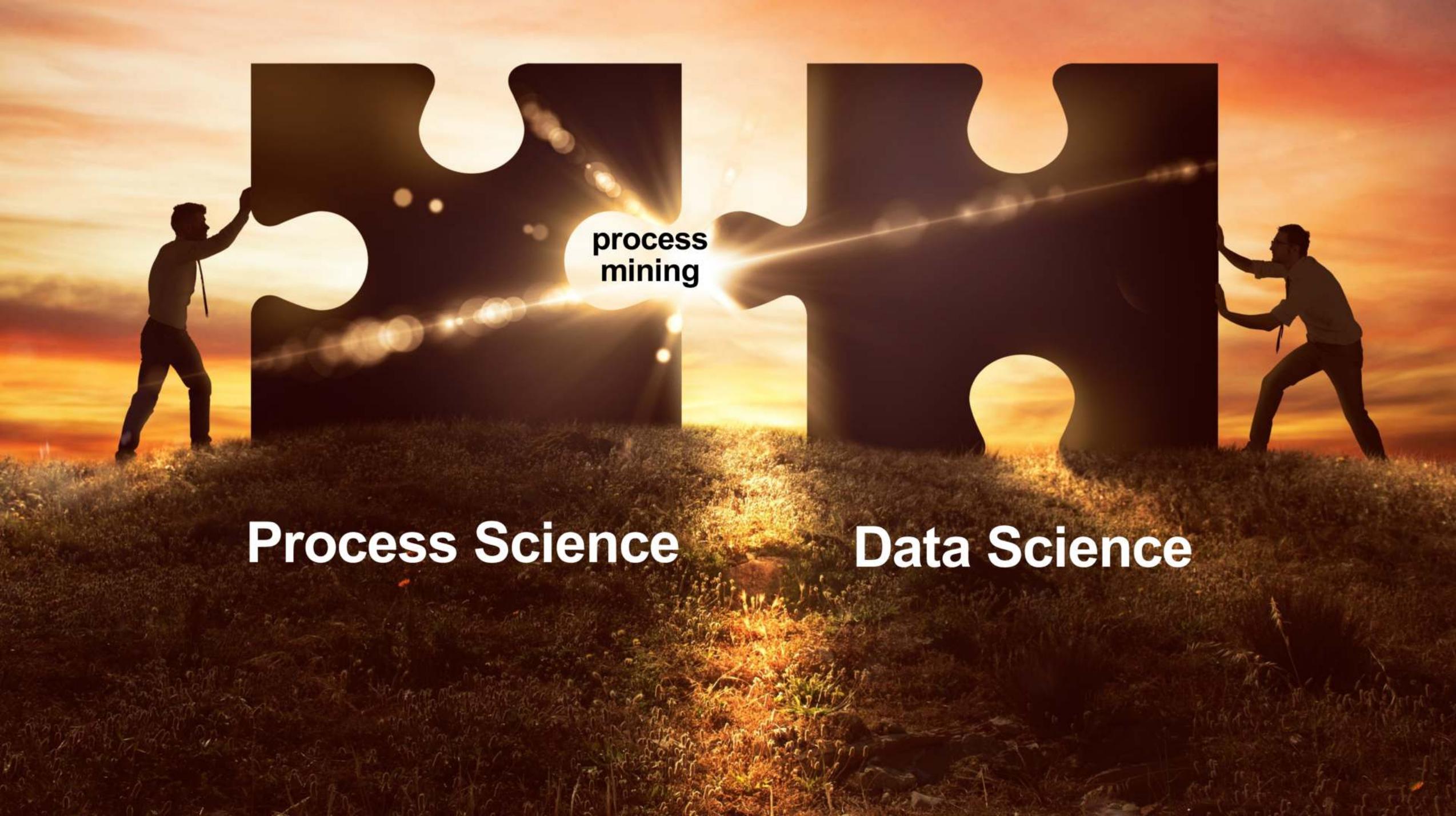


Perception of some and suggestions in media



Positioning process mining





process
mining

Process Science

Data Science

Process mining is awesome!

(but I do not need to explain this here)

Back to the skills needed

To drive a car, do you need to know how the powertrain works?



powertrain = process mining engine

Powertrain Porsche 718

To drive a car, do you need to know how the powertrain works?

It depends on how fast you want to go and whether you want to select your own car.

Understanding helps to operate the car!



To drive a car, do you need to know how the powertrain works?

In any case, one should not deny the existence of cars.

Most professions will change dramatically in the coming years!



pwc.co.nz/automation

Will robots really steal our jobs?

An international analysis of the potential long term impact of automation

Key findings: impact of automation

Financial services jobs could be relatively vulnerable to automation in the shorter term, while transport jobs are more vulnerable to automation in the longer term

In the long run, less well educated workers could be particularly exposed to automation, emphasising the importance of increased investment in lifelong learning and retraining

Figure 1 – Potential job automation rates by industry across waves

% of existing jobs at potential risk of automation

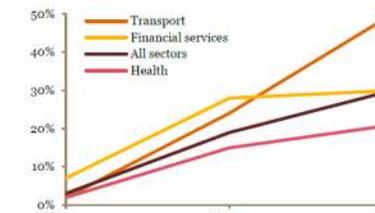


Figure 2 – Potential job automation rates by education level across waves

% of existing jobs at potential risk of automation

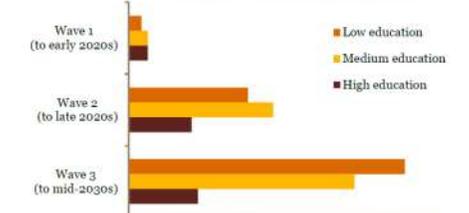
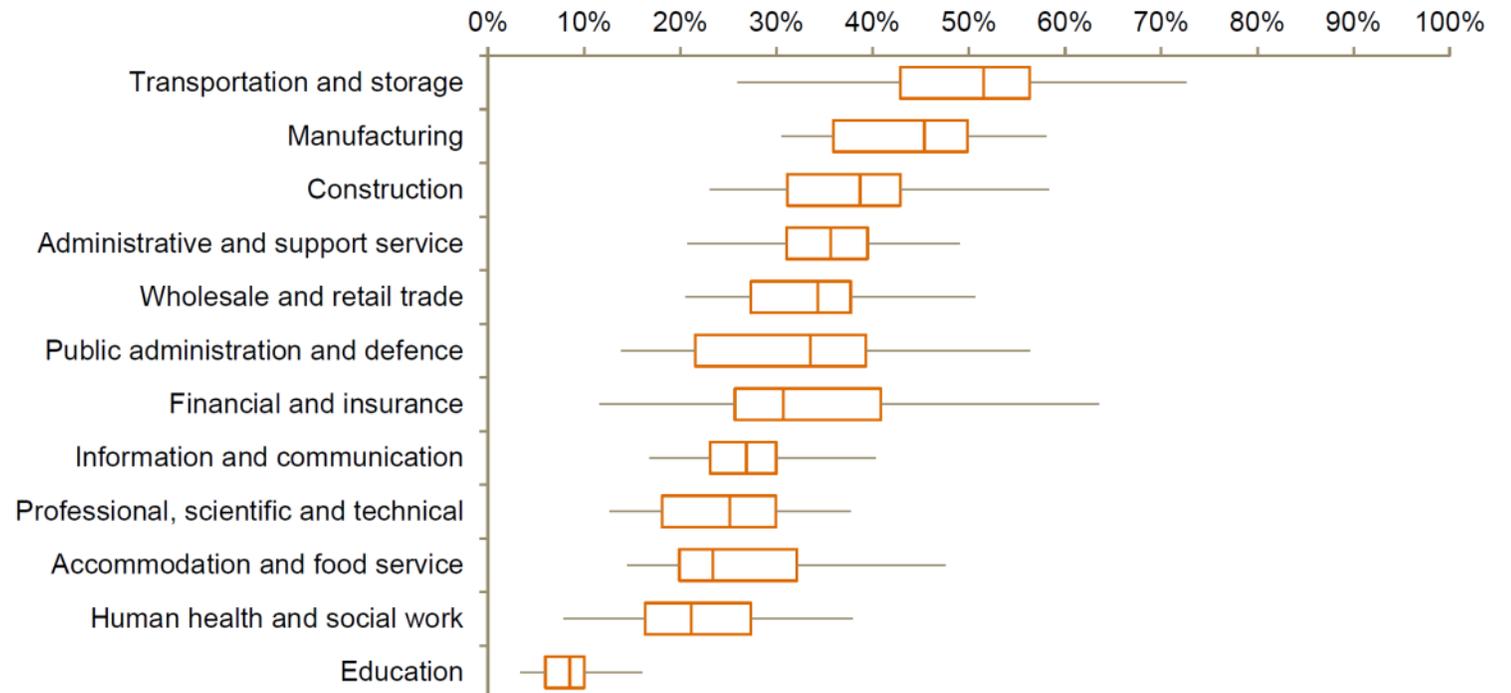


Figure 4.1 – Share of jobs with potential high automation rates by industry

Potential jobs at high risk of automation

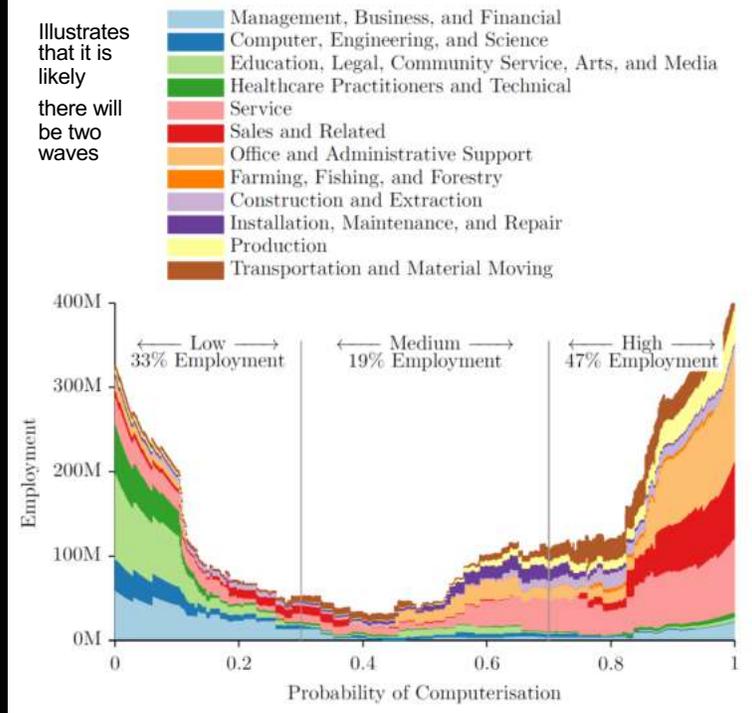


The Future of Employment

Carl Benedikt Frey & Michael Osborne



“According to our estimates around **47 percent of total us employment is in the high risk category**. We refer to these as jobs at risk – i.e. jobs we expect could be automated relatively soon, perhaps over the next decade or two.”



Occupations and probability of computerization (sample from 702 occupations):

- Healthcare Social Workers 0.35%
- Firefighters 17%
- Statisticians 22%
- Accountants and Auditors 94%
- Bookkeeping and Accounting 98%
- Tax Preparers 99%

How about Lean Six Sigma professionals, IT Auditors, Business Consultants, etc. ?

- These jobs will remain for some time, but will definitely change.
- Changes will not only relate to efficiency/cost reduction, but also quality improvement.
- A more analytical approach in many professions is inevitable.
- **Process mining is yet visible/widely adopted, because:**
 - Most of the time is still wasted on data preprocessing.
 - Lack of training and still a lack of awareness.
 - Hiding behind privacy arguments, avoiding transparency, etc.
 - It takes time to change.
- **How many of these factors still hold in 10 years from now?**



process mining stereotypes



The process mining pipeline



- extract from data source
- clean data
- filter data
- ...

- process discovery
- conformance checking
- bottleneck analysis
- ...

- explain to stakeholders
- challenge outcomes
- make decisions
- ...

The process mining pipeline

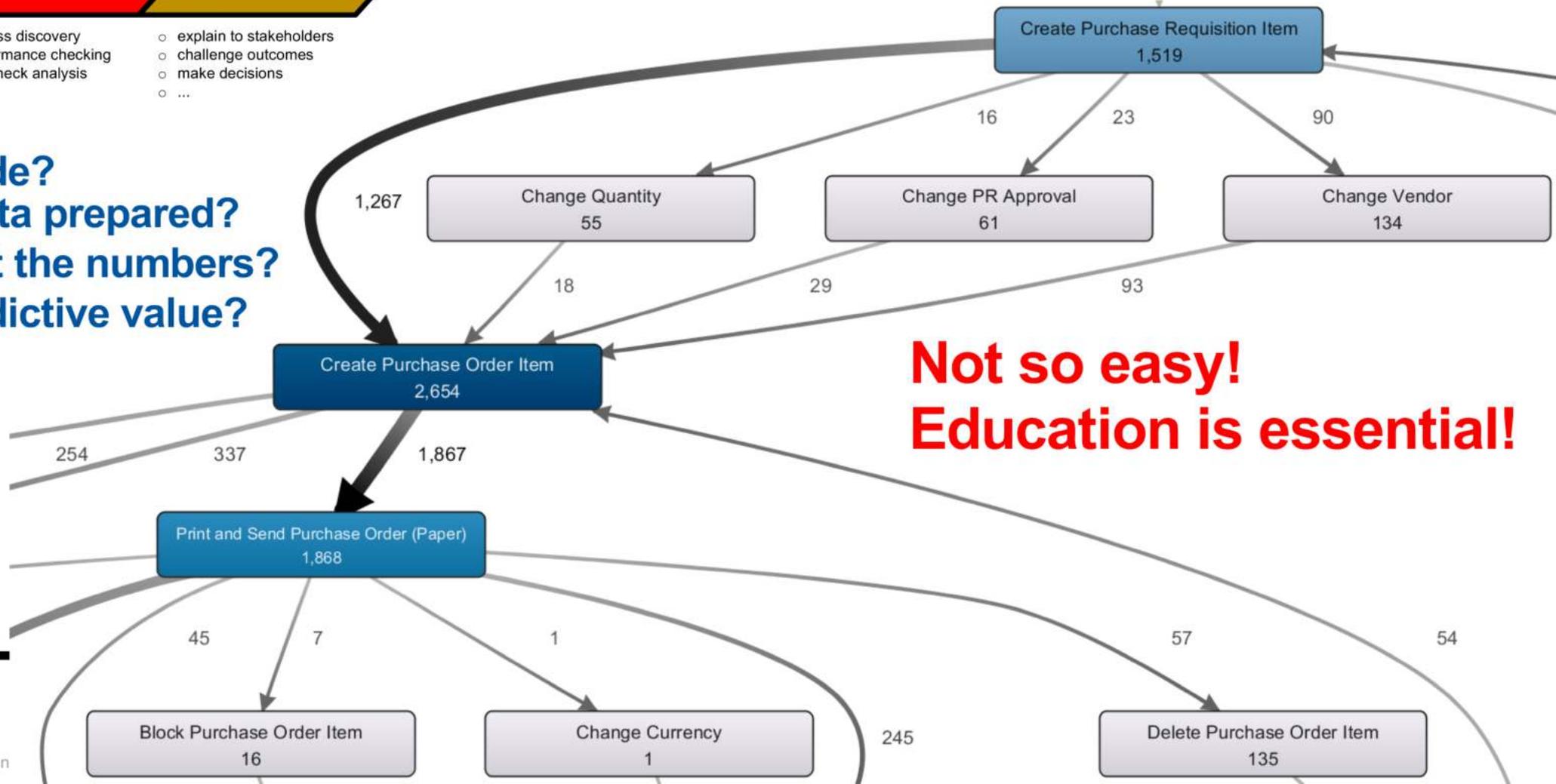


- extract from data source
- clean data
- filter data
- ...

- process discovery
- conformance checking
- bottleneck analysis
- ...

- explain to stakeholders
- challenge outcomes
- make decisions
- ...

What to conclude?
How was the data prepared?
How to interpret the numbers?
What is the predictive value?



**Not so easy!
Education is essential!**



data
science
focused

domain /
business
focused





data
science
focused

1: PM expert

domain /
business
focused

2: Domain expert



data
science
focused

domain /
business
focused

**3: Domain expert
able to
apply/interpret PM**



data
science
focused

domain /
business
focused

**4: PM expert able to do
specific types of analysis**



data
science
focused

domain /
business
focused

Where are you now?
Where do you want to go?



Chair of Process
and Data Science

RWTHAACHEN
UNIVERSITY

Thank you!

Foundations
of Process
Mining

Dealing
with XXXX
Event Data

Automated
Operational
Process
Improvement

Responsible
Process
Mining

 **Fraunhofer**
FIT

www.pads.rwth-aachen.de - www.vdaalst.com - [@wvdaalst](https://twitter.com/wvdaalst)