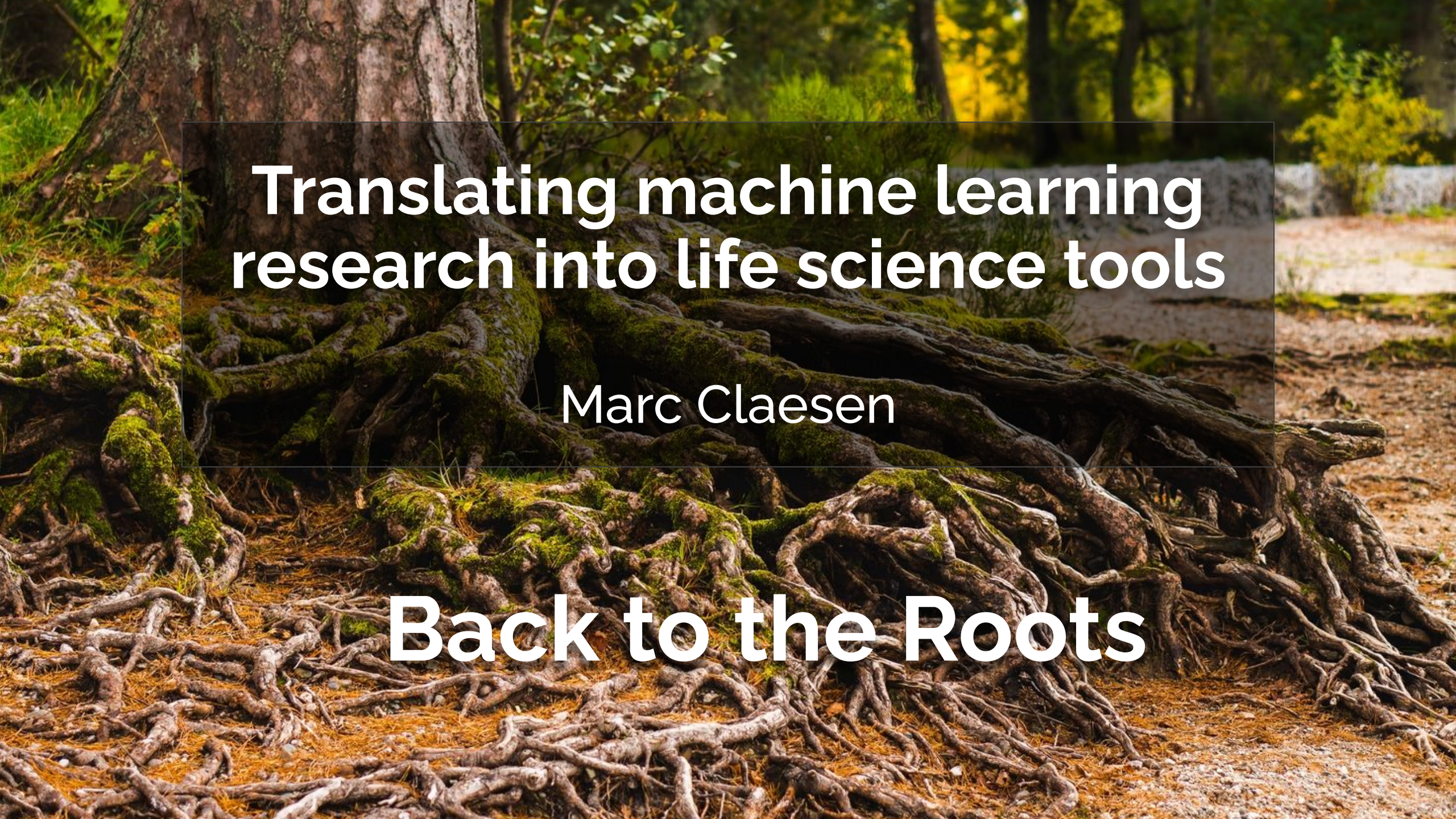

Translating machine learning research into life science tools
Marc Claesen
Back to the Roots

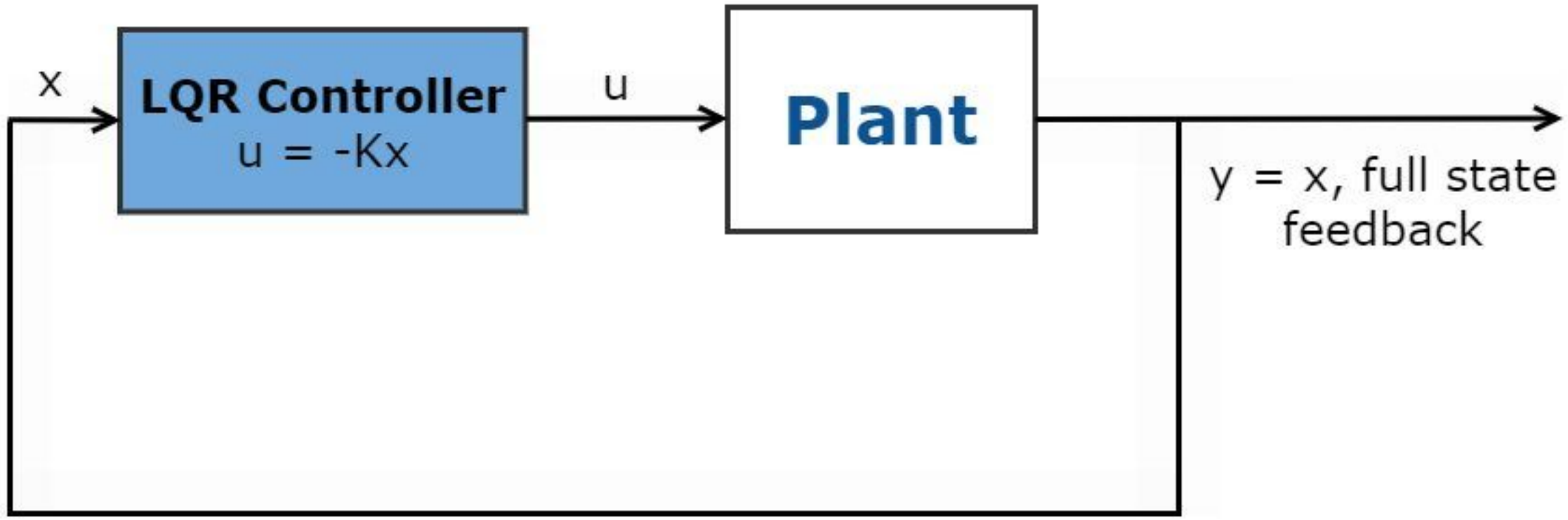
Outline

1. How I met your host
2. Research at STADIUS
3. From research to industry
4. Same same, but different
5. Conclusion

How I met your host

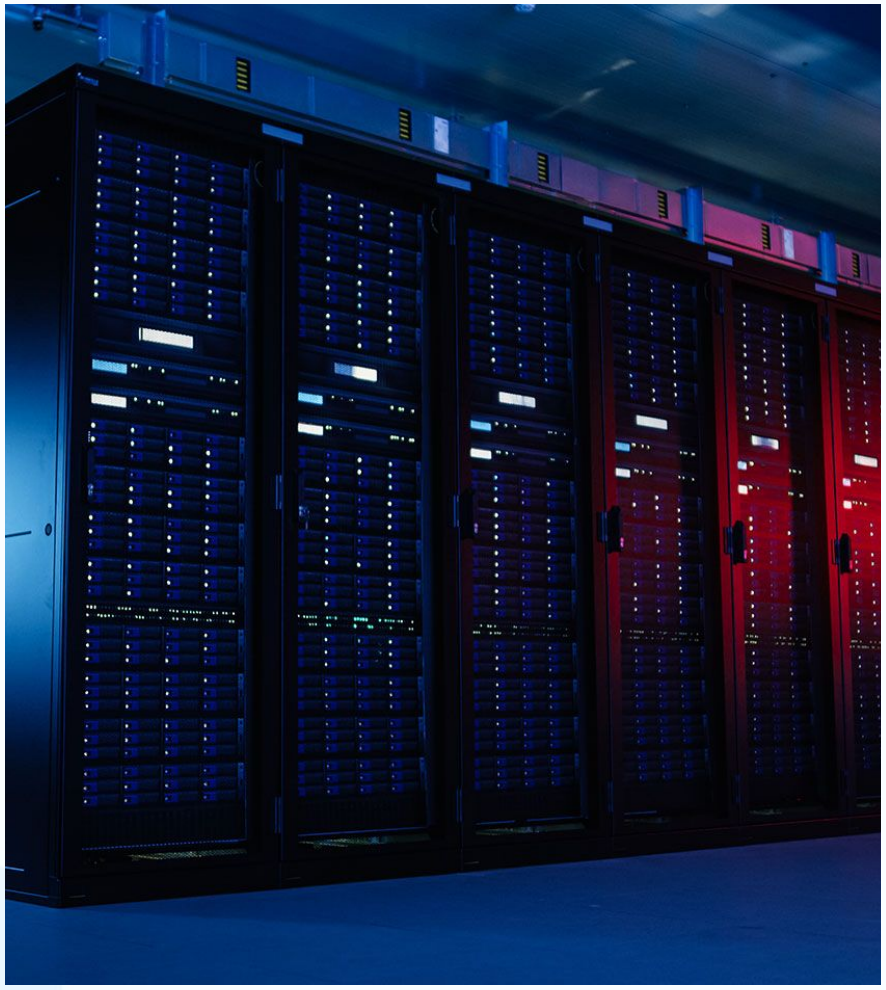
Back to the roots of my interactions with Bart

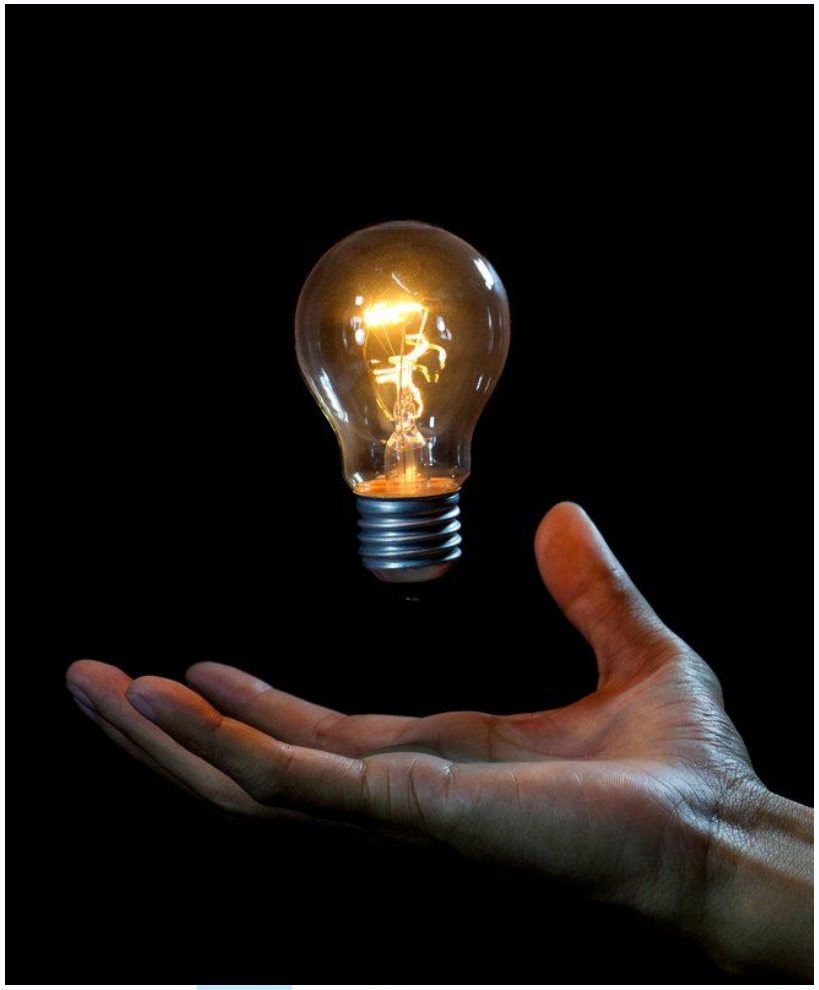
Control theory



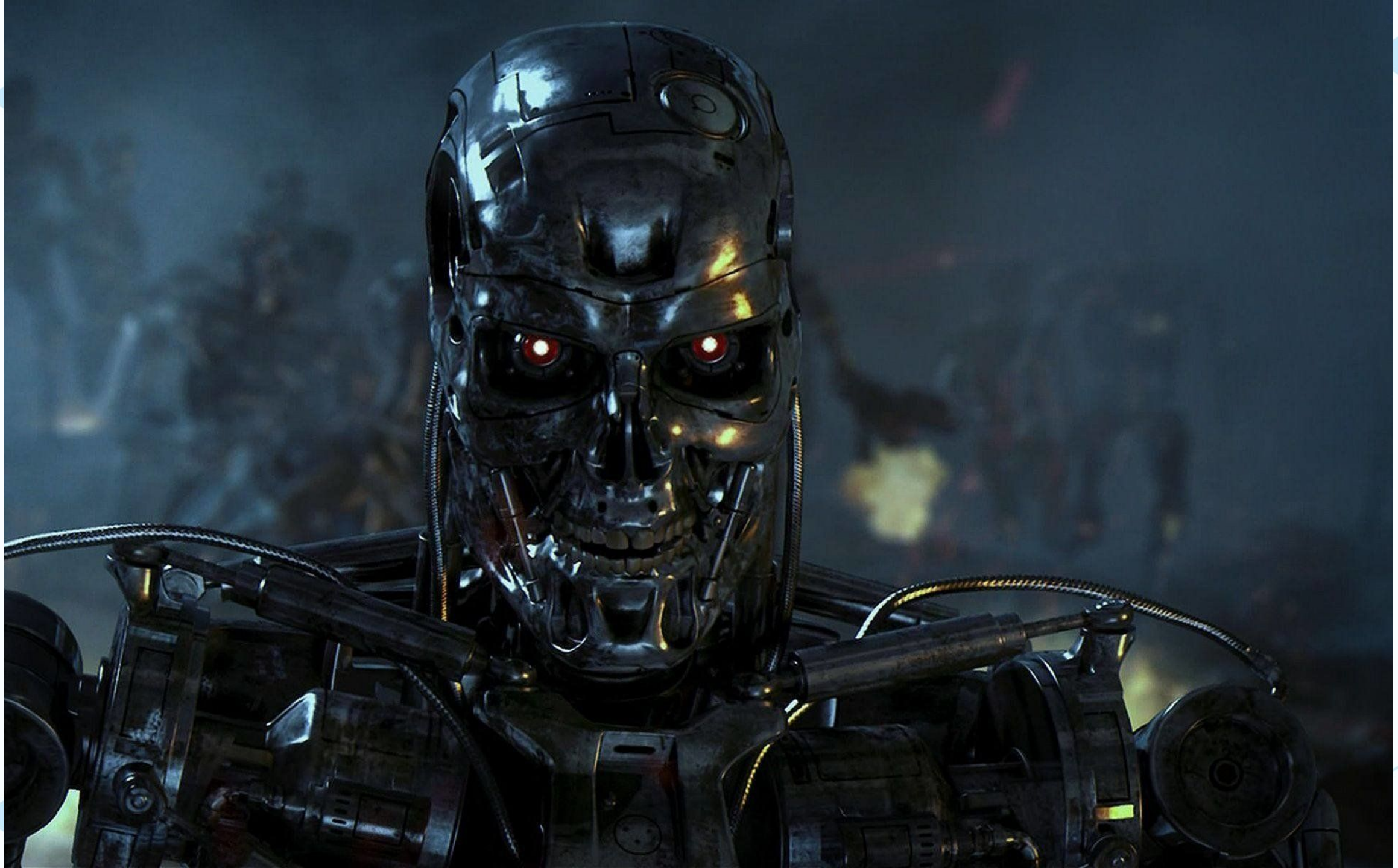
Control theory

Personal interests





PhD in machine learning?



Let's investigate health records!

Research at STADIUS

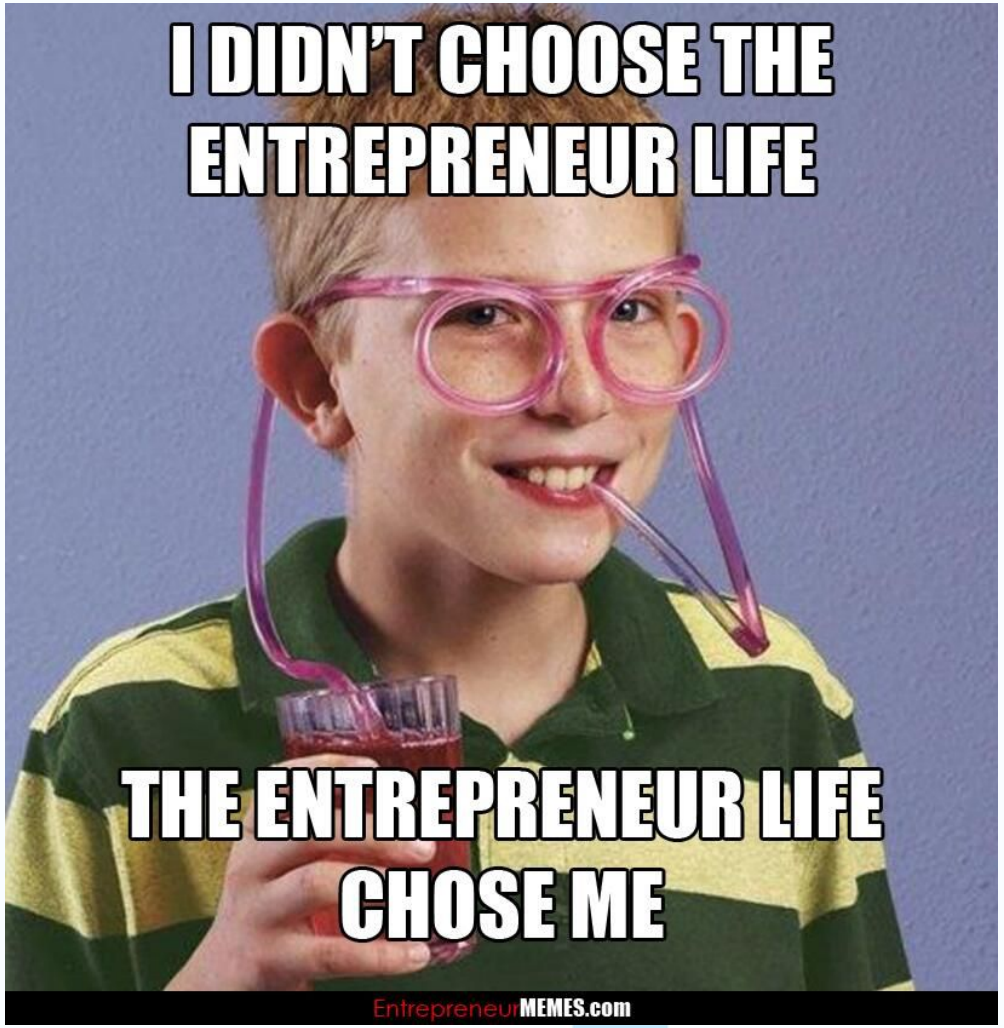
My PhD on machine learning, promoted by Bart De Moor

Project: text mining health insurance records

Project: text mining health insurance records

Project: ???

Project: ???


I DIDN'T CHOOSE THE ENTREPRENEUR LIFE
THE ENTREPRENEUR LIFE CHOSE ME
EntrepreneurMEMES.com

Project: diabetes screening via insurance records

Bart De Moor (promotor)

Frank De Smet (copromotor)

Chantal Mathieu

Johan Suykens

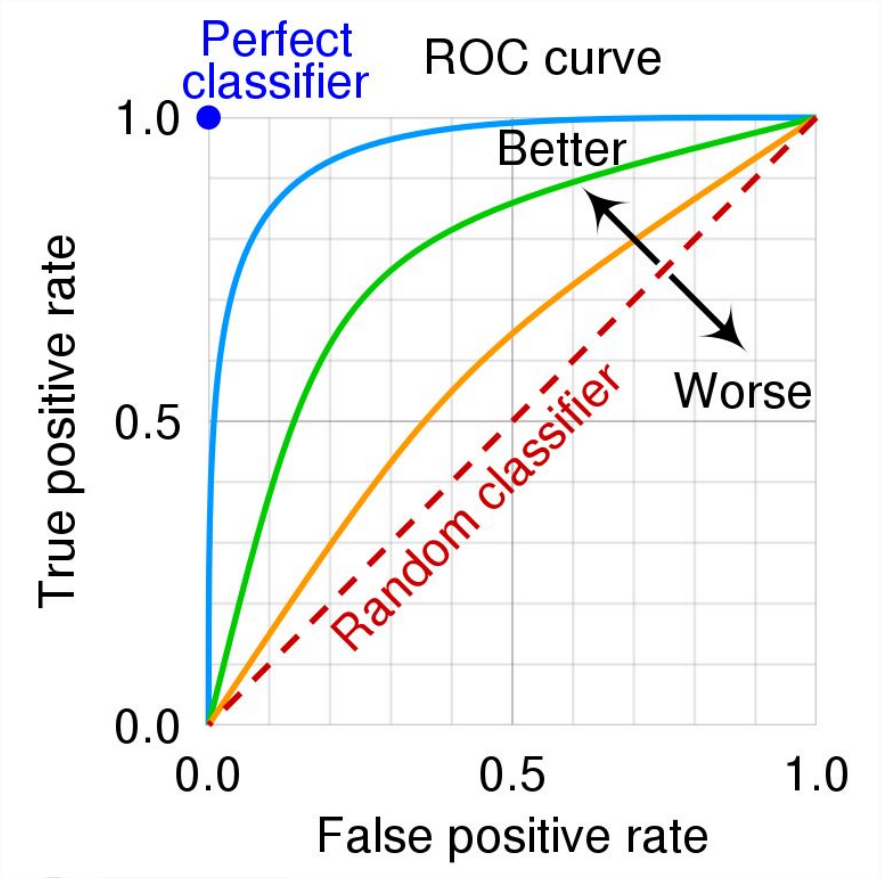
Jesse Davis

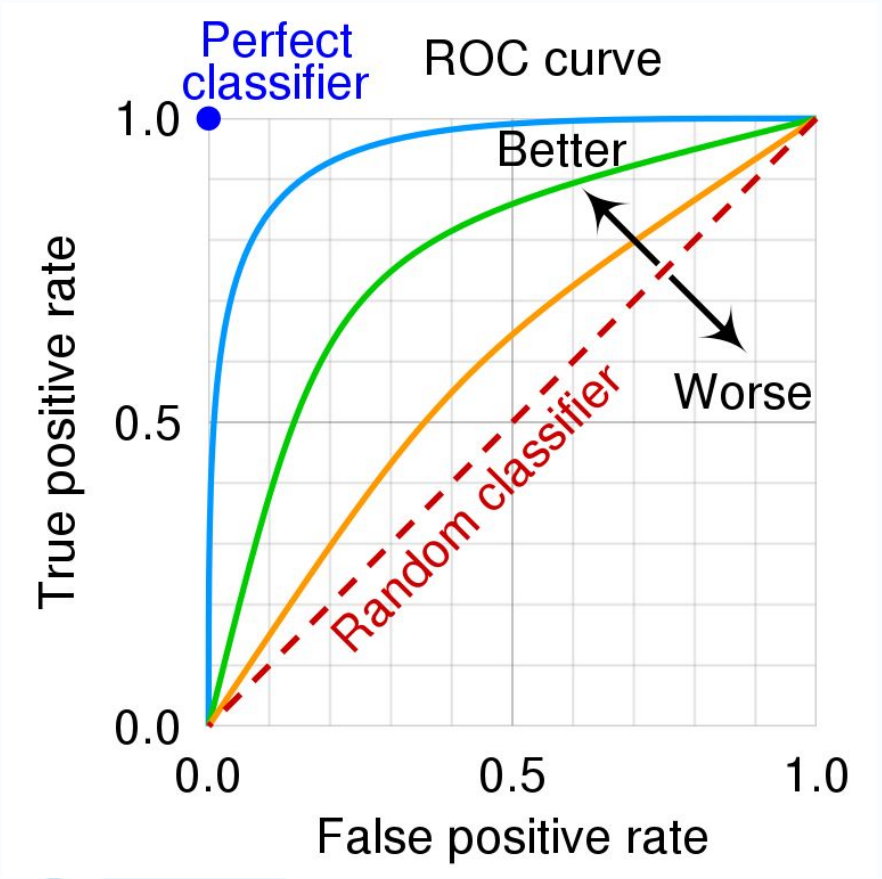
Marc = Piranha

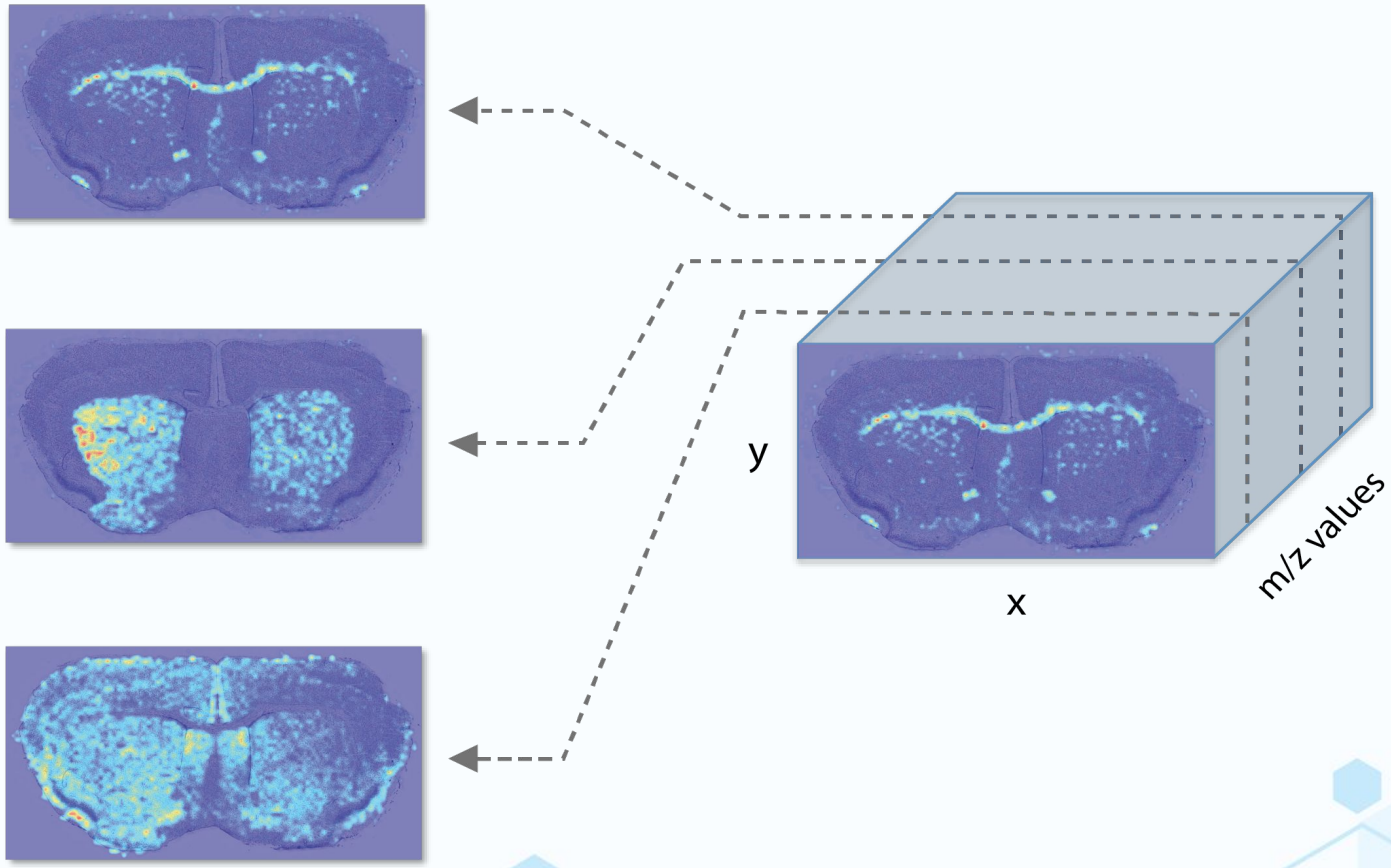
From research to industry

A brief overview of Aspect Analytics.


In Digital We Trust !
Make sure to make impact!
Prof. Dr. Ir. Bart De Moor
KU LEUVEN


Perfect classifier
ROC curve
1.0
Better
Worse
0.5
Random classifier
True positive rate
0.0
0.0
0.5
1.0
False positive rate


Perfect classifier
ROC curve
1.0
Better
Worse
Random classifier
True positive rate
0.5
0.0
0.0
0.5
1.0
False positive rate


y
x
m/z values

Aspect Analytics

Beyond bioinformatics.

AGENTSCHAP INNOVEREN & ONDERNEMEN

Vlaanderen is ondernemen

eit Accelerated by Health

NVIDIA INCEPTION PROGRAM

eurostars™

KU LEUVEN


FAVORITE POSITION?
I WOULD LIKE TO BE THE CEO.
EntrepreneurMEMES.com

Aspect Analytics in a nutshell

Aspect Analytics

Mission

- Dedicated software solutions for mass spectrometry imaging (MSI)
  - focus on high-throughput applications
  - enabling integration with other imaging modalities
  - scalable by design if and when necessary

Our core offering

- Bioinformatics & machine learning
- Cloud-based software platform
- Customized workflows & integrations


EVER
PROJECT WITH
TERABYTES OF
EXPERIMENTAL DATA
MY LAPTOP USED
FOR DATA ANALYSIS

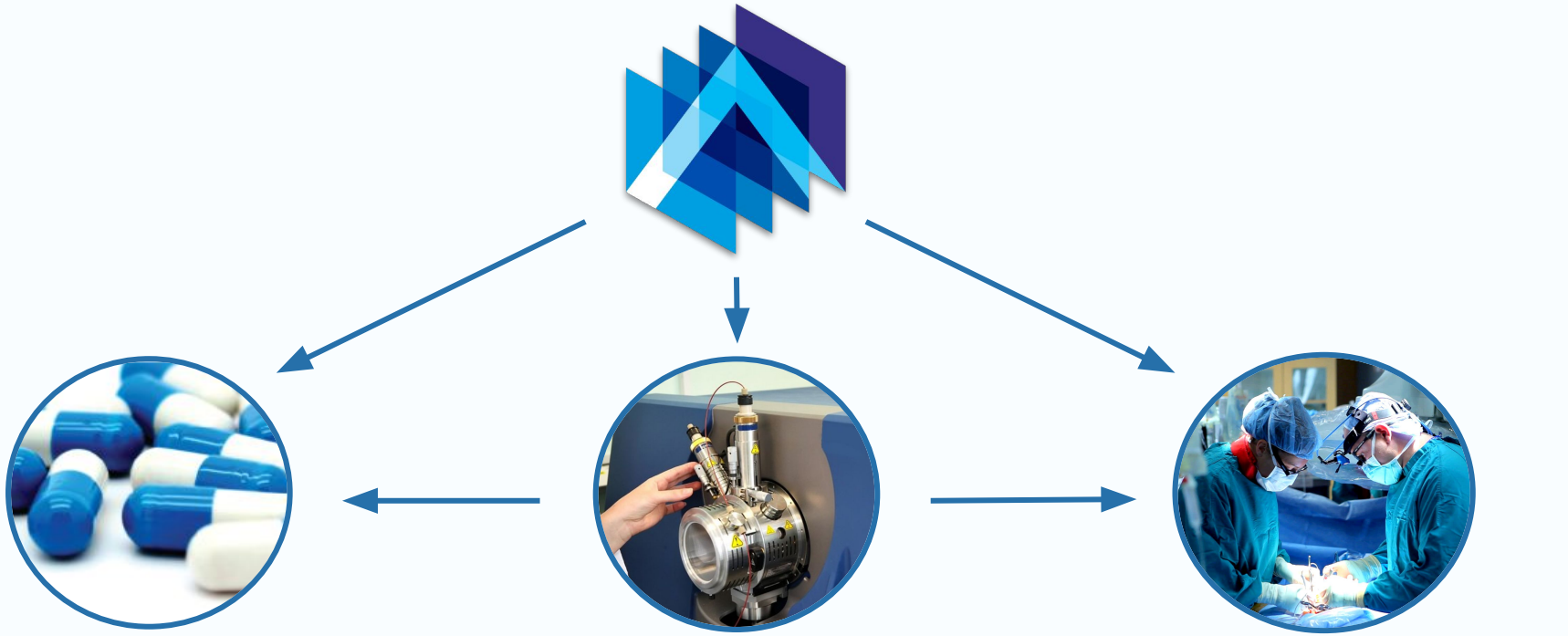
Same same, but different

A brief overview of Aspect Analytics.

Customer segmentation

Aspect Analytics

Focus on **industrial, high-throughput** applications of MSI technology.



Big Pharma & biotech companies

Instrument vendors & service providers

Biomarker discovery & clinical diagnostics

The pharmaceutical R&D pipeline

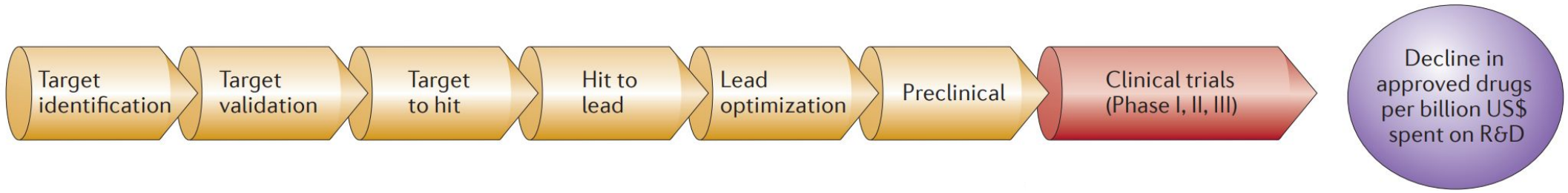


Figure taken from Scannell, Jack W., et al. "Diagnosing the decline in pharmaceutical R&D efficiency." Nature reviews Drug discovery 11.3 (2012): 191.

The pharmaceutical R&D pipeline

Figure taken from Scannell, Jack W., et al. "Diagnosing the decline in pharmaceutical R&D efficiency." Natu iews Drug discovery 11.3 (2012): 191.

Only 12% of compounds entering clinical trials are ultimately approved by the FDA.[1]

- inadequate efficiency (57%)[2]
- safety concerns (17%)[2]
- commercial reasons (22%)[2]

1. Pharmaceutical Research and Manufacturers of America. 2016 biopharmaceutical research industry profile.
2. Hwang, Thomas J., et al. "Failure of investigational drugs in late-stage clinical development and publication of trial results." JAMA internal medicine 176.12 (2016): 1826-1833.

The pharmaceutical R&D pipeline

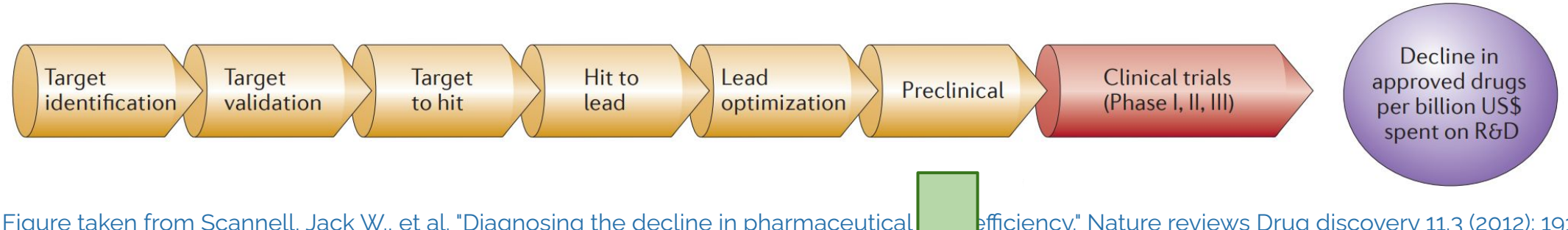


Figure taken from Scannell, Jack W., et al. "Diagnosing the decline in pharmaceutical [illegible]efficiency." Nature reviews Drug discovery 11.3 (2012): 191.

Opportunities in preclinical research:

- ***"fail fast"***: culling unfit compounds early avoids major costs
- **deeper biological insights** can improve entire process

Key challenges in high-throughput

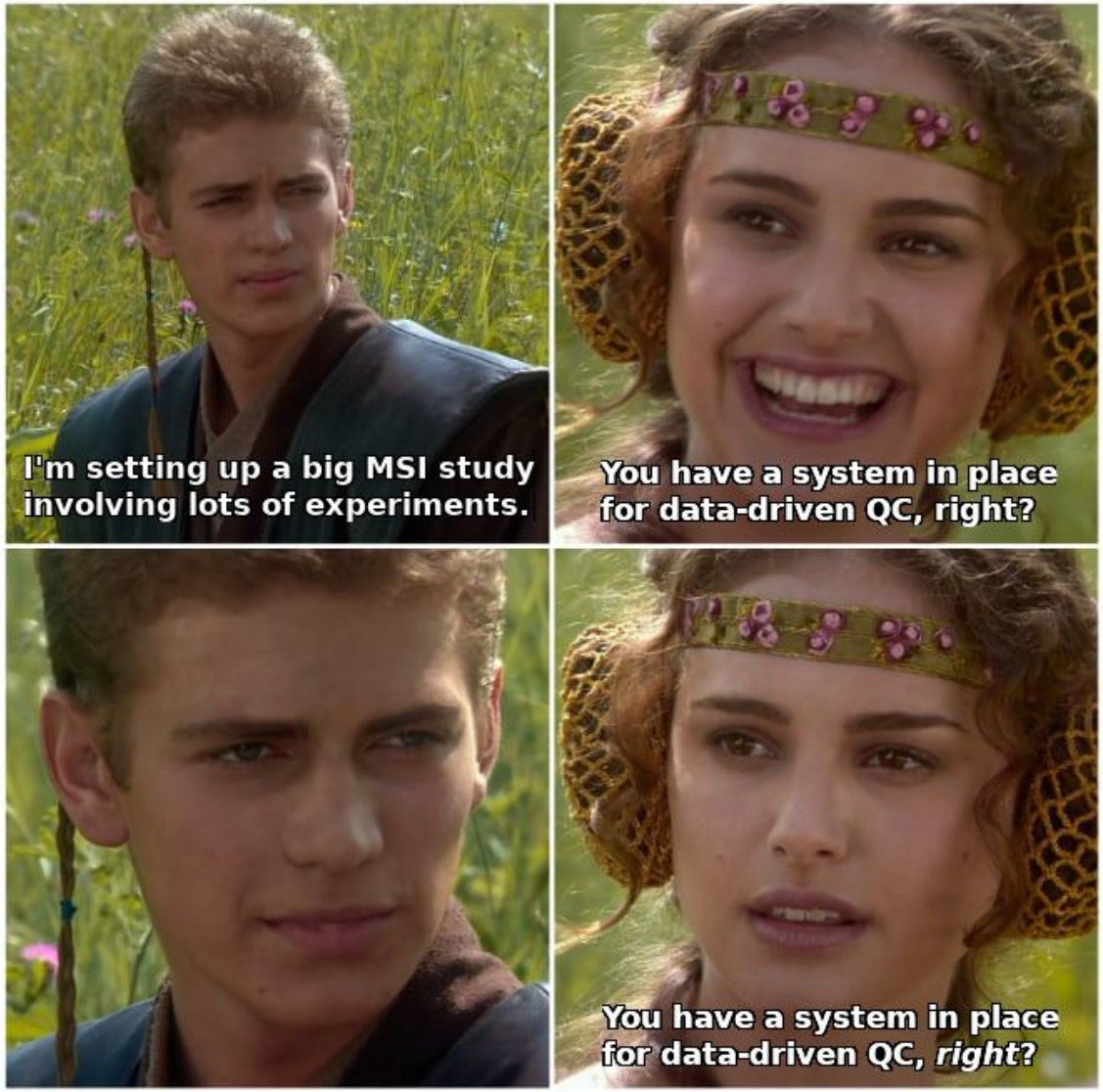
Aspect Analytics

Optimizing human efficiency via workflows

- Full digital pipeline is split into work items
- Individual roles based on expertise
- Streamline communication & collaboration
- Automate data analysis steps when possible

Detecting potential issues via data-driven QC

- Pivotal for robust, high-throughput usage
- Requires deep knowledge of application
- Supported by detailed metadata
- Ideally fully automated



When you try to fail and you succeed …

"When you operate at scale, failure is the norm." - distributed computing adage

In the context of MS imaging, *failures* can come in many forms, e.g.:

- **bad data**: problems with sample / sample prep, instrument issues, …
- **poor data analysis**: information leaks, problematic assumptions, bad fit, …
- **human error**: errors in data input / copying, communication issues, …

How to deal with impending failures

- **observability**: can we identify the presence of certain issues? e.g., bad spectral quality
- **traceability**: can we identify the root cause of an issue? e.g., broken laser
- **prevention**: how to avoid this issue moving forward? e.g., laser QC protocol

Observability reduces risk

Aspect Analytics

	high observability (issues can be detected reliably)	**low observability** (issues cannot be detected reliably)
low impact (probably does not affect conclusions)		
high impact (may significantly affect conclusions)		

Data-driven QC is key

Aspect Analytics

Spend time once to gain time permanently

- Check key assumptions for each sample / batch in an automated way
- Run data-driven QC as a background task, notify team upon anomalies only

Improve reliability and confidence in core outputs

- Rest assured that large swaths of problems are checked by default
- Avoid structural failures that can compromise long-term success

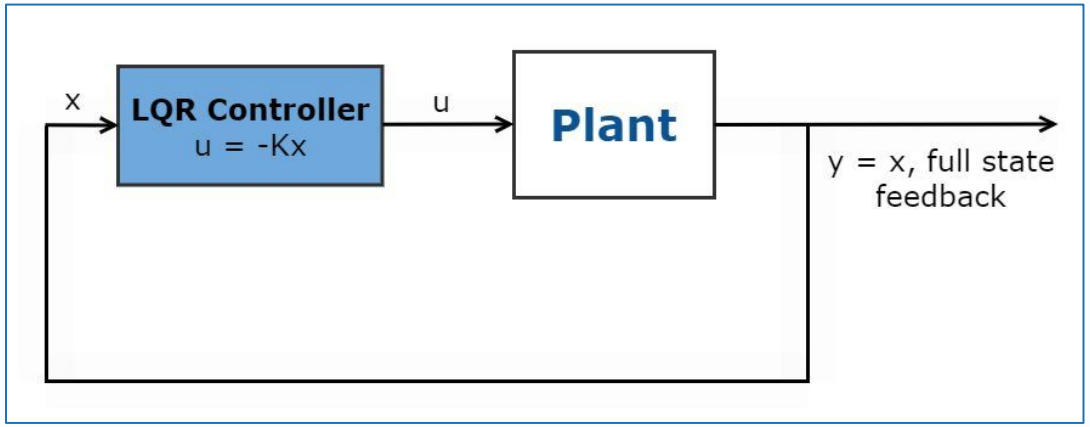
Data-driven QC requires **well-curated meta-data**

- info on all relevant levels: sample / ROI / measurement / …
  - e.g., control vs. treated animal, technical/biological replicates, # laser shots, …
- data lineage to enable traceability

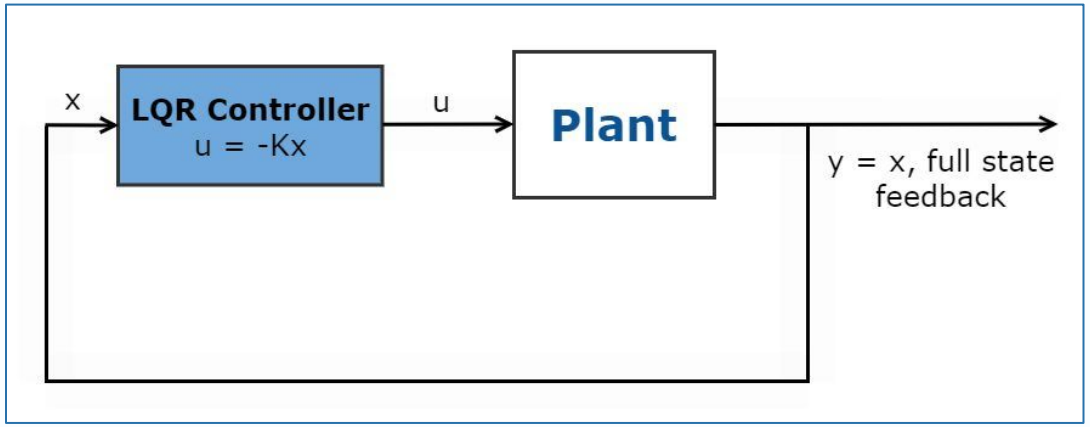
Conclusion

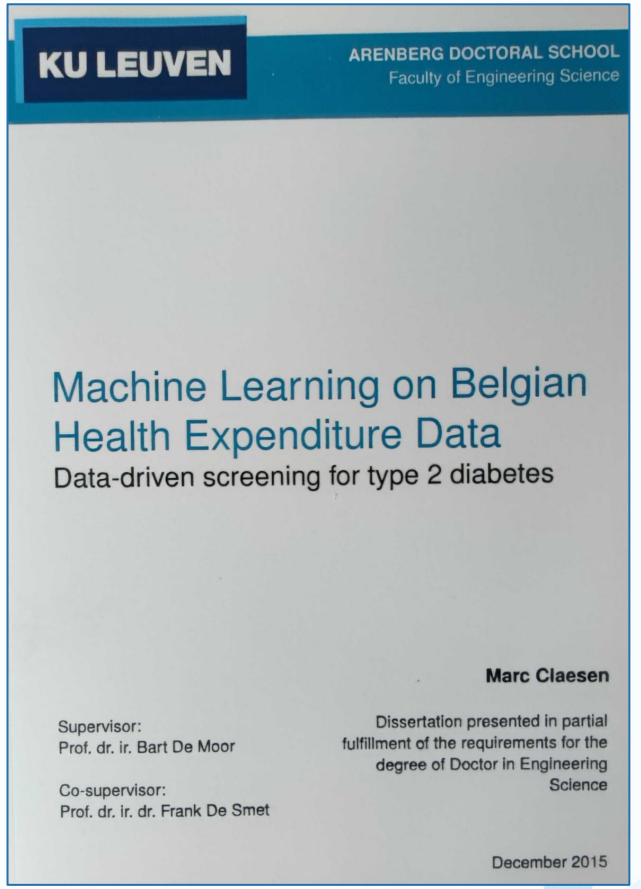
What a long, strange trip it's been

What a long, strange trip it's been

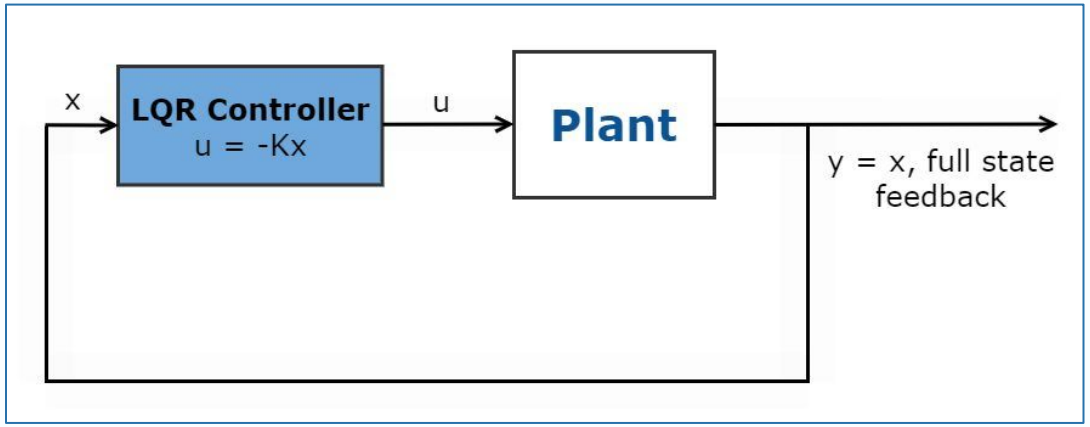


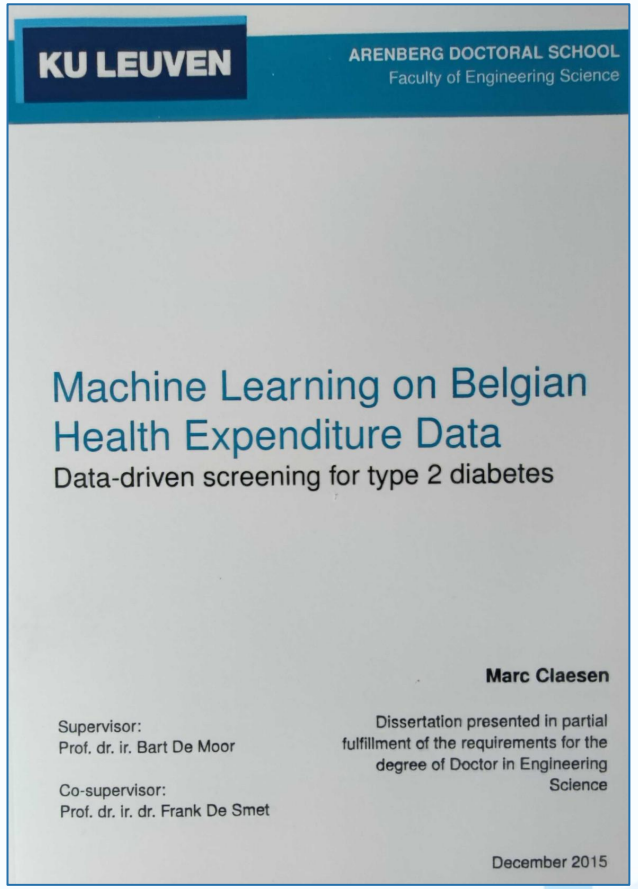
What a long, strange trip it's been





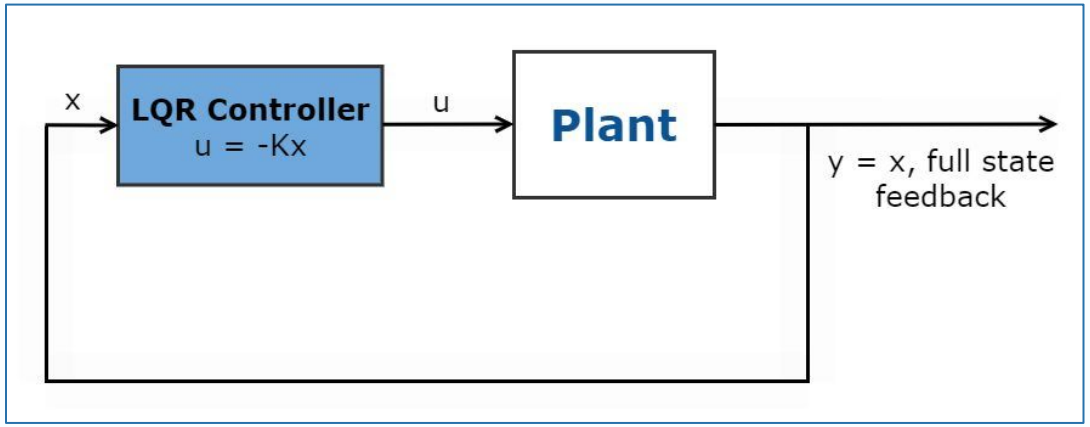
What a long, strange trip it's been

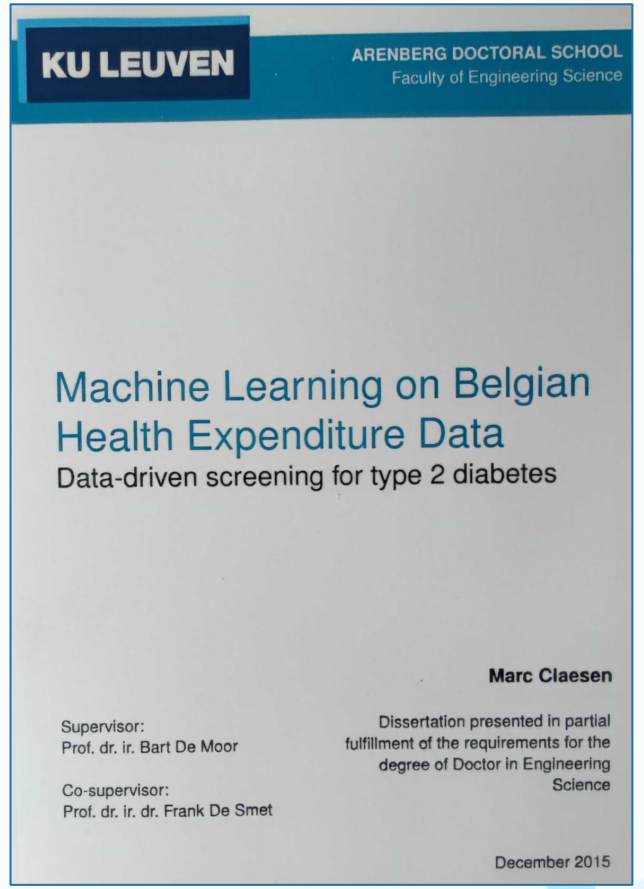




Aspect Analytics

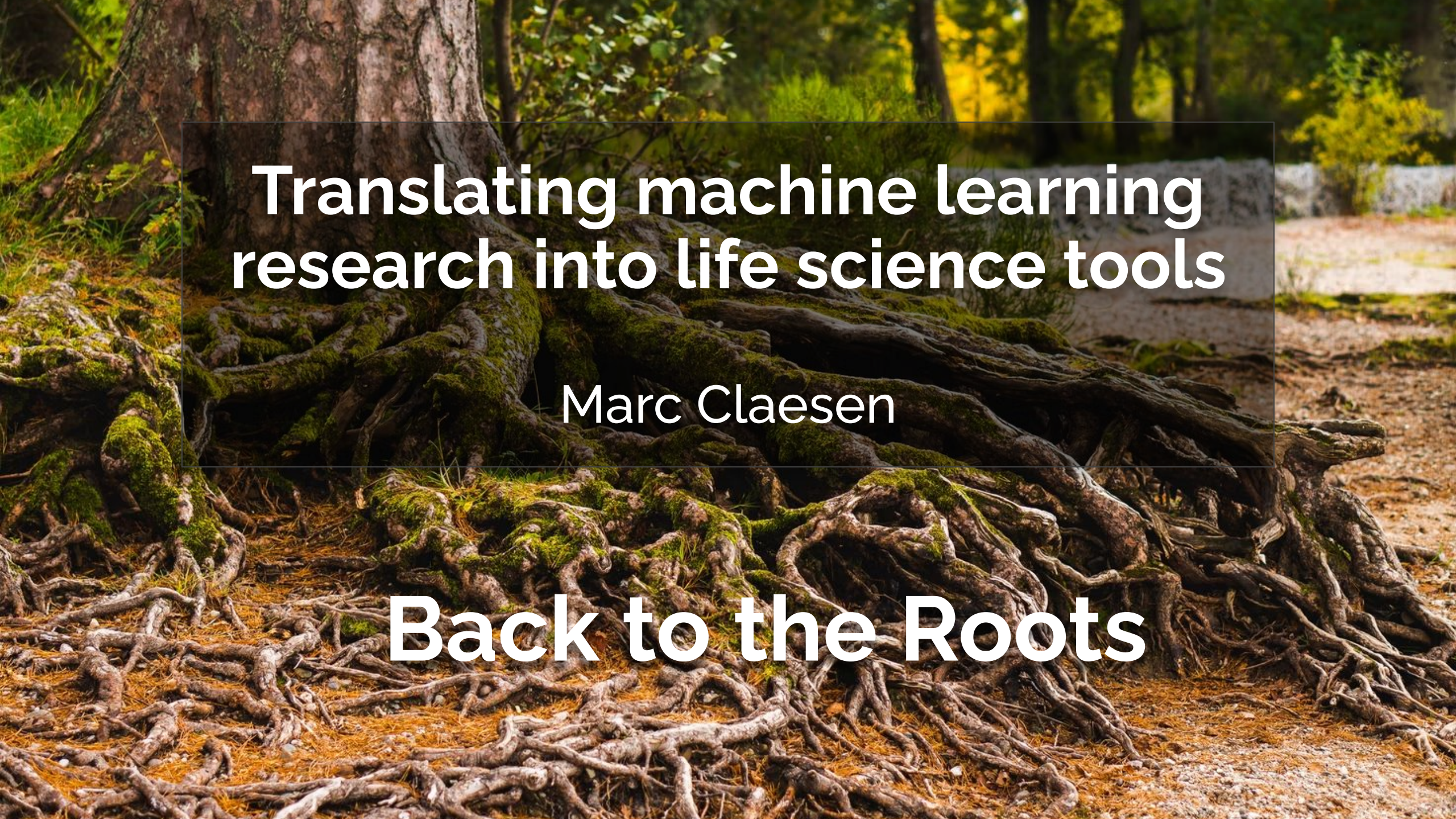
What a long, strange trip it's been





Let's go back to the roots!

Aspect Analytics


Translating machine learning research into life science tools
Marc Claesen
Back to the Roots