Receiving fairness from machine intelligence (fairness. bias. transparency)

Favoritenstr. 9/193-1, 4. th floor

A-1040 Vienna, AUSTRIA

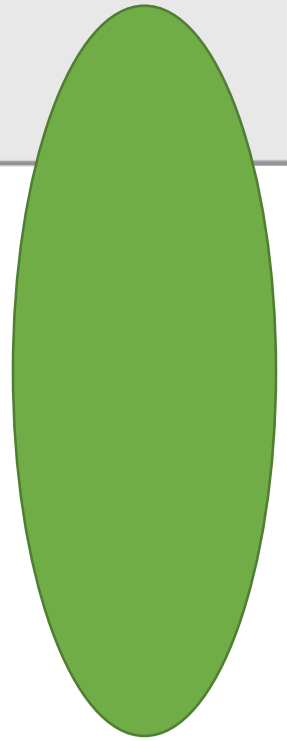
Phone: +43-1-58801-18364

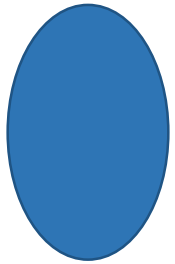
Fax: +43-1-58801-18399

www.cvl.tuwien.ac.at/

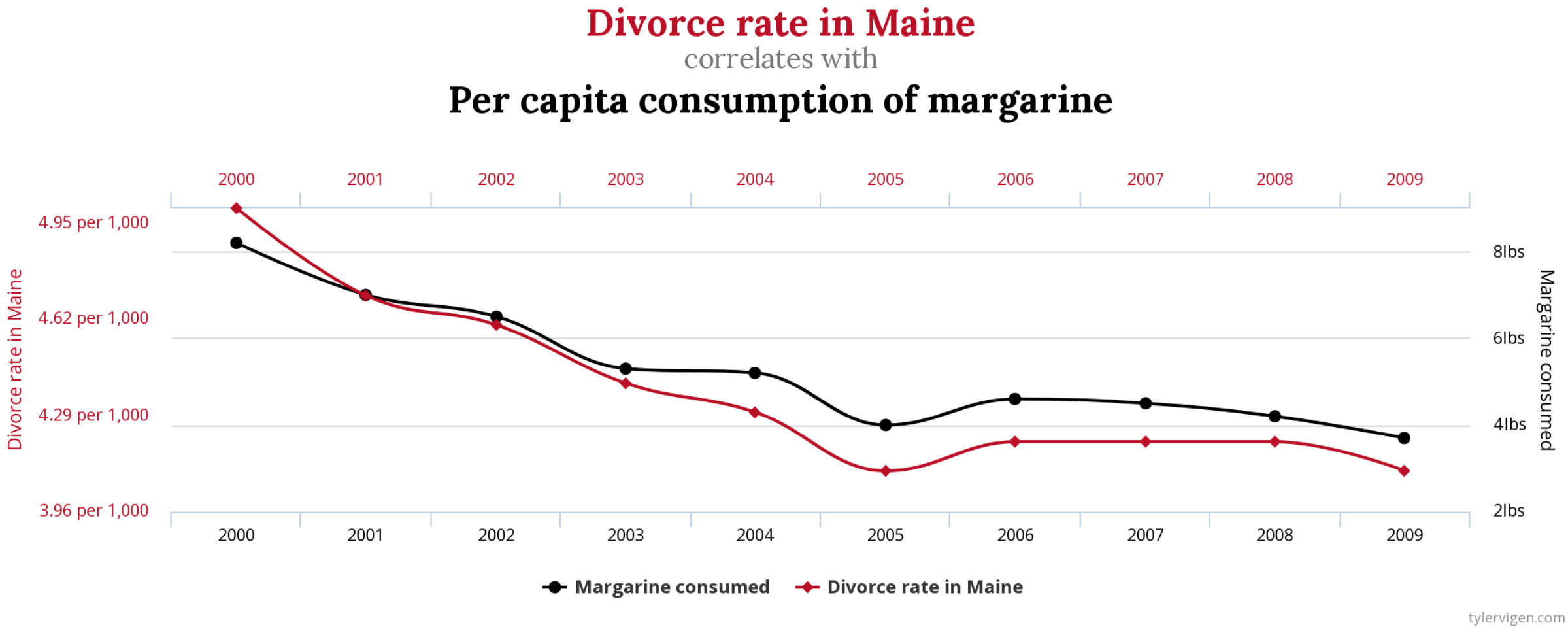
Agenda

- Introduction and Motivation
- Examples
- MIT Moral Machine
- It s the Data
  - Behavior modelling (fall detection)
  - cancer research
  - Detection of suicidal activities
- Views from the developer



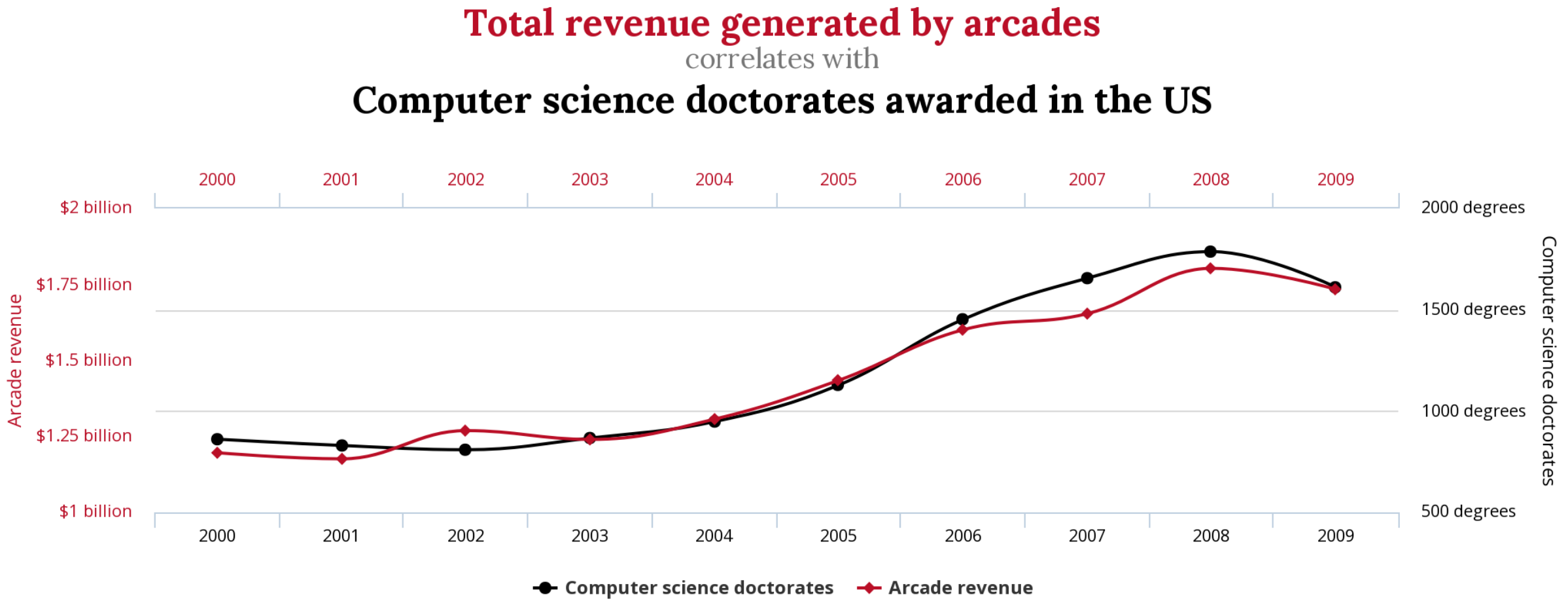


… learning from data .



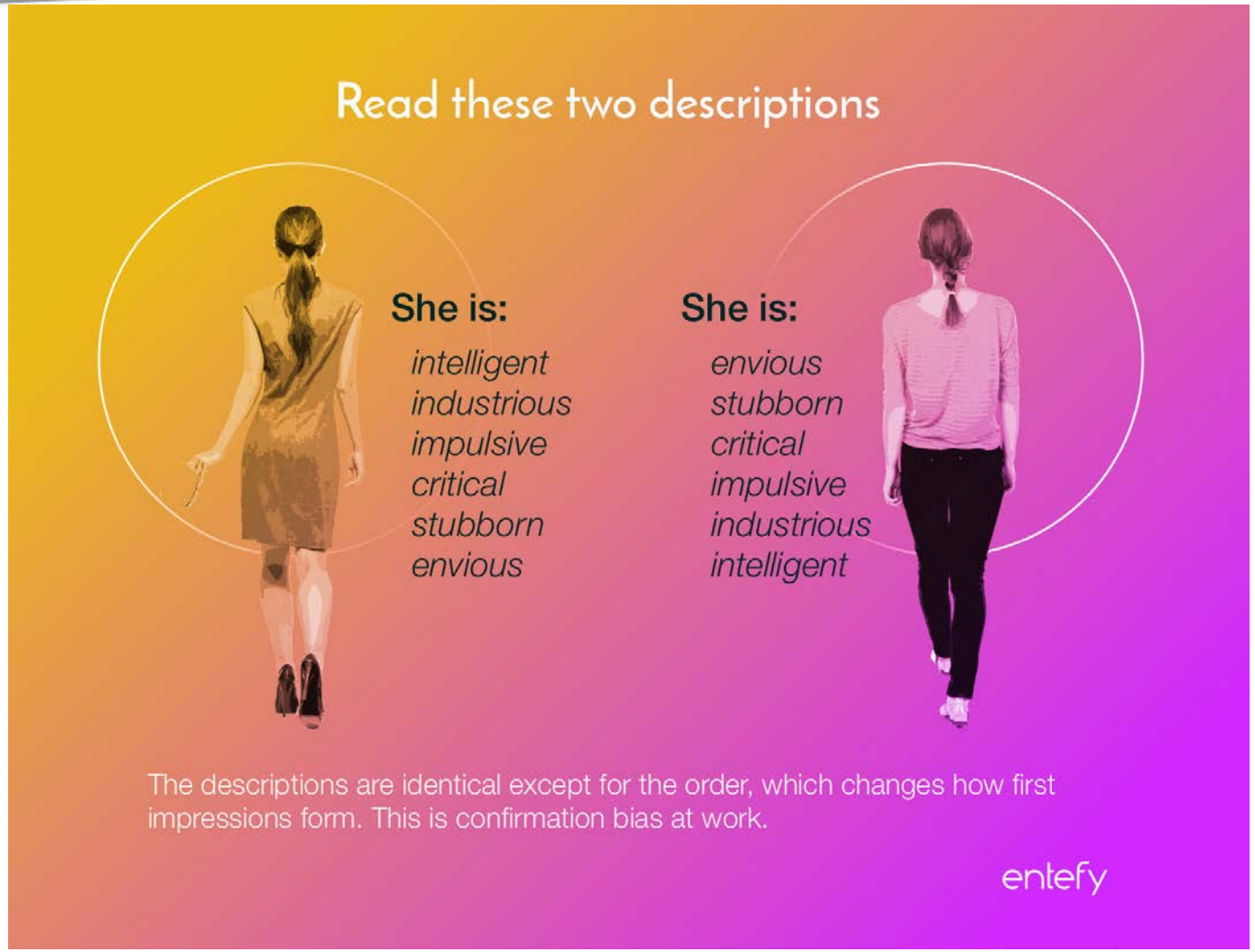
http://www.tylervigen.com/

… learning from data



http://www.tylervigen.com/

Bias at work …



https://www.entefy.com/blog/post/315/the-hazards-of-confirmation-bias-in-life-and-work

Goal of this lecture

Raising awareness towards

Fairness

Bias

and **transparency**

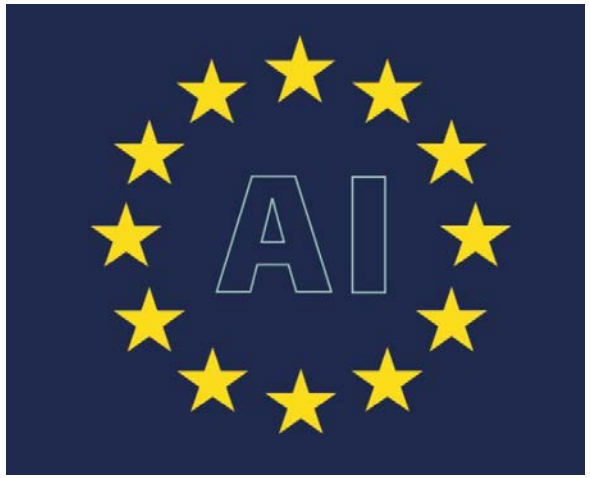
of (learning) algorithms

https://www.wired.com/story/what-does-a-fair-algorithm-look-like/

BIAS

Bias is a prejudice for or against something or somebody, that may result in unfair decisions.

(Ethics Guidelines for trustworthy AI, EU Expert Group AI, Dec 2018)



- Humans are **biased** in their **decision making**. Since AI systems are designed by humans, it is possible that humans inject their bias into them, even in an unintended way.
- Many current AI systems are based on machine learning **data-driven techniques**. Therefore a predominant way to inject bias can be in the **collection and selection of training data**.
- If the training data is not inclusive and **balanced enough**, the system could learn to make **unfair decisions**. At the same time, AI can help humans to identify their biases, and assist them in making less biased decisions.

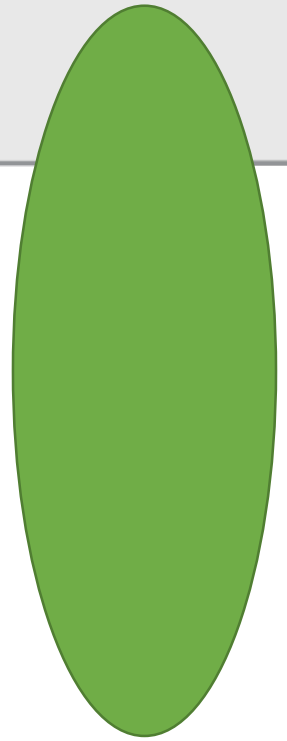
- Entails the capability to **describe**, **inspect** and **reproduce** the mechanisms through which AI systems make decisions and learn to adapt to their environments
- Transparency is key to building and maintaining **citizen's trust** in the developers of AI systems and AI systems themselves
- **Technological transparency** implies that AI systems be auditable, comprehensible and intelligible by human beings at varying levels of comprehension and expertise.
- **Business model transparency** means that human beings are knowingly informed of the intention of developers and technology implementers of AI systems.

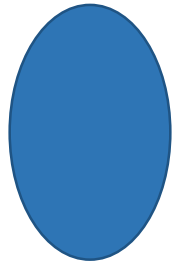
Fairness

- Individual fairness: If we're being individually fair, then similar individuals get treated similarly.
  We define "similarity" in a way that ignores protected categories (like race, gender, etc.)

Agenda

- Introduction and Motivation
- Examples
- MIT Moral Machine
- It‘s the Data
  - Behavior modelling (fall detection)
  - Cancer research
  - Detection of suicidal activities
- Views from the developer





Nikon: Face detection (from 2010)

- Nikon-cameras (Nikon Coolpix S630) provide face detection
- Asian faces not detected appropriately
  - Message from the camera „Did someone blink?“

- Problem: Schmales und halb geschlossenes Auge schwer zu erkennen
  - Auge ist nur wenige Pixel groß + Downsampling der Kamera
- Webcam von HP Laptop: Problem mit Face-Tracking
  - Erkennt Gesichter von schwarzen Personen nicht

Google Translate: Gender Bias

- Online Translation
  - Übersetzt Wörter, Sätze und Webseiten
  - Über 100 Sprachen verfügbar
- Gender Bias noted
  - Anwendung auf ursprünglich geschlechtsneutrale Begriffe
  - z.B.: „the secretary“ -> „die Sekretärin“, „the boss“ -> „der Chef“
- NOW: Google Translate learns to reduce gender bias

- Test: Direktor -> director; Direktorin -> Headmistress
- Geschäftsführer -> Executive Director
- Geschäftsführerin –> manager

Google Translate: Gender Bias

- Basieren aus Daten aus der Vergangenheit
  - Algorithmen erschweren es aus Stereotypen auszubrechen
  - Daten kommen direkt aus den Online-Aktivitäten der Gesellschaft
  - Gesellschaft ist vorbelastet -> generierte Daten sind vorbelastet -> Algorithmen sind vorbelastet
- Google ist Reflektion der vorherrschenden Bias
- Google hat Potential um auf Bias aufmerksam zu machen und dagegen zu wirken

https://towardsdatascience.com/a-gentle-introduction-to-the-discussion-on-algorithmic-fairness-740bbb469b6

Northpointe/Equivant: Predicting Crimes

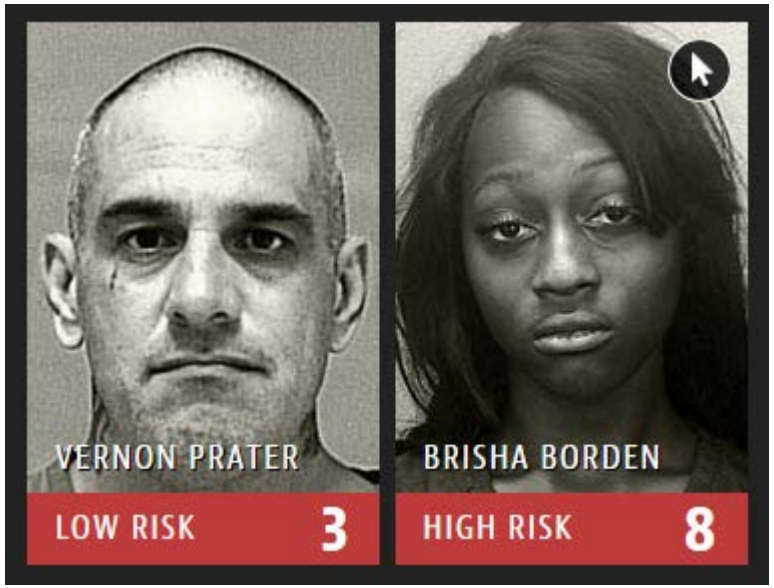
- COMPAS tool is widely used to assess a defendant's risk of committing more crimes
  - Risc Assessment to predict crime and re-offending
  - Beeinflusst u.a. Freilassung und Kautionsbeträge
  - Entscheidung über Ausmaß für Rehabilitationsmaßnahmen
- Einsatz zukünftig bei jedem Schritt eines strafrechtlichen Prozesses

Northpointe: Risikoanalyse bei Straftätern

- Northepointe ist weit verbreitetes Bewertungstool für Straftäter
- Bewertet Wiederholungstäter unter anderem nach Hautfarbe
  - Wahrscheinlichkeit für Wiederholungstäter bei schwarzen Personen höher
  - Schwarze Personen doppelt so oft falsch als Wiederholungstäter eingeschätzt als weiße Personen
  - Weiße Personen öfter falsch als ungefährlich eigeschätzt als schwarze Personen

Northpointe: Risikoanalyse bei Straftätern

	Vernon Prater	Brisha Borden
Verhaftungsgrund	Kleiner Diebstahl	Kleiner Diebstahl
Vorstrafen	• 3 bewaffnete Raubüberfälle • 1 versuchter bewaffneter Raubüberfall	4 Jugendstraftaten
Spätere Straftaten	1 großer Diebstahl	-
Bewertung	**GERINGES RISIKO: 3**	**HOHES RISIKO: 8**

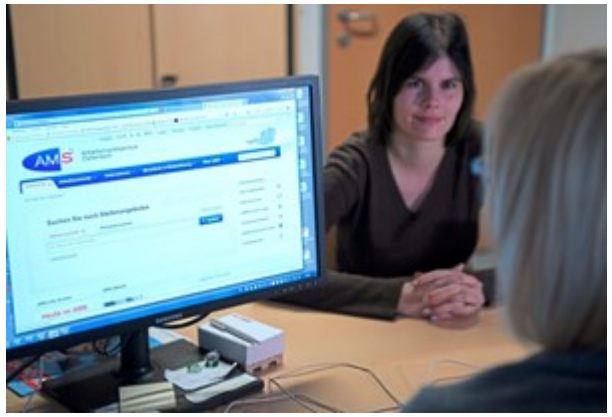


Google Fotos: Gesichtserkennung und Tagging

- **Google Fotos** ermöglicht Speicherung von Bildern und Videos
- Bilder und Videos werden automatisch mit Tags versehen
- Mehrere Vorfälle von **diskriminierenden Tags**
  - **Personen mit dunkler Haut als Gorillas getaggt**
  - Personen mit weißer Haut als Hunde oder Robben getaggt
- Öffentliche kulturell unangebrachte Annahmen

AMS – Arbeitslosenklassifikation (Austria 2018)

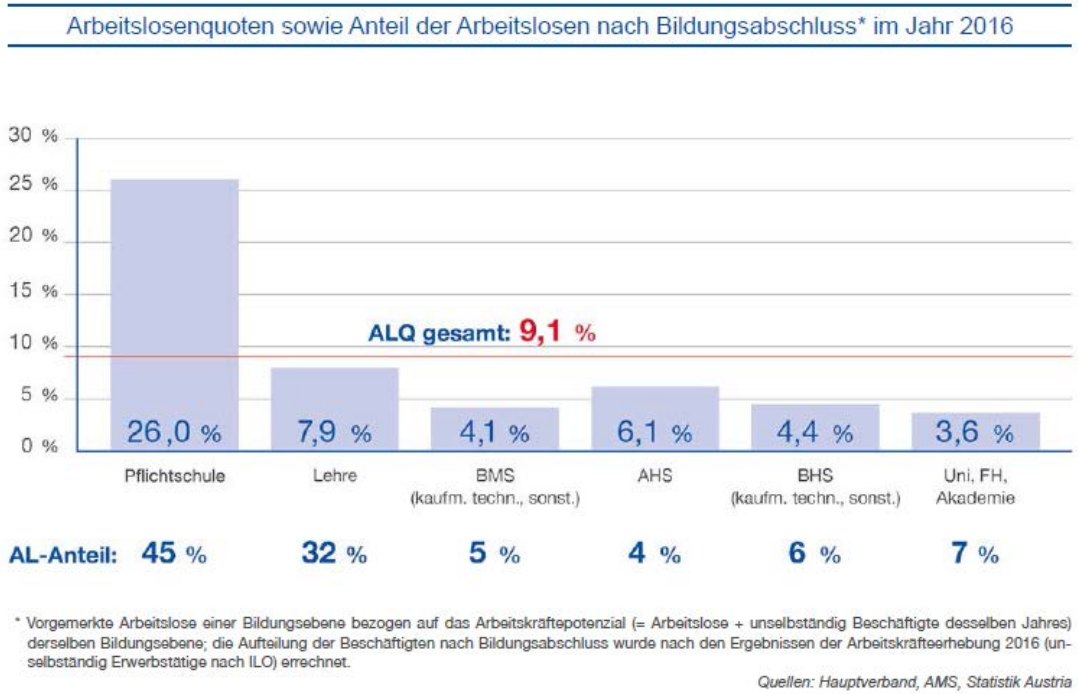
- Soll das Potenzial von Arbeitslosen flächendeckend bestimmen
- Arbeitsuchende werden in drei Kategorien eingeteilt:
  - hohe Chancen
  - mittlere Chancen
  - niedrige Chancen
- Chancen am Arbeitsmarkt einen Job zu finden.
- Motivation: Ressourcen der Arbeitsmarktpolitik langfristig effizient einsetzten



AMS - Arbeitslosenklassifikation

Persönliche Merkmale und der bisherige Erwerbsverlauf und vorangegangene AMS-Geschäftsfälle werden für den Algorithmus berücksichtigt

- Alter
- Geschlecht
- Staatsbürgerschaft
- Ausbildung
- Gesundheitliche Einschränkungen
- Bisherigen Berufe
- Ausmaß einer Beschäftigung
- Die Häufigkeit und die Dauer von Geschäftsfällen



AMS - Arbeitslosenklassifikation

Aus dem AMS-Arbeitsmarkt-chancen-Modell:

BE_INT
= f (0,10
– 0,14 x GESCHLECHT_WEIBLICH
– 0,13 x ALTERSGRUPPE_30_49
– 0,70 x ALTERSGRUPPE_50_PLUS
+ 0,16 x STAATENGRUPPE_EU
– 0,05 x STAATENGRUPPE_DRITT
+ 0,28 x AUSBILDUNG_LEHRE
+ 0,01 x AUSBILDUNG_MATURA_PLUS
– 0,15 x BETREUUNGSPFLICHTIG
– 0,34 x RGS_TYP_2
– 0,18 x RGS_TYP_3
– 0,83 x RGS_TYP_4
– 0,82 x RGS_TYP_5
– 0,67 x BEEINTRÄCHTIGT
+ 0,17 x BERUFSGRUPPE_PRODUKTION
– 0,74 x BESCHÄFTIGUNGSTAGE_WENIG
+ 0,65 x FREQUENZ_GESCHÄFTSFALL_1
+ 1,19 x FREQUENZ_GESCHÄFTSFALL_2
+ 1,98 x FREQUENZ_GESCHÄFTSFALL_3_PLUS
– 0,80 x GESCHÄFTSFALL_LANG
– 0,57 x MN_TEILNAHME_1
– 0,21 x MN_TEILNAHME_2
– 0,43 x MN_TEILNAHME_3)

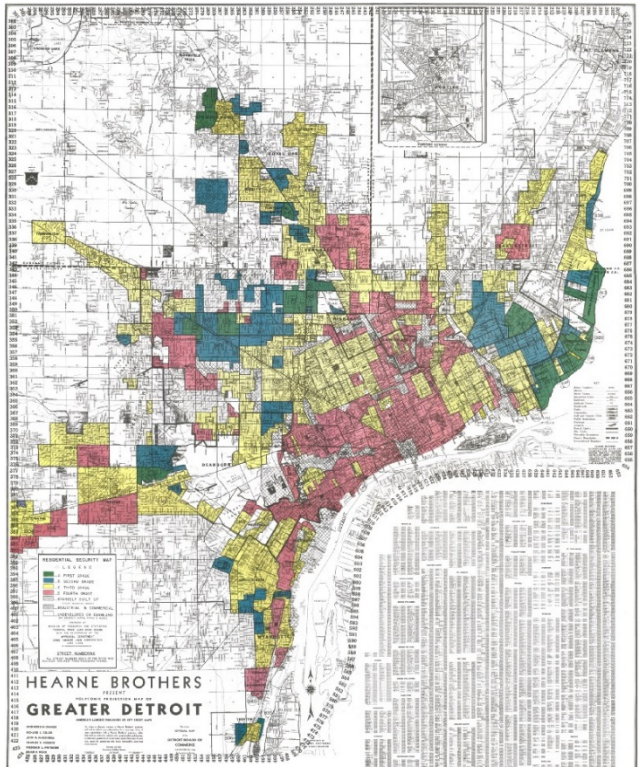
-0,14 x Geschlecht weiblich
-0,7 x Altersgruppe 50 plus
-0,67 x gesundheitliche Beeinträchtigung

+0,16 x Staatengruppe EU
+0,28 x Ausbildung Lehre

Loan algorithms/ red lining

- Analyzing SMSs, utility & credit bill payments, social media profiles, e-commerce purchase patterns, mobile phone usage and behavioral patterns to evaluating educational and professional backgrounds of individuals, algorithms today are directing almost all consumer technology.

- “Data as a weapon”: Between 1934 and 1968 the US Federal Housing Administration systematically denied loans to black people by using entire neighbourhoods, colour-coded by perceived risk factor, as their decision-making metric.
- “AI is upgrading institutional racism”



Detroit 1938

James A Berkovec, Glenn B Canner, Stuart A Gabriel, and Timothy H Hannan. 1994. Race, redlining, and residential mortgage loan performance. Journal of Real Estate Finance and Economics 9, 3 (1994), 263–294

Social Credit System - 社会信用体系; shèhuì xìnyòng tǐxì)

- National reputation system from Chinese government
- Assessment of citizens' and businesses' economic and social reputation and mass surveillance
- Parameters include bad driving, smoking in non-smoking zones, buying too many video games and posting fake news online, but also education, postings, etc

- Output: Restriction to travel (buying train or flight tickets), job openings, etc

How I'm fighting bias in algorithms | Joy Buolamwini

https://youtu.be/UG_X_7g63rY

Visual Computing Trends 2019, 31.01.19

Talk **Andrew Glassner**, The Best of Algorithms, the Worst of Algorithms

Will you get a promotion or raise at your job?

Will you be accepted into the university you're hoping for?

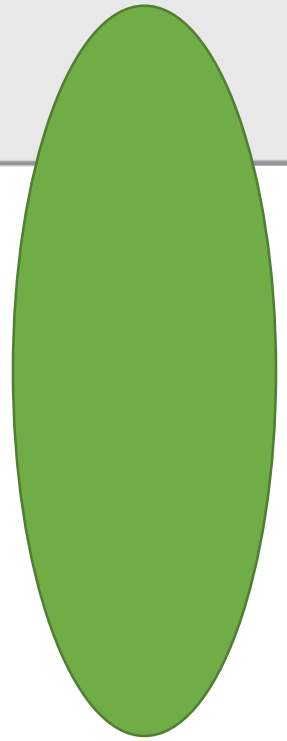
Should you have surgery for a medical problem?

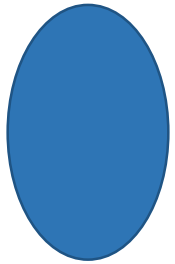
Will you get that loan you need to fix up your home?

- answered by a new class of learning algorithms.
- used by companies, governments, educational institutions, hospitals, banks, courthouses, and other organizations that make decisions that directly affect our lives.
- algorithms are efficient, and presumed to be fair.
- In fact, it has become clear that these algorithms are routinely and inherently unfair, because they solidify, perpetuate, and amplify human biases and prejudices at an unprecedented scale. This creates a potential for enormous harm. If we acknowledge the problem and strive to address it with care and foresight, we can produce better tools that respect and benefit all citizens.

Agenda

- Introduction and Motivation
- Examples
- MIT Moral Machine
- It s the Data
  - Behavior modelling (fall detection)
  - cancer research
  - Detection of suicidal activities
- Views from the developer

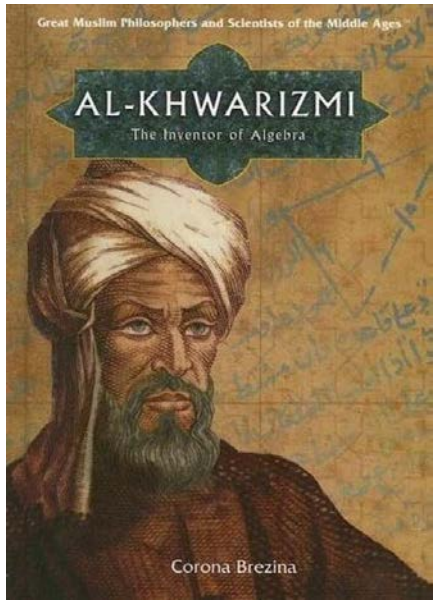


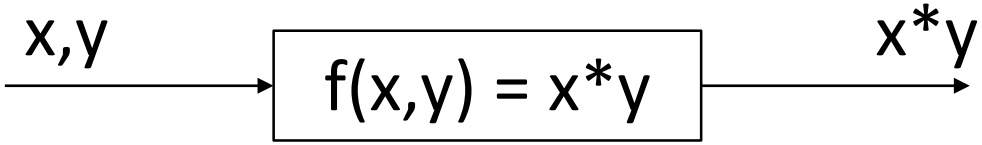


Algorithm

A set of instructions for solving a problem

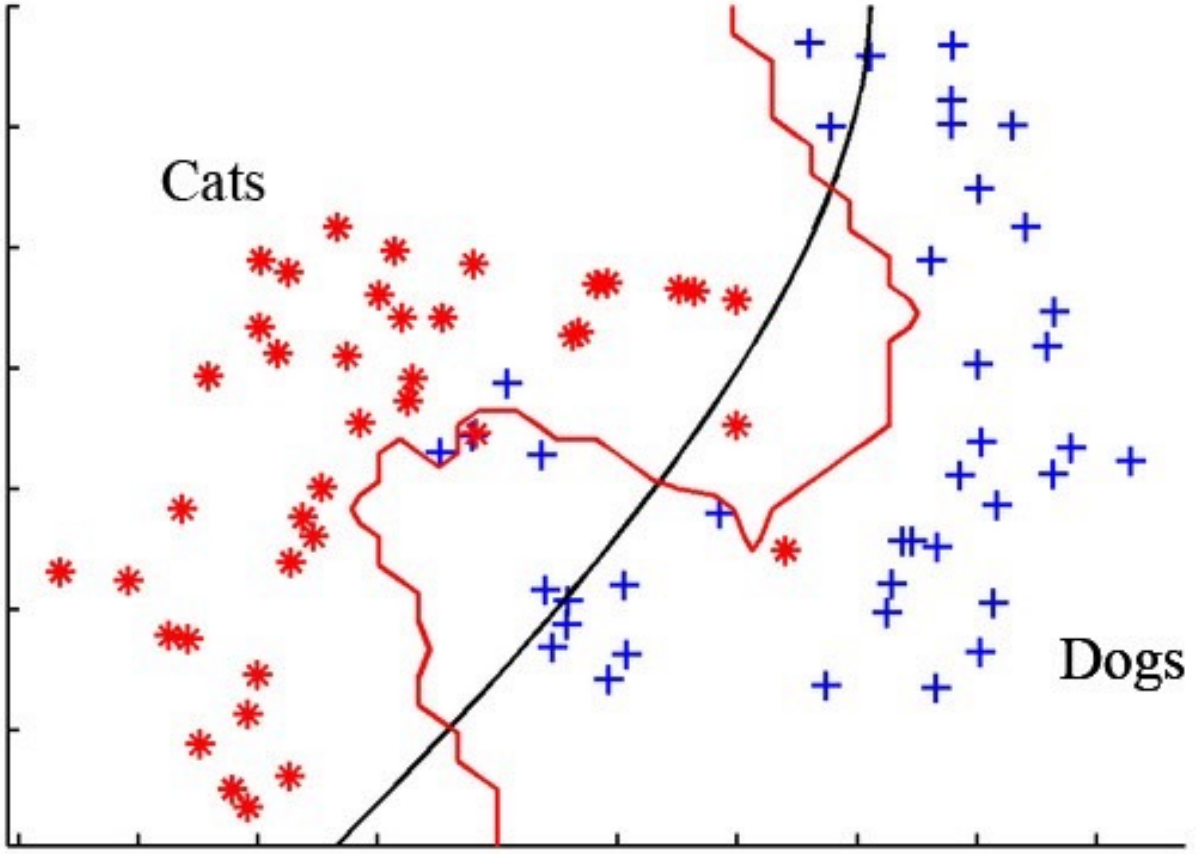
Input → Set of Rules (defining a function) → Output





https://www.scriptol.com/programming/algorithm-definition.php

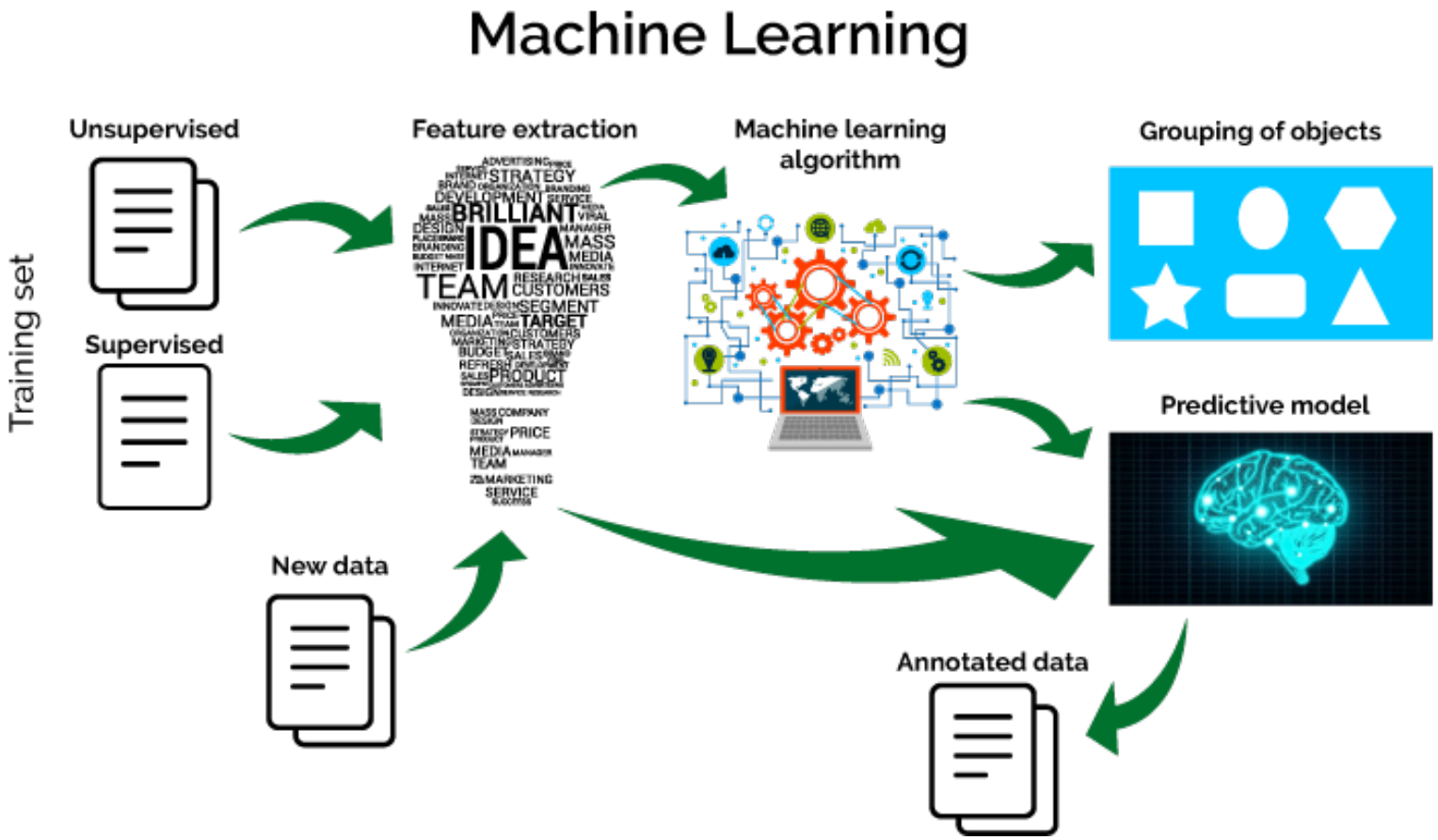
Basic Machine Learning Algorithm



1. Get lots of images depicting cats and dogs
2. Label the images with the correct category
3. Find a separator that can determine if an image depicts a dog or a cat

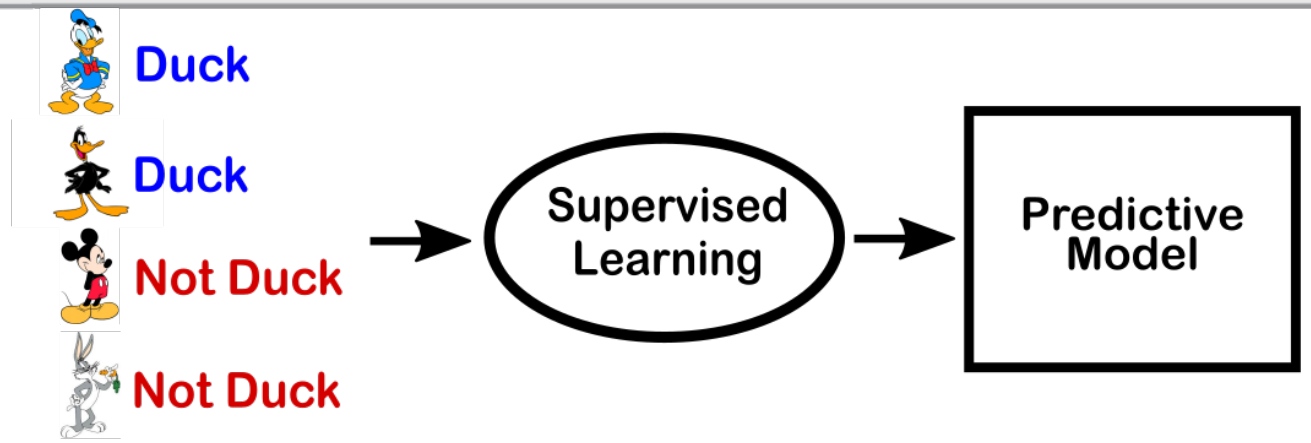
Machine Learning

It’s representation, probability, and algorithms.



Supervised Learning

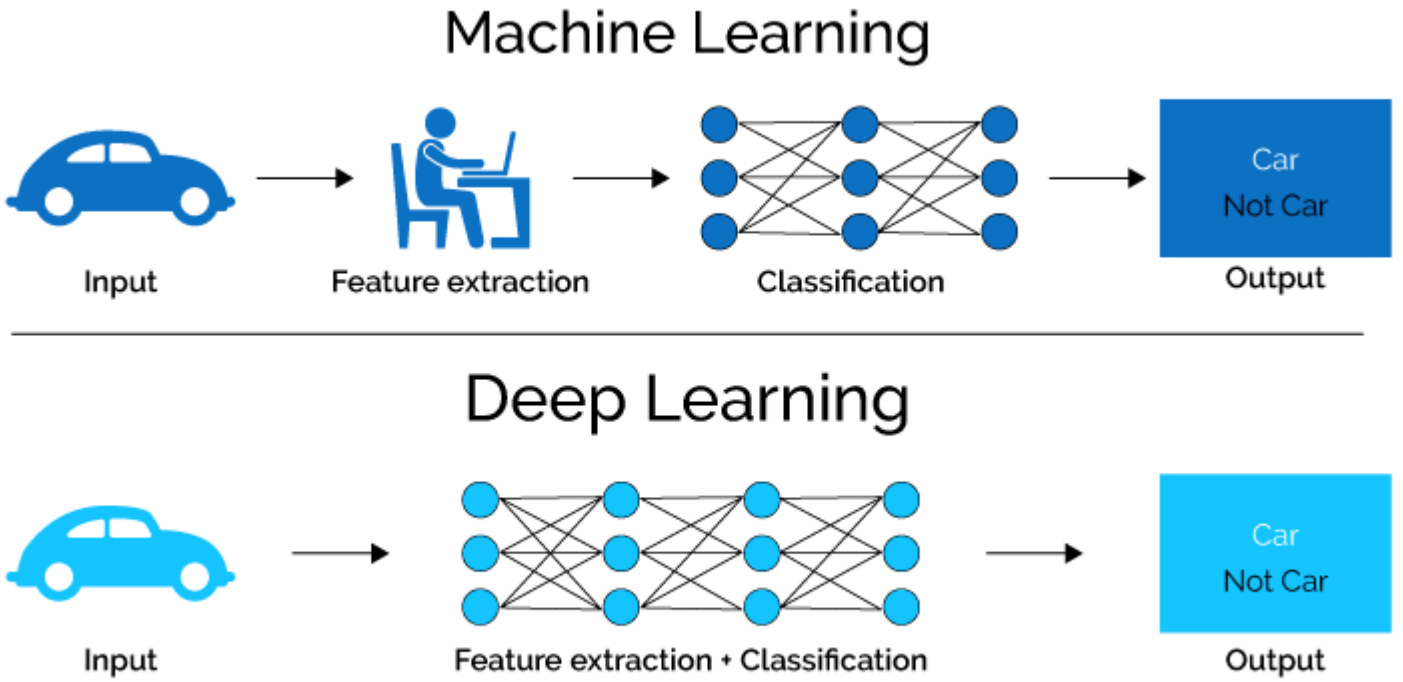
Task: Classify in dog, bird and human



1. Collect a high number of a wide range of “differing” images from the three categories
2. Annotate the images: each image is manually pre-classified
3. Learning starts with the first image of an “eagle”, and the system classifies in dog (30%), bird (50%) and human (20%) in the first round.
4. Training phase: the result is optimized (Feedback Signal) by learning from the variety of training images, resulting in dog (2%), bird (96%) and humen (2%)
5. Shortcut: the developer programs a hint (“pecker for an eagle”), to fasten the process.
6. Testing phase: Check if the system works for unknown images

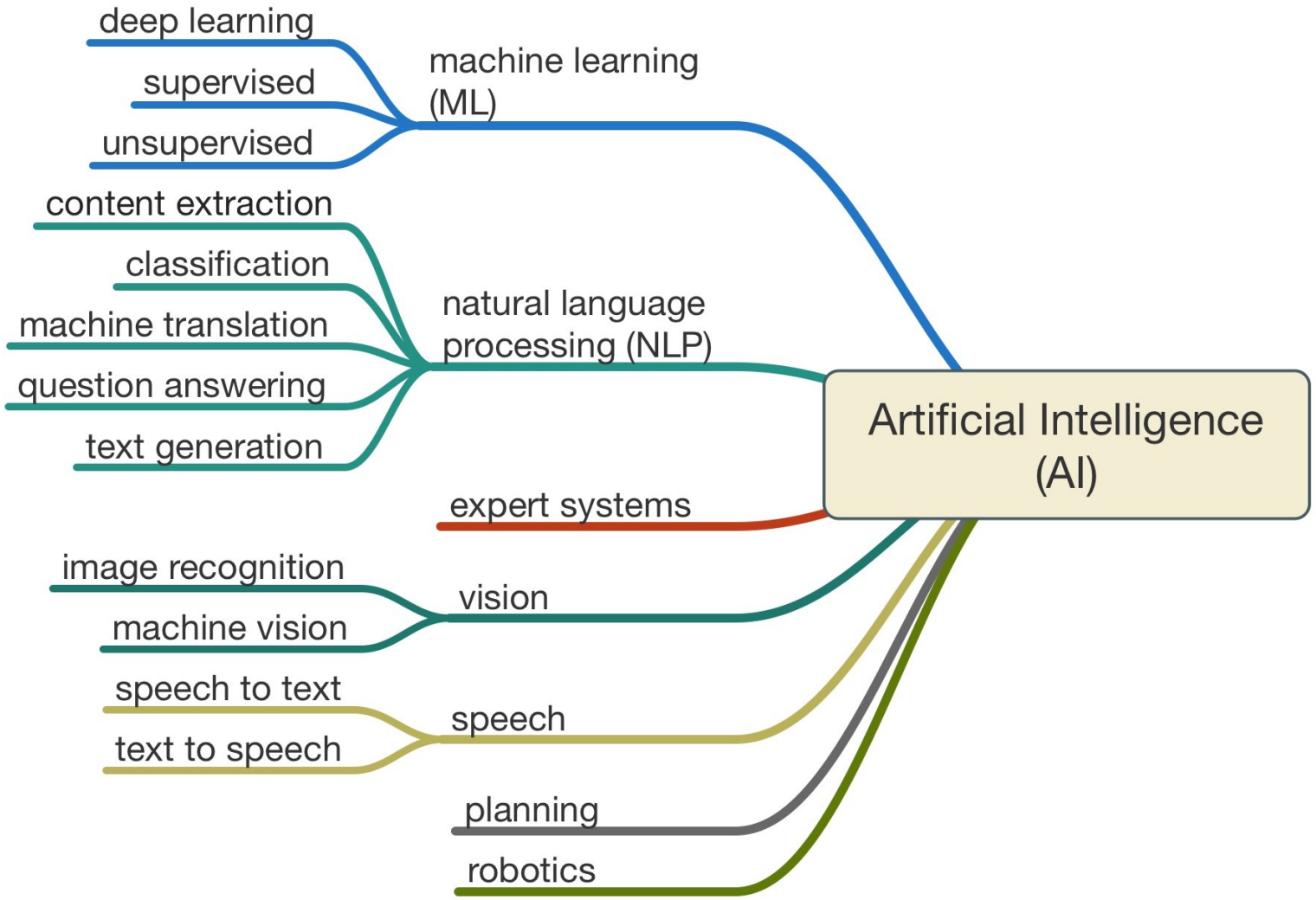
Machine Learning – More General

1. Getting computers to learn and act like humans do
2. Learn from data without relying on rule-based programming
3. Data is in the form of observations and real world interactions
4. Learning improves over time in an autonomous fashion



Source: XenonStack

AI


Artificial Intelligence (AI)
machine learning (ML)
deep learning
supervised
unsupervised
natural language processing (NLP)
content extraction
classification
machine translation
question answering
text generation
expert systems
vision
image recognition
machine vision
speech
speech to text
text to speech
planning
robotics

Ad hoc Learning Problems

1. No Background knowledge (context)
2. Conclusion from the past to future, frequent update of (training) data necessary
3. Catastrophic forgetting (lacks the ability to generalize)
4. Many images with cats and dogs, but no images with car accidents killing persons
5. Collecting images e.g. representing men and women in a stereotypic way, leads to machine learning results representing men and women in a stereotypic way

Where are chairs?

Research project: The Profiler

The Profiler - Objectives

- Research the usage of images of 10-15 years olds
  - Personal images online
  - Awareness for possibilities and risks: fairness and transparency …
- Develop a learning tool for face analysis
  - explore and evaluate own personal online profile
  - based on acquisition, analysis and linking of digital images
- Research, implementation and evaluation of state-of-the-art algorithms to detect, recognize and classify persons, age, gender and emotions in large amounts of image data taken in unconstrained environments

There is no privacy – deal with it.
(Roger Chesley, 2009)

There is no fairness and no transparency

(N.N., 2018)

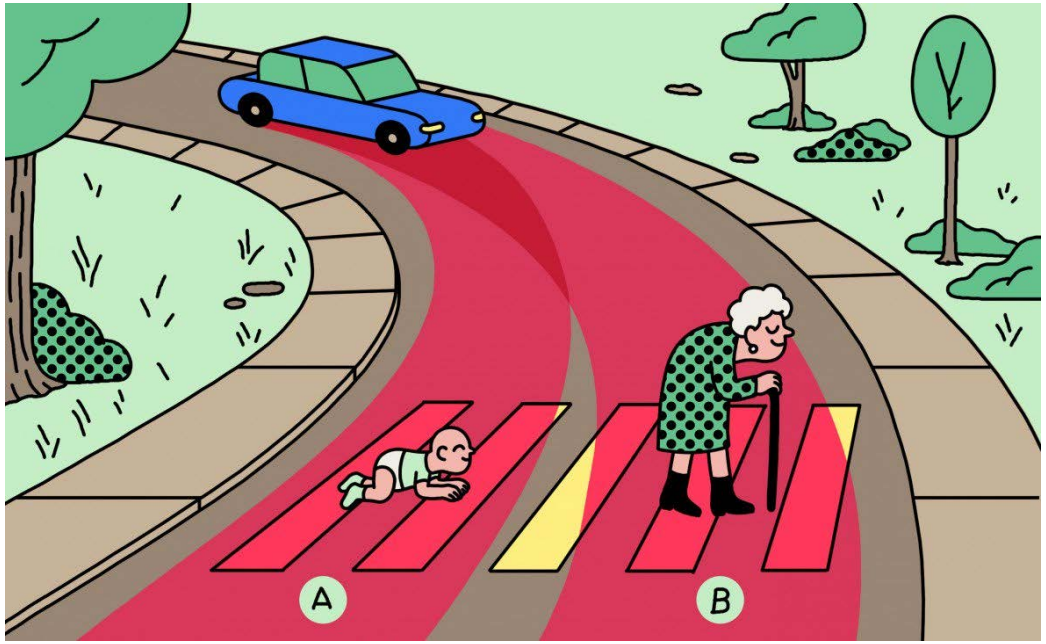
Profiler Tool

https://profiler.cvl.tuwien.ac.at/#/main/init

MIT Moral Machine

1. Crowd-sourced picture of human opinion on how machines should make decisions when faced with moral dilemmas
2. What should the self driving car do?
3. Referred to the trolley problem, a thought experiment in ethics.

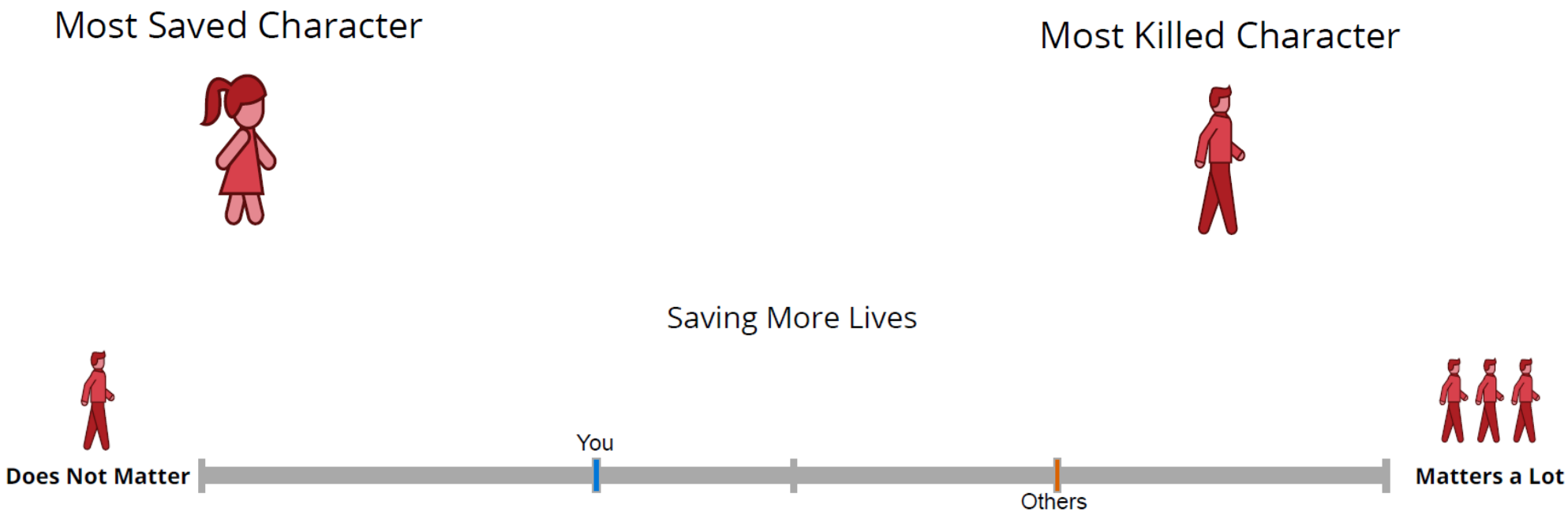
http://moralmachine.mit.edu/



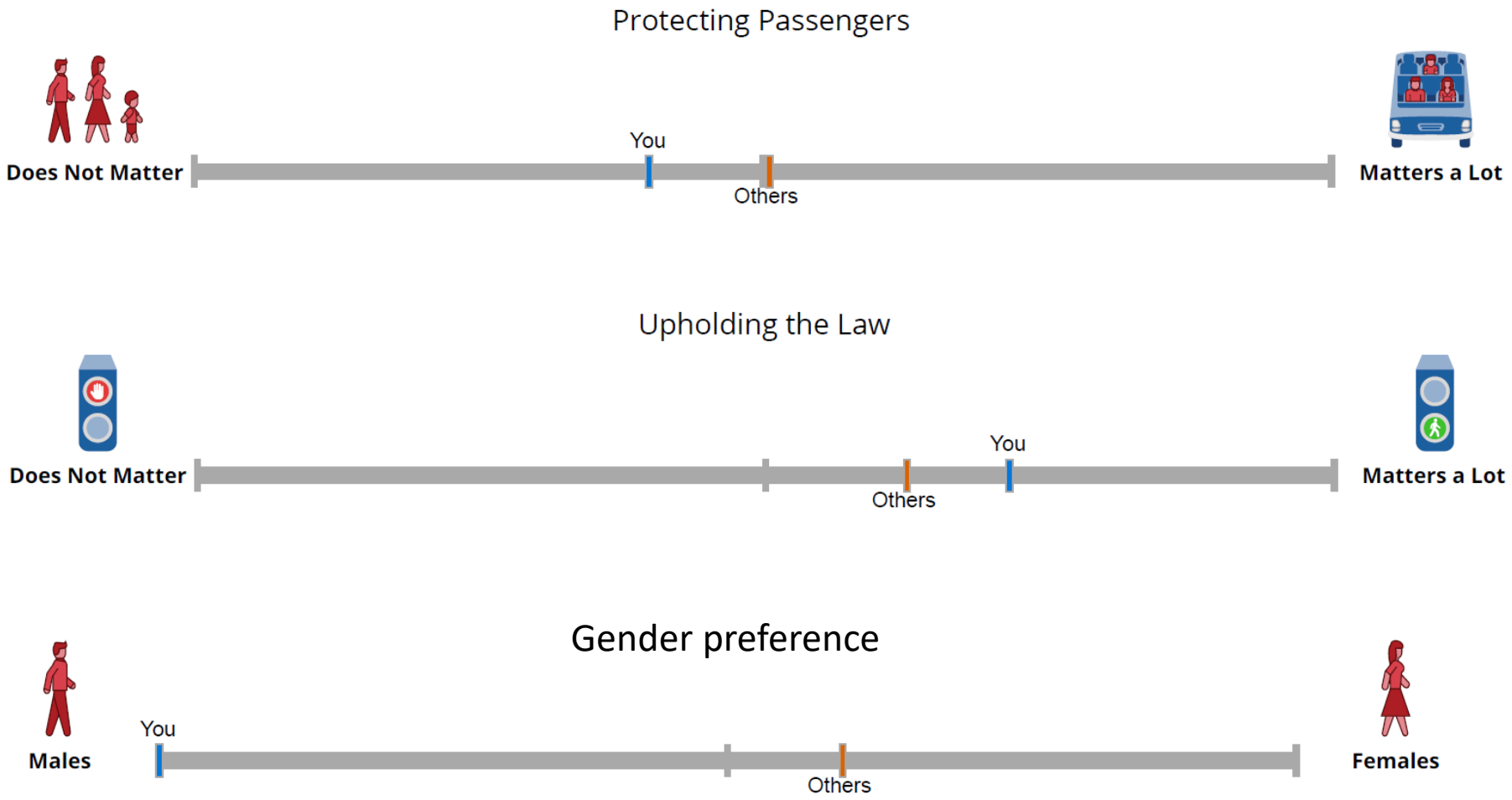
SIMON LANDREIN

MIT Moral Machine – Human perspectives on Ethics

http://moralmachine.mit.edu

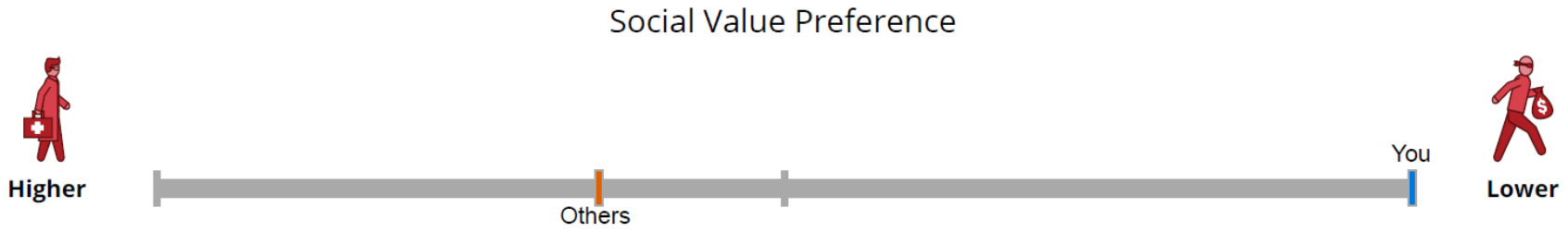


MIT Moral Machine



MIT Moral Machine

MIT moral machine



If autonomous vehicles cause less accidents then humans, then the technology could be superior

MIT moral machine

Questions from the developer:

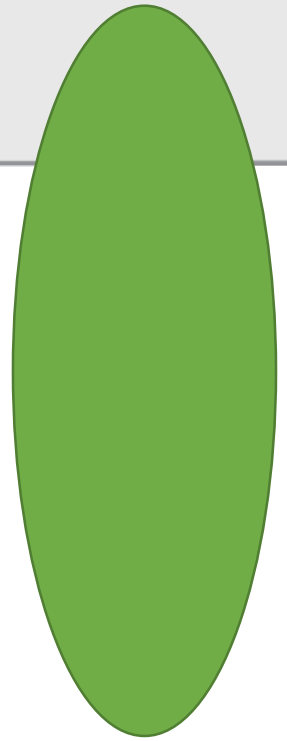
The Moral machine does not deal with ethics and moral in an appropriate way:

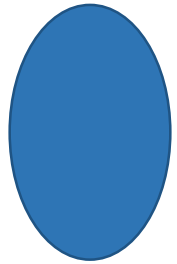
- Ethical aspects are represented in a simplified way.
- Prejudices are reinforced, categories of persons are not appropriate.

How should we program the car?

Agenda

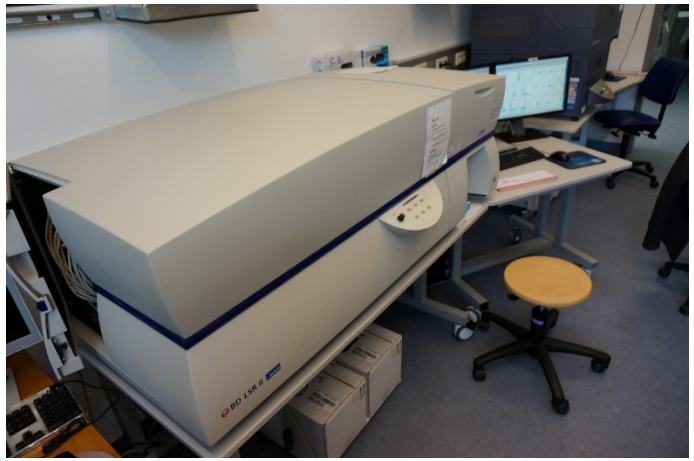
- Introduction and Motivation
- Examples
- MIT Moral Machine
- It s the Data
  - Behavior modelling (fall detection)
  - cancer research
  - Detection of suicidal activities
- Views from the developer





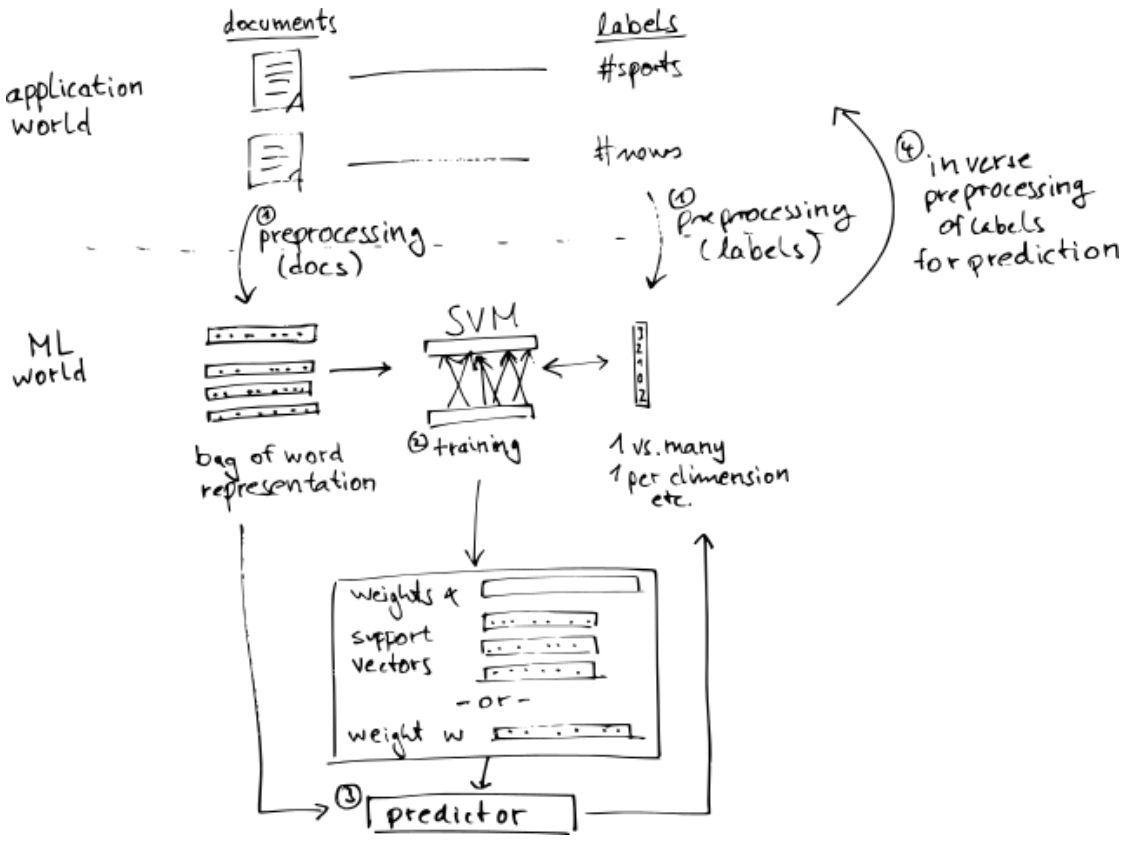
Agenda

- Data Acquisition for Active Assisted Living
  - Data Description
    - Action / Event / „obnormal“ behavior recognition
    - 3D
  - Data Modelling and Creation
  - Motion Capture
  - Training and Evaluation
- Data Acquistion in cancer research
  - Data Description
    - Leukaemia
    - Flowcytometry
  - Data Representation and Creation
  - Metadata and Biomarker Extraction
  - Data Analysis and Evaluation
- Data Acquisition for detecting pre- suicidal activities in prisons
  - Setup and data description



Motivation – Machine Learning Project design

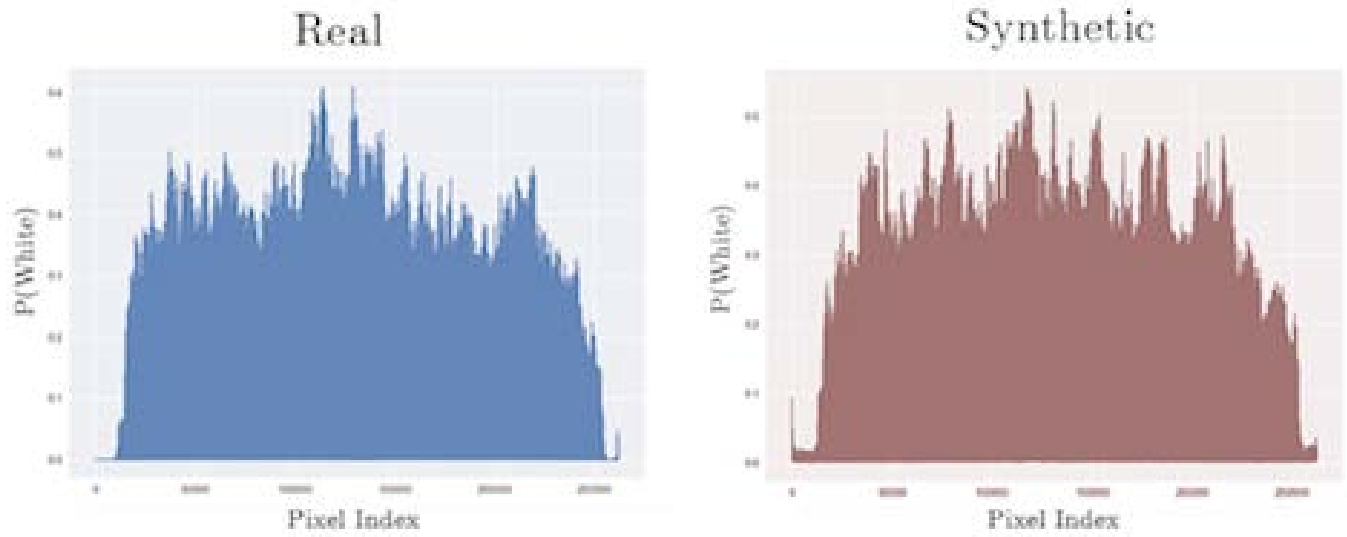
1. Learning about the discipline (5%)
2. Learning about the data (5%)
3. Data acquisition (40%)
4. Data analysis (30%)
5. Refinement and Evaluation (20%)



Data for training and evaluation

Real data: coming from sensors mounted in the Real world

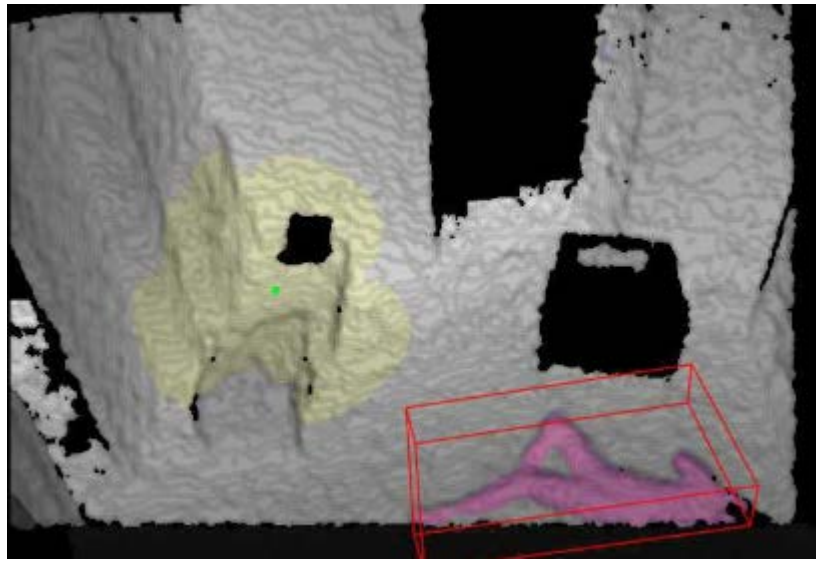
Synthetic data: data created and drawn with software

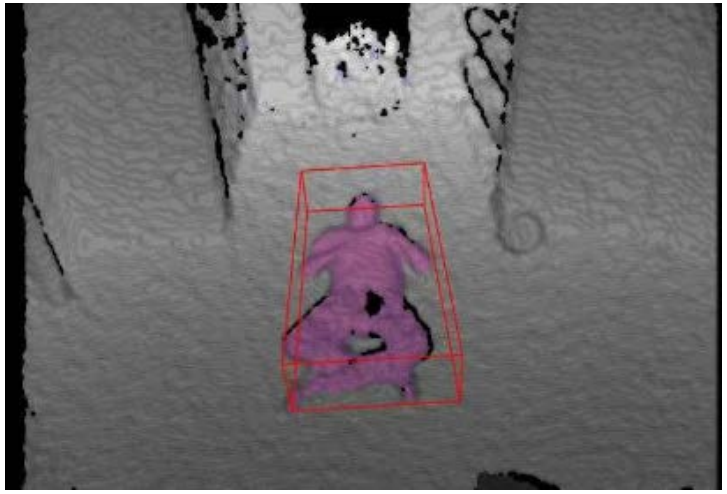


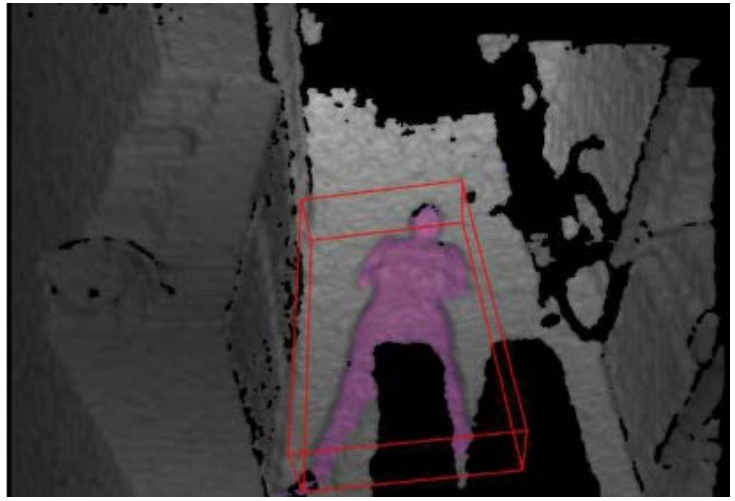
Data from the web: existing datasets for benchmarking

Acquisition device

Range Data







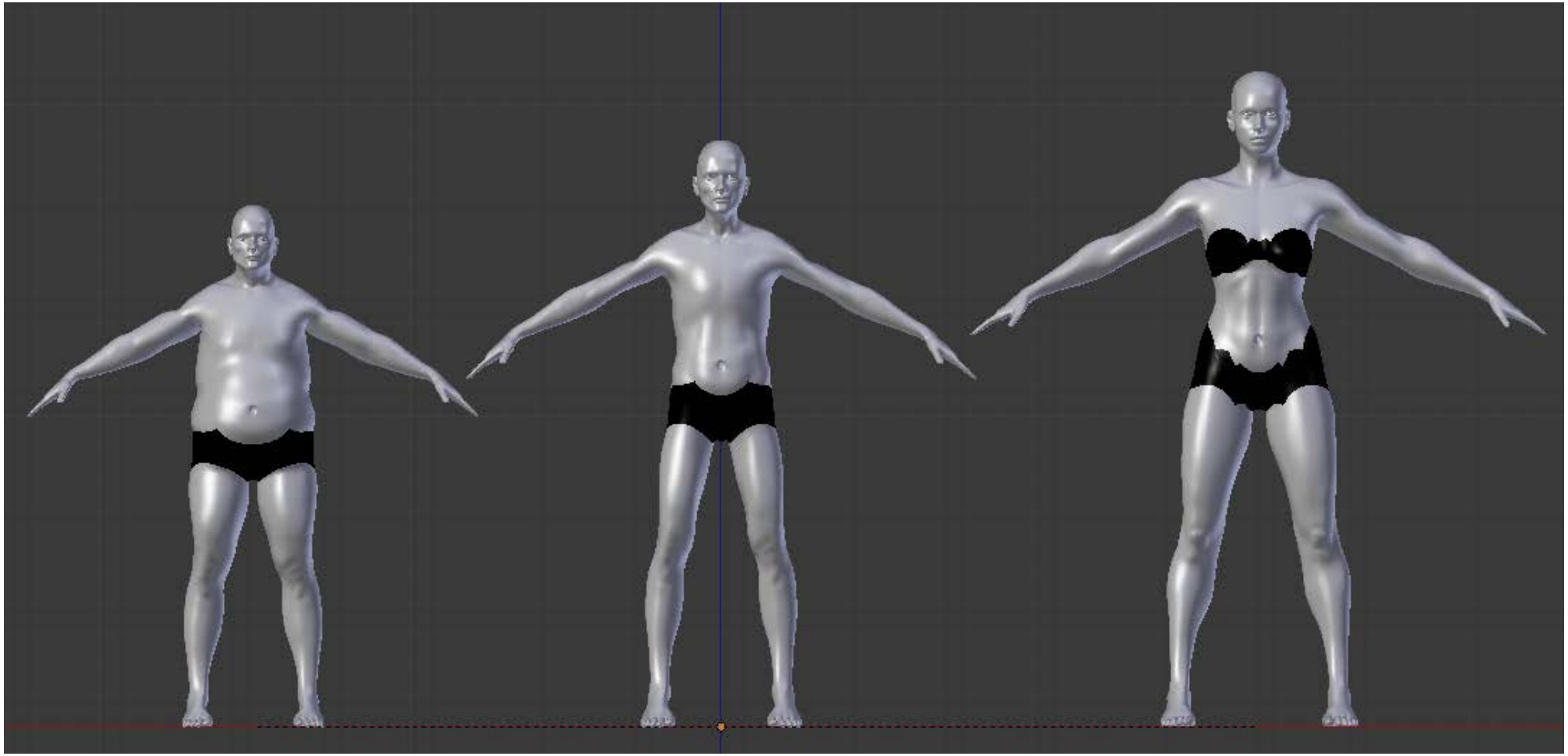
Data

1. Gesamte aufgenomme Zeit: 4800h (~200d)
   79kB pro Bild
2. entspricht 259.2 Millionen Frames
3. davon 1194 Sequenzen extrahiert und gelabeled
4. 118h oder 5.5 Millionen Frames
5. 18.962 Labels, 92 davon echte Stürze

Data acquisition

1. Motion based recording: most of the time, nothing is happening
2. -> 10s before and after event are bufferd and saved.
3. -> 30s after as additional data
4. Own file format defined: -> with compression ca 60MB per minute
5. Recording with external storage
6. Targeted persons: elder and frail persons
7. Data is recorded for evaluation and testing purposes: e.g. missing falls

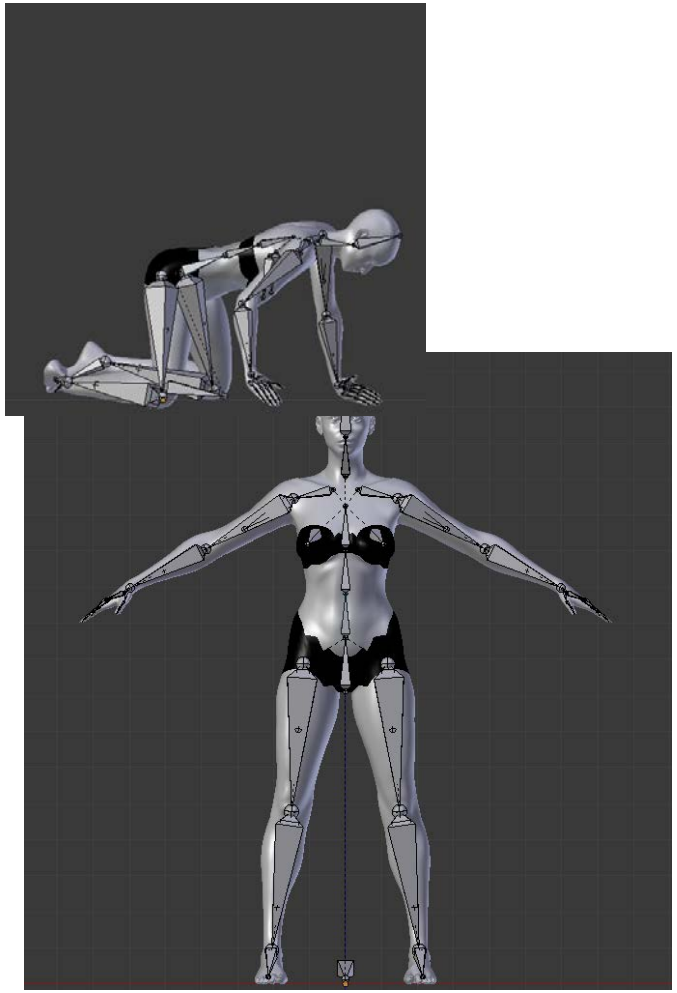
Synthetic data



Data Modelling

1. Modellierung mit Blender und dem Plugin "Manuel Bastioni Lab"-
2. Variieren Größe, Gewicht, Muskulatur, Geschlecht- etc.
3. keine Kleidung / keine Haare, da zu komplex und Daten soweit runterskaliert, dass es fast keinen Unterschied macht-
4. System kann für konkrete Zielgruppe trainiert werden (ältere Personen zw 1.5m und 2.0m) - zB auch Kinder möglich

Motion Capture Suits



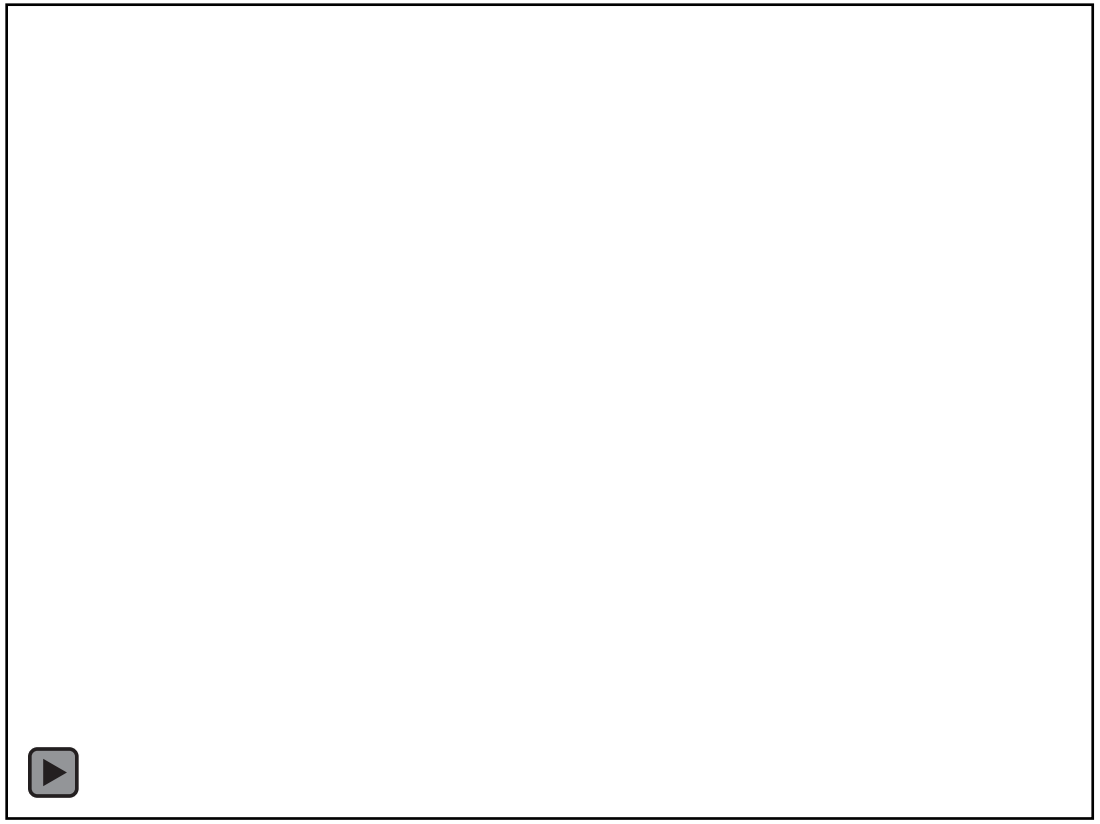
Motion Capture

1. Aufnahme mit dem Rokoko Smart Suit- 19
2. Sensoren mit jeweils 3 Komponenten (gyrometer, accelerometer, magnetometer)
3. Aufgenommene Daten werden auf die Avatare übertragen (Trainingsdaten)
4. Sehr genau, jedoch nicht "hollywood reif". Temporäre Fehler sind allerdings nicht schlimm für unsere Zwecke, da wir frame-basiert trainieren.

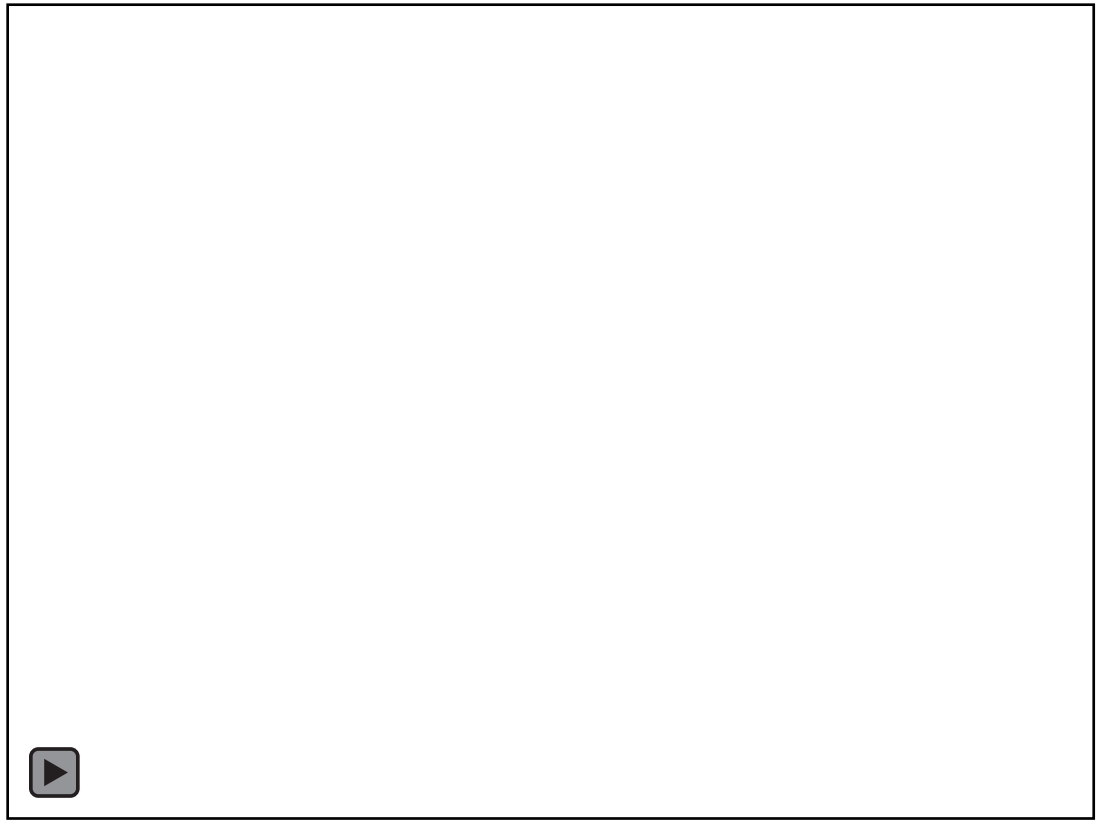
Developed with real life data

- 4,800 hours monitoring of elderly recorded
- > 5 million frames annotated (e.g. sitting, standing, lying, fallen)
- 92 real falls
- Automatic evaluation of algorithms on test database
- Robust detection of various falls
  - adjustable notification threshold per client
  - >95% of falls are detected

Variety of falls

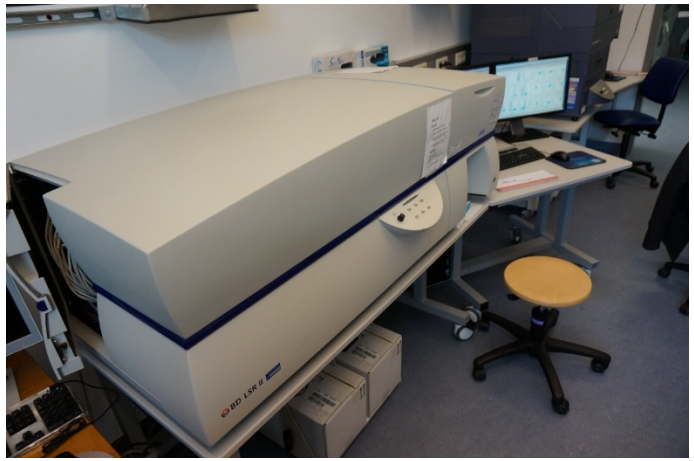


Variety of falls



Agenda

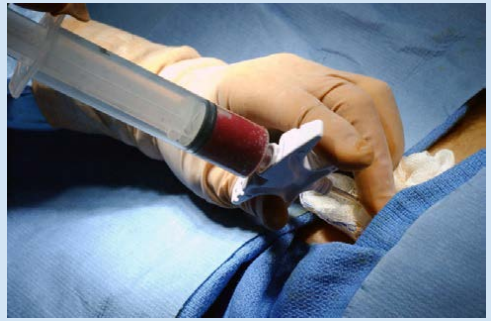
- Data Acquisition for Active Assisted Living
  - Data Description
    - Action / Event / „obnormal“ behavior recognition
    - 3D
  - Data Modelling and Creation
  - Motion Capture
  - Training and Evaluation
- Data Acquistion in cancer research
  - Data Description
    - Leukaemia
    - Flowcytometry
  - Data Representation and Creation
  - Metadata and Biomarker Extraction
  - Data Analysis and Evaluation
- Data Acquisition for detecting pre- suicidal activities in prisons
  - Setup and data description



Data Flow in Cancer Research

Study Design

- Definition
- Research Question
- Definition study cohort size
- Definition of SOP
- Treatment Protocol
- Antibodypanel
- Ethics
- ….



Data Acquisition

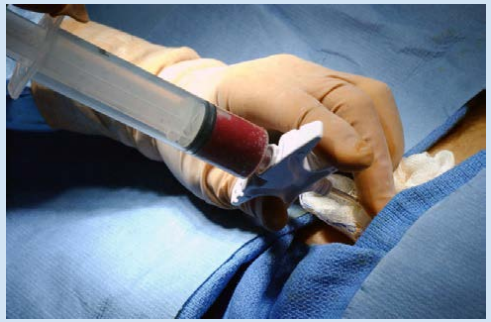
- Patient recruitment
- Informed consent
- Data acquisition within the hospital
- Analysis and diagnosis within the hospital or externerly

Data-Clearing

- Anonymization /Pseudonymization
- Quality check
- Ethics check
- Distribution to research Institutions by agreement

Data Acquisition in Leukaemia Research

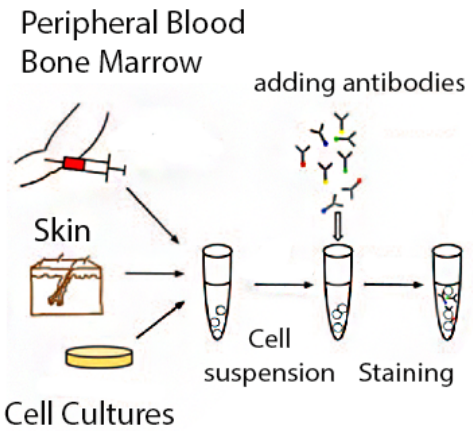
St. Anna Kinderkrebsforschung
CHILDREN'S CANCER RESEARCH INSTITUTE

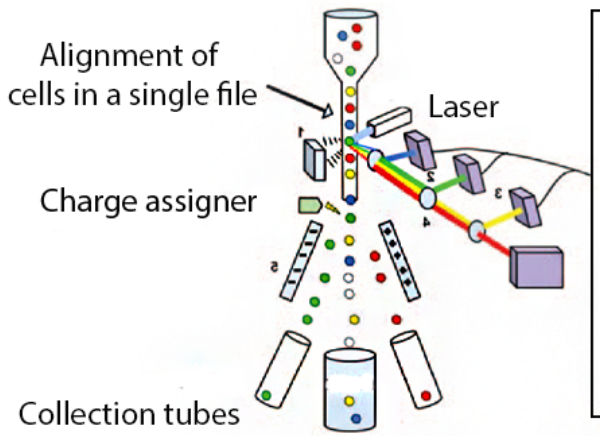


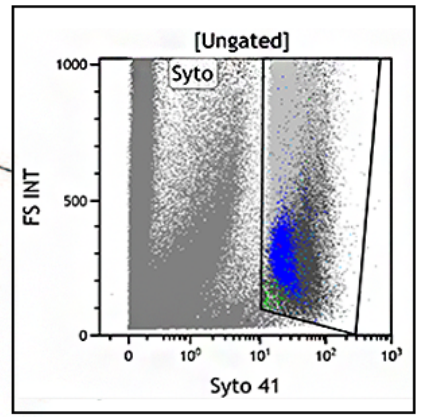
labdia
labordiagnostik
St. Anna Forschungsinstitut

MEDIZINISCHE UNIVERSITÄT WIEN

https://www.meduniwien.ac.at/web/ueber-uns/organisation/gremien/daten-clearingstelle/





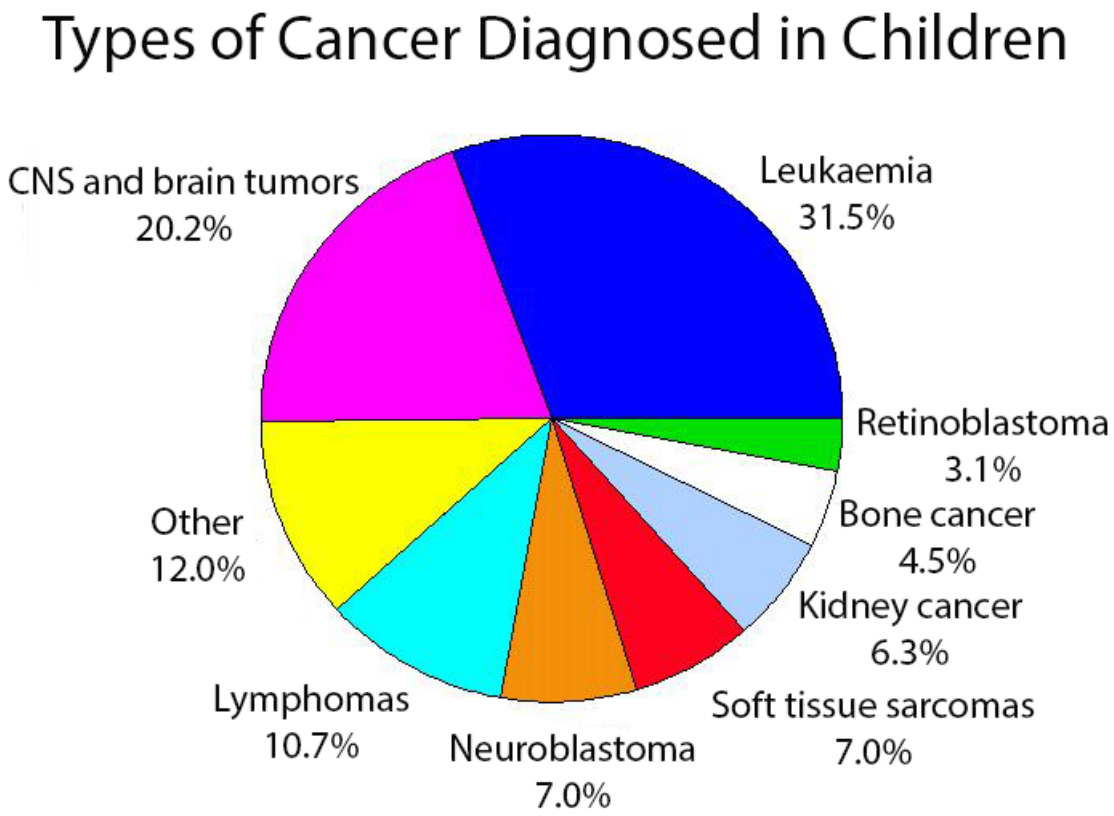


TU WIEN
TECHNISCHE UNIVERSITÄT WIEN
Vienna University of Technology

Flowcytometry

CVL Computer Vision Lab TU WIEN

Data Description

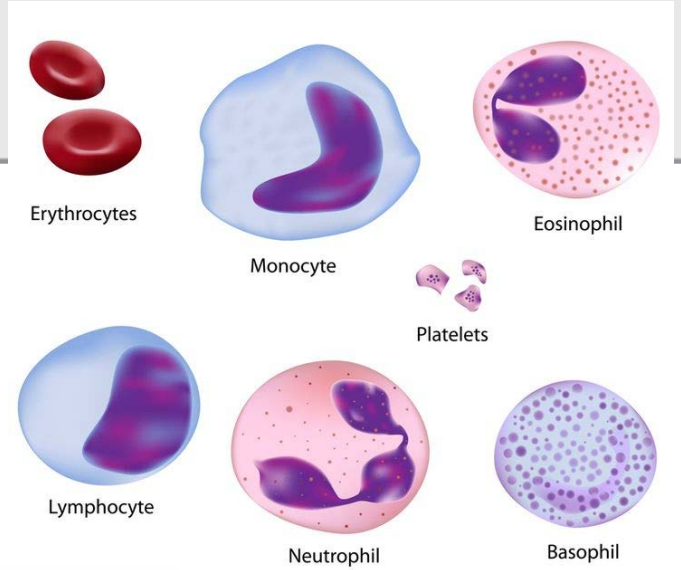


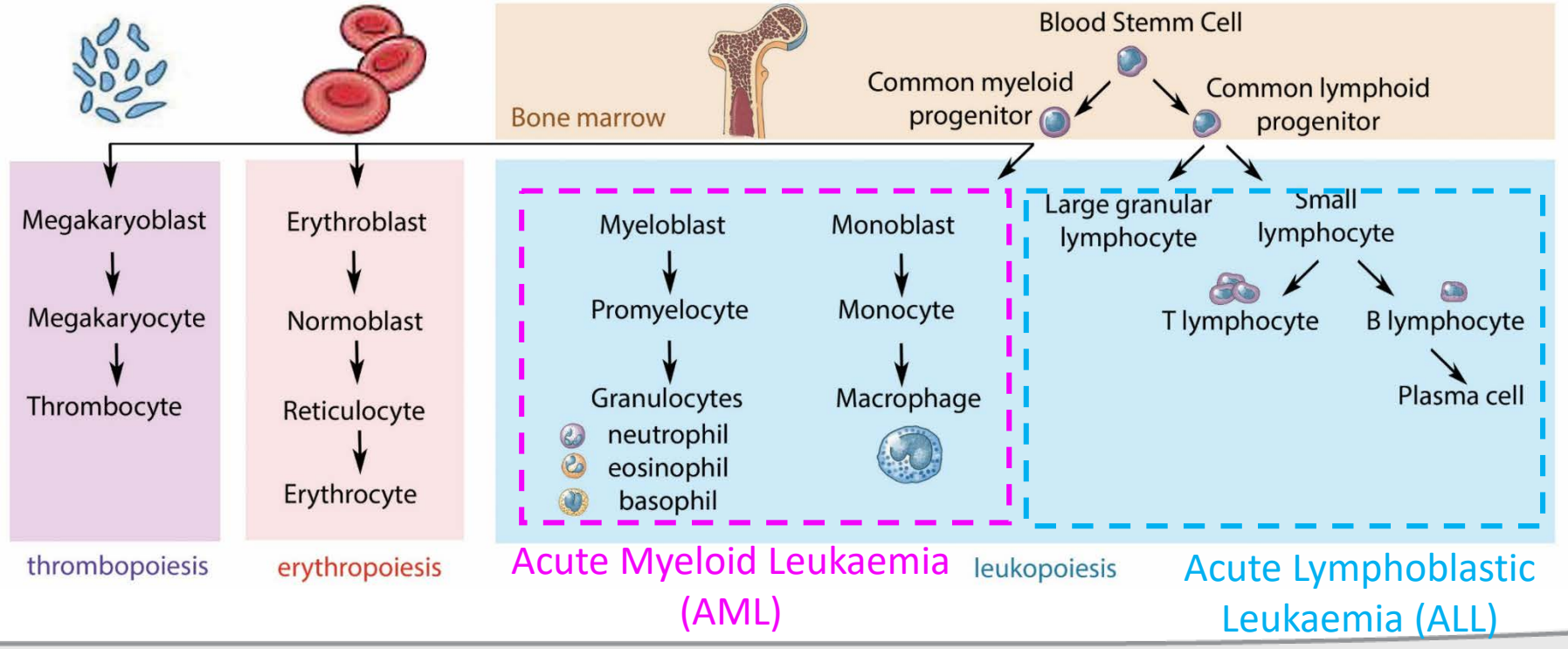
Leukaemia

- Acute Lymphoblastic Leukaemia (ALL)
  - 80-85%
  - Incidence 3,3 / 100 000 (< 15 years)
  - Peak: 2-5 years
- Acute Myeloid Leukaemia (AML)
  - 15-20%
  - Incidence 1,84 / 100 000 (< 15 years)
  - Peak: 0-2 years and >13 years

Data Description

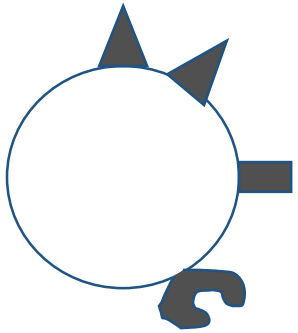
1. Measuring blood cells in leukaemia
2. Minimal Residual Diseases for therapy guidance





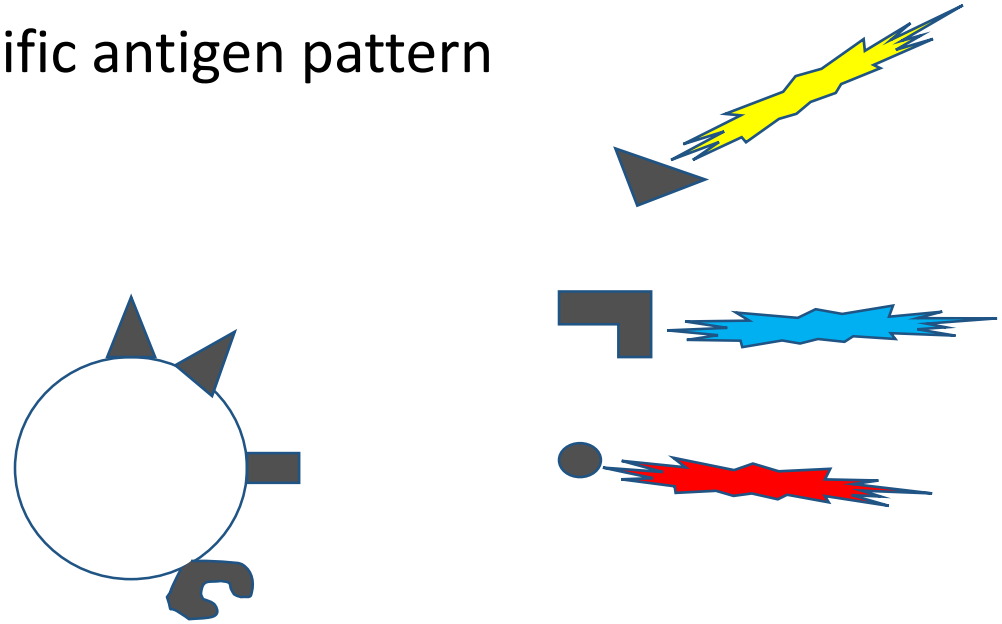
Data Acquisition Flow Cytometry

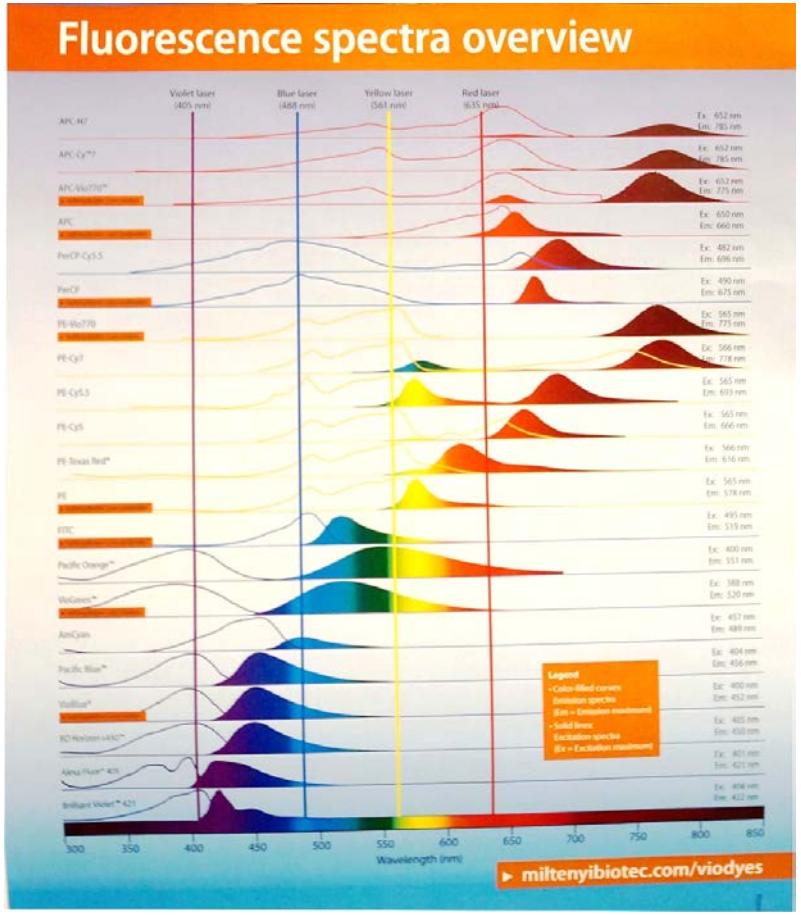
Cell-specific antigen pattern



Data Acquisition Flow Cytometry

Cell-specific antigen pattern

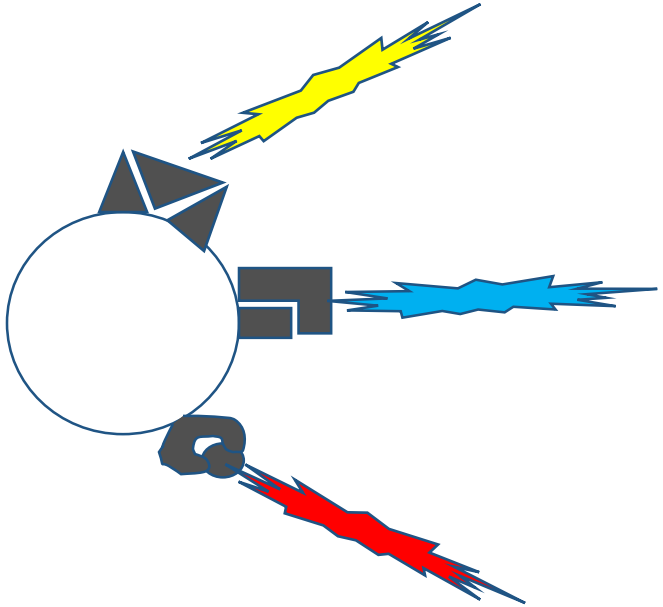




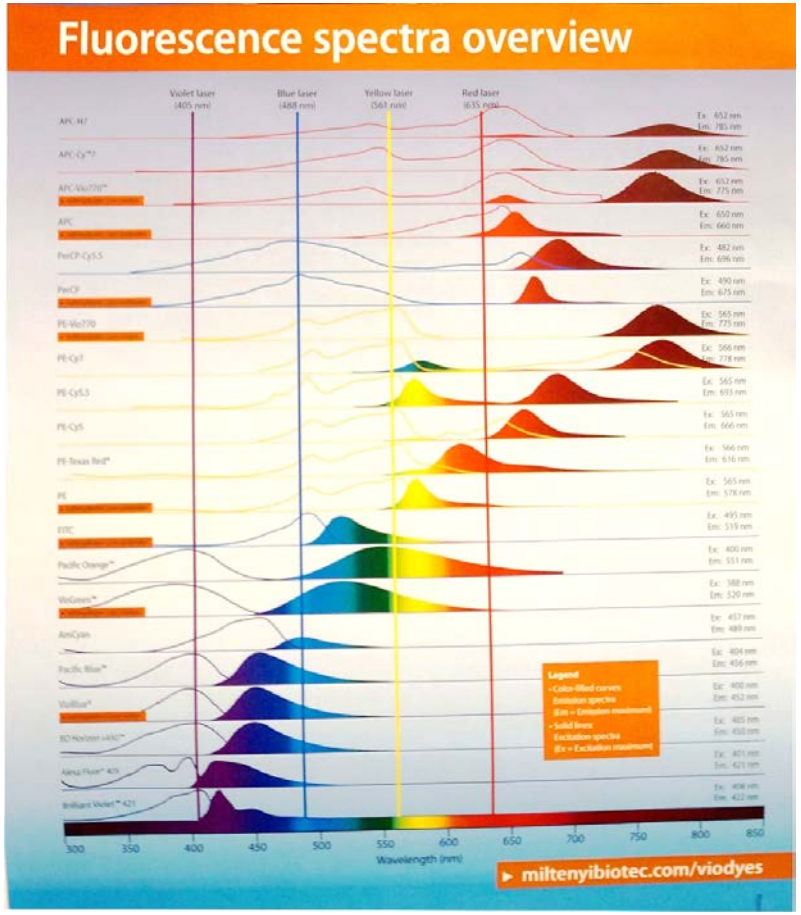
Staining using antibodies
marked with fluorochromes

Data Acquisition Flow Cytometry

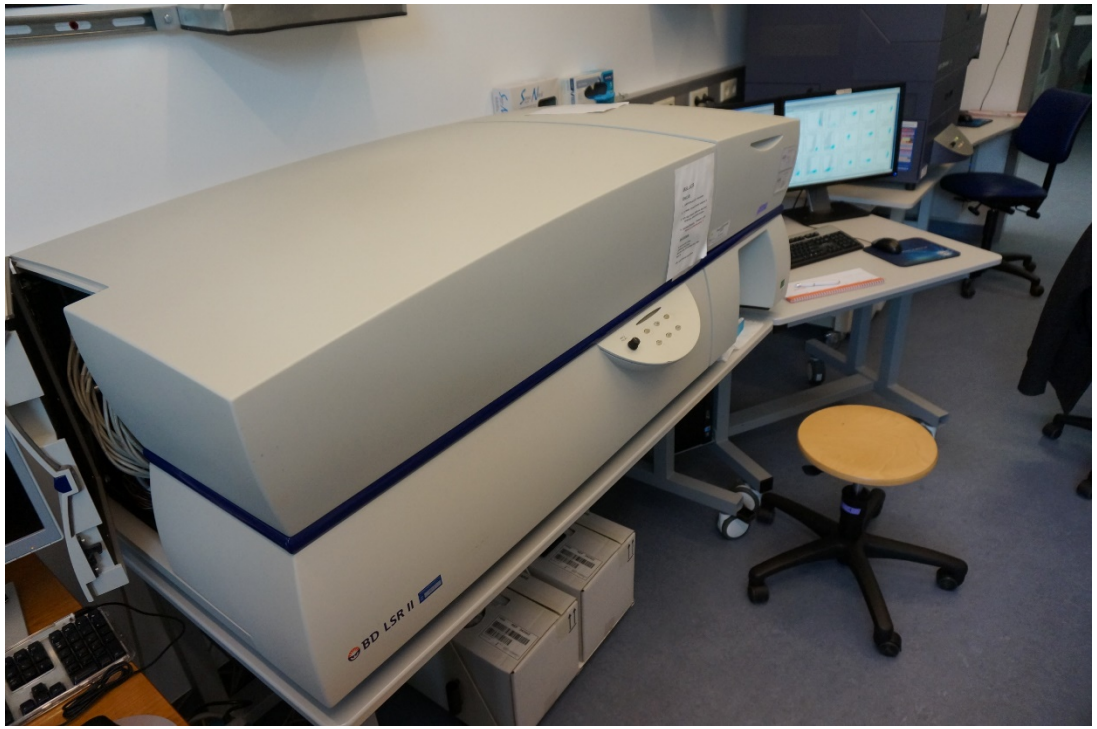
Cell-specific antigen pattern

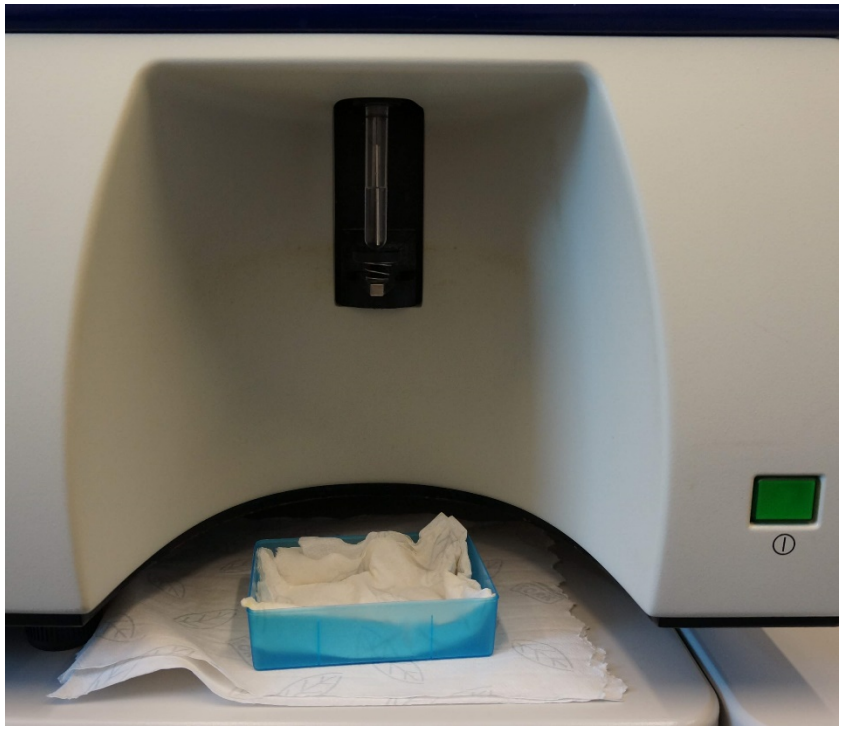


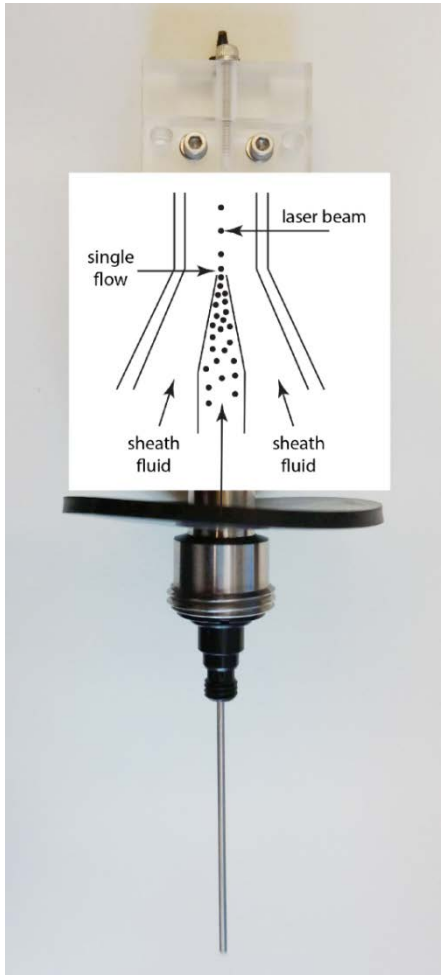
Staining using antibodies
marked with fluorochromes



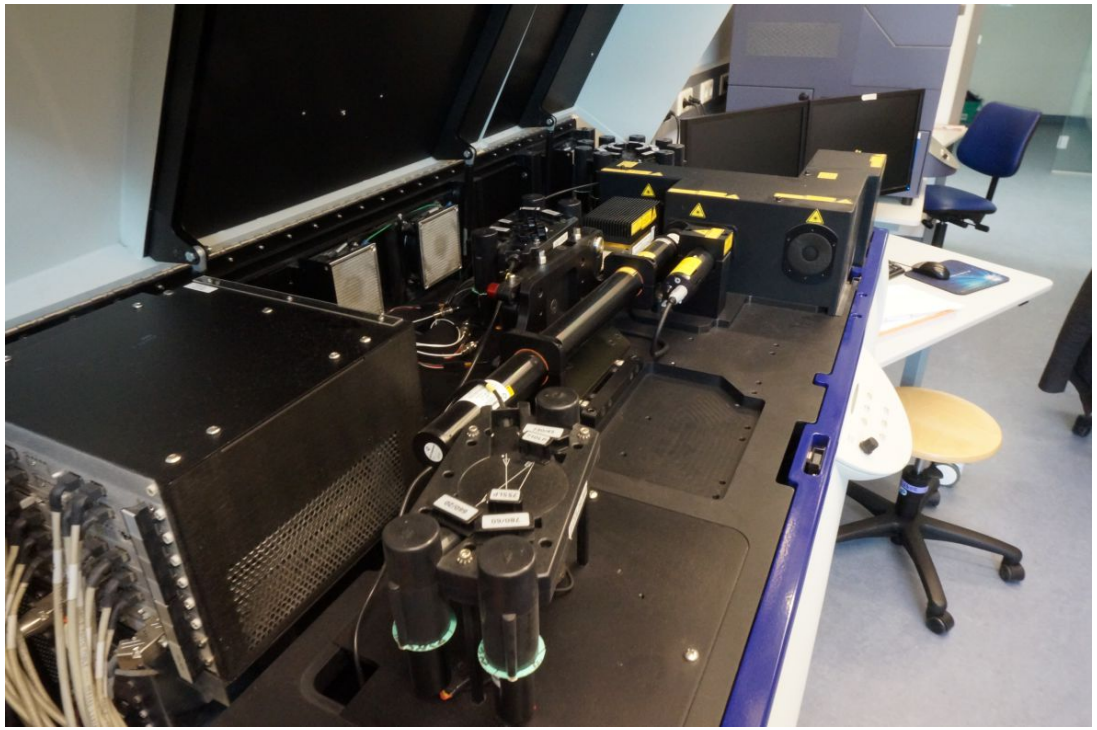
Data Acquisition Flow Cytometry

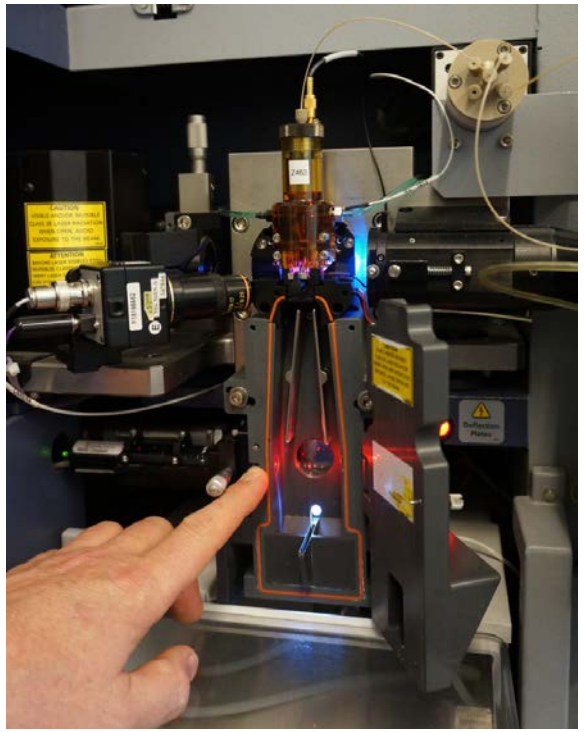


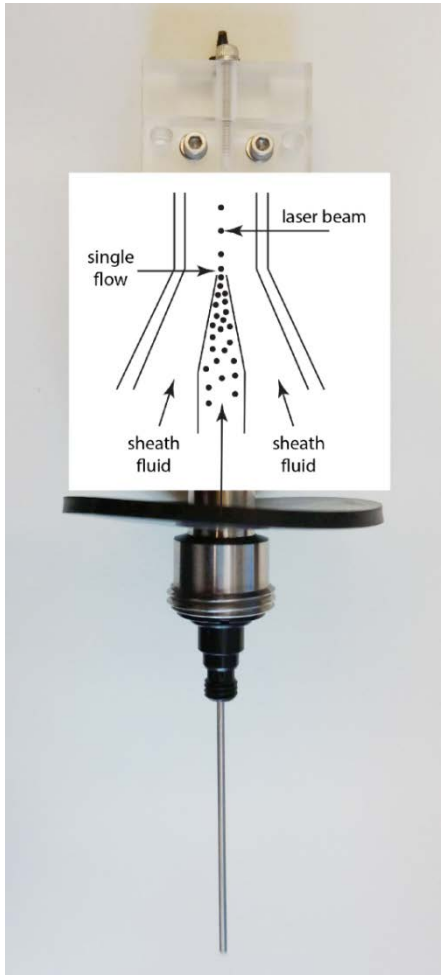




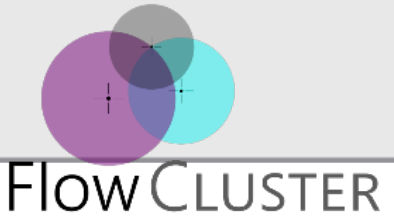
Data Acquisition Flow Cytometry







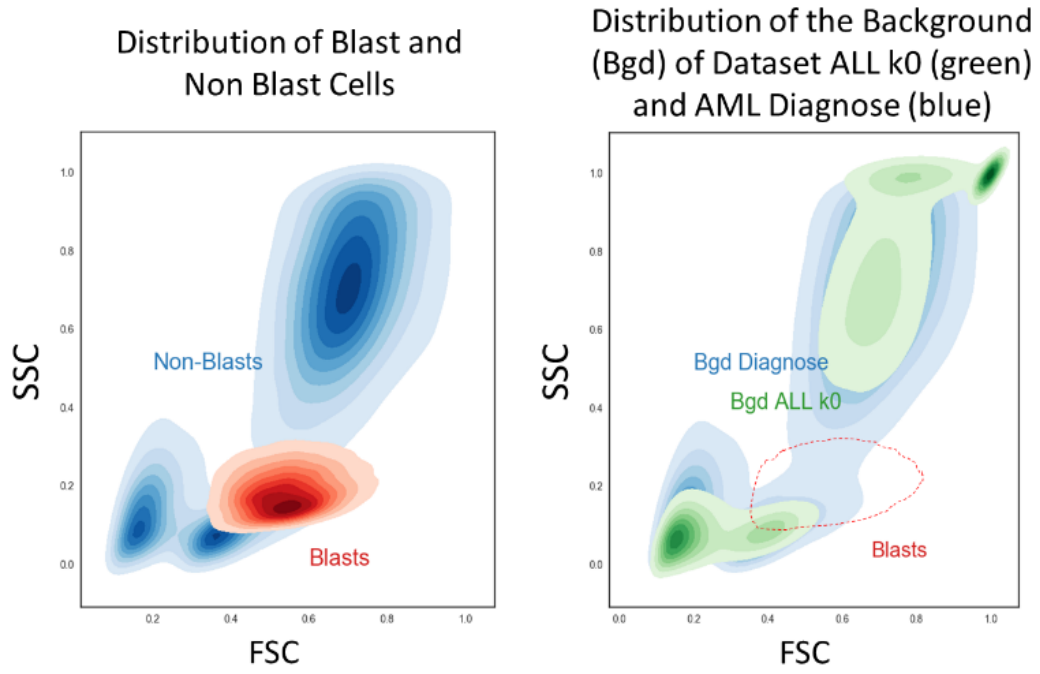
Dataset Example AML



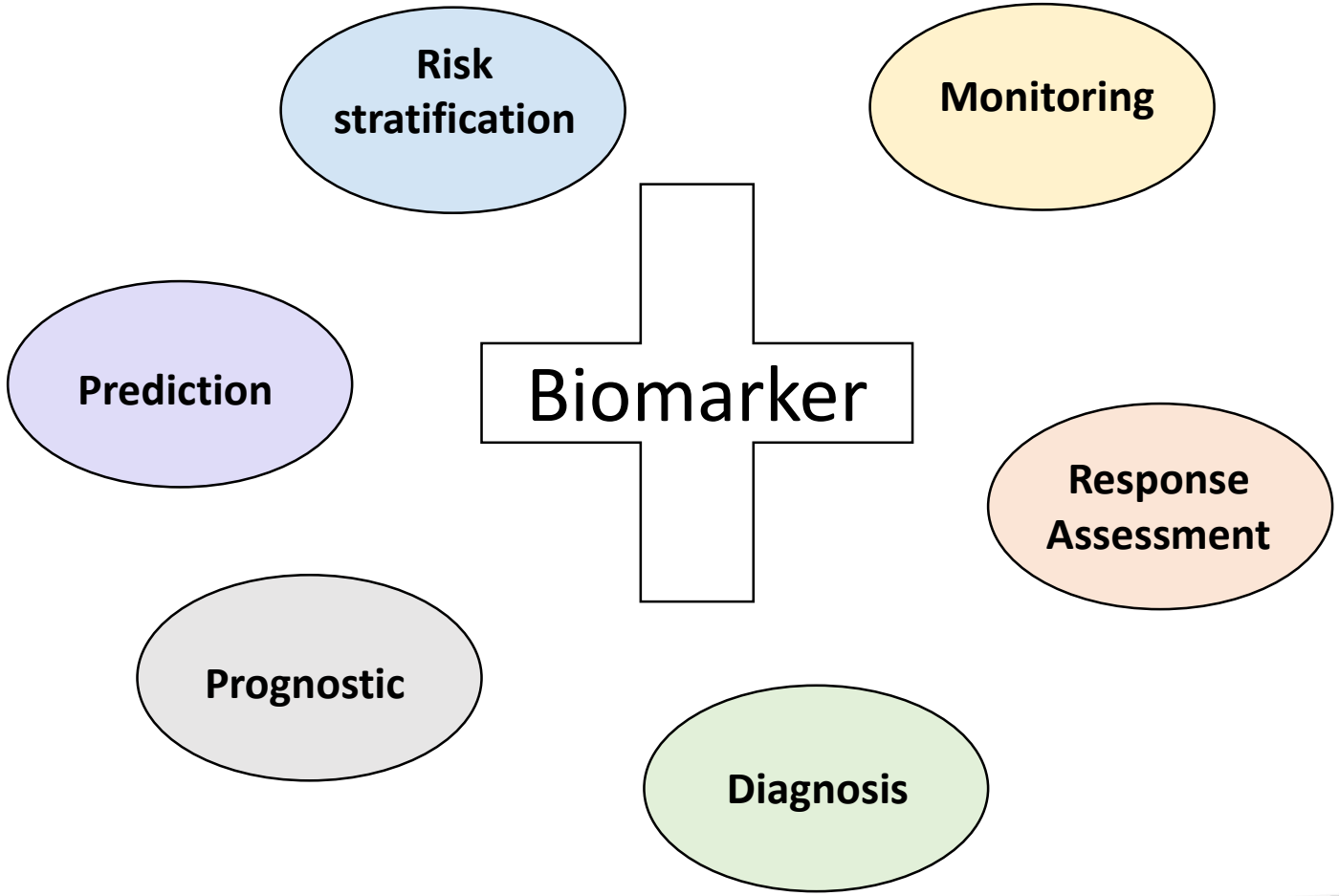
- Dataset AML Diagnose: 13 (Blasts, Background)
- Dataset ALLk0: 30
- 9 Features observed per cell
  - 2 Physical features: FS INT, SS INT
  - 7 Antibodies: CD38, CD34, CD117, CD33, CD123, CD45RA, CD45
- $3*10^5 – 10^6$ cells observed per subject
- Treatment protocols
  - AML-BFM 2004
  - AIEOP-BFM 2009

Data Representation

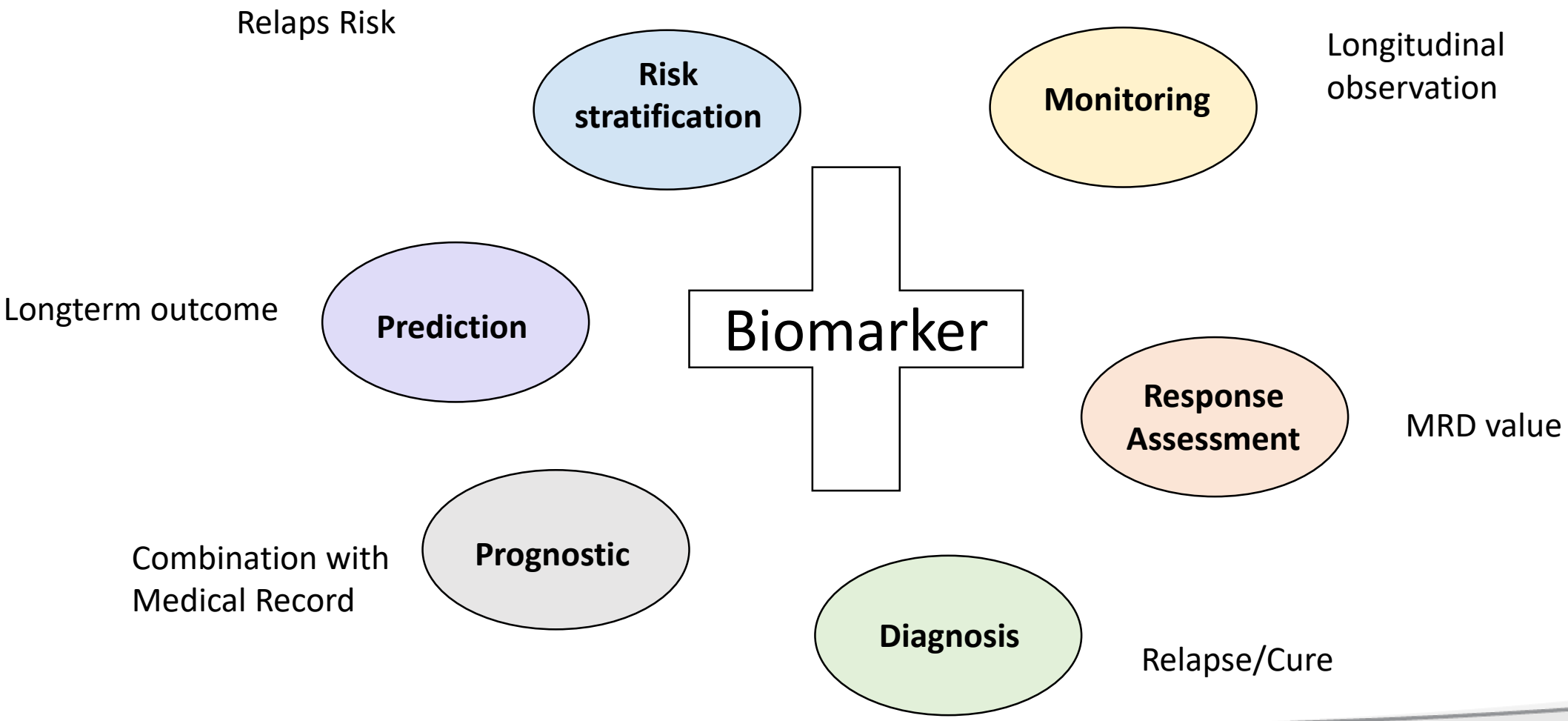
1. Modelling of Distributions
2. Lower Dimensional Embedding Spaces
3. Forming of Synthetic Dataclouds



Metadata- and Biomarker Extraction

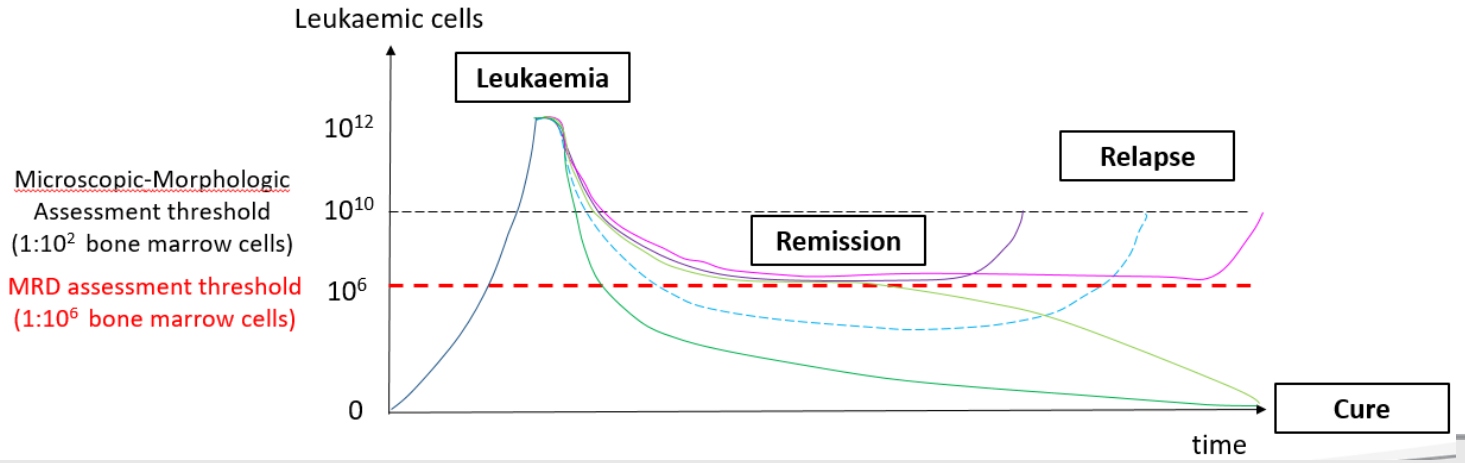


Metadata- and Biomarker Extraction



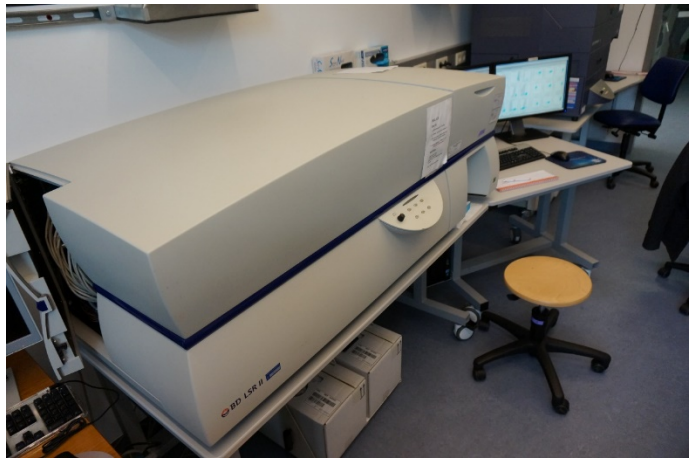
Data Analysis and Evaluation

- Training, Validation and Testdata Design
  - Crossvalidation
- Statistical Analysis
- MRD Value Computation
- Extraction and Analysis of Temporal Trajectories



Agenda

- Data Acquisition for Active Assisted Living
  - Data Description
    - Action / Event / „obnormal“ behavior recognition
    - 3D
  - Data Modelling and Creation
  - Motion Capture
  - Training and Evaluation
- Data Acquistion in cancer research
  - Data Description
    - Leukaemia
    - Flowcytometry
  - Data Representation and Creation
  - Metadata and Biomarker Extraction
  - Data Analysis and Evaluation
- Data Acquisition for detectiong pre- suicidal activities in prisons
  - Setup and data description



Aufnahmesystem

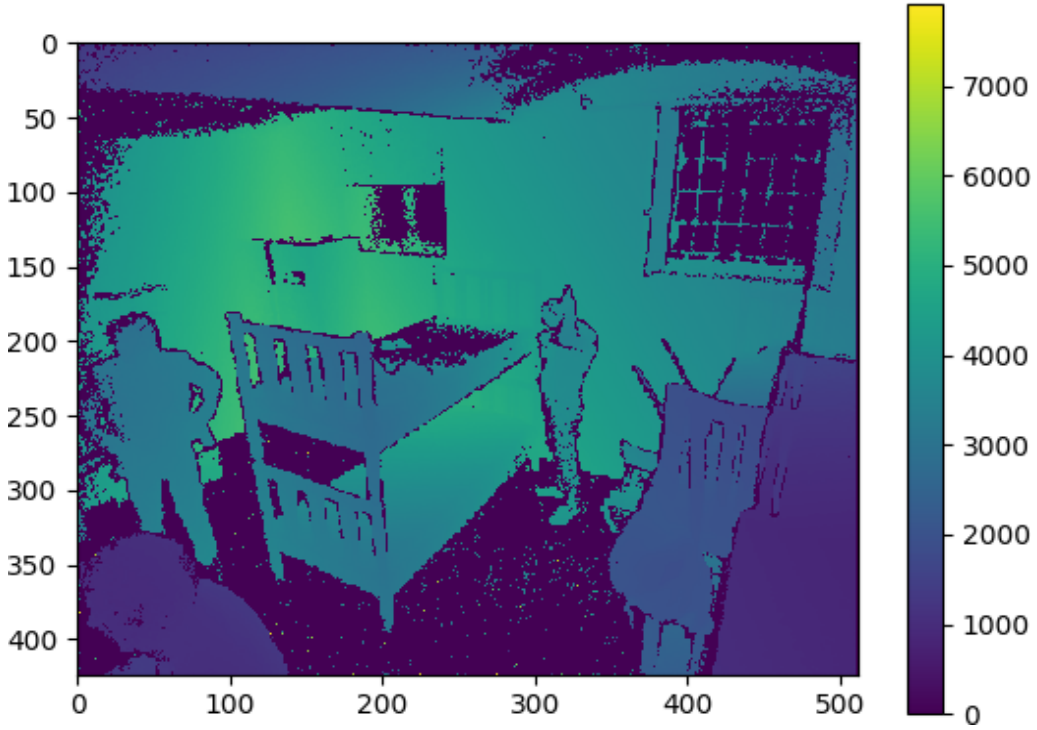
Sensoren

- 2 Kinect Tiefensensoren (Zelle und Nassraum)

Aufnahmesystem

Sensoren

- Distanzmessung in mm (hier farblich dargestellt)



detect-sitting.mp4

Aufnahmesystem

Sensorgehäuse

Version 1 – bruchsicheres Glas, konfigurierbare Sensorneigung

Aufnahmesystem

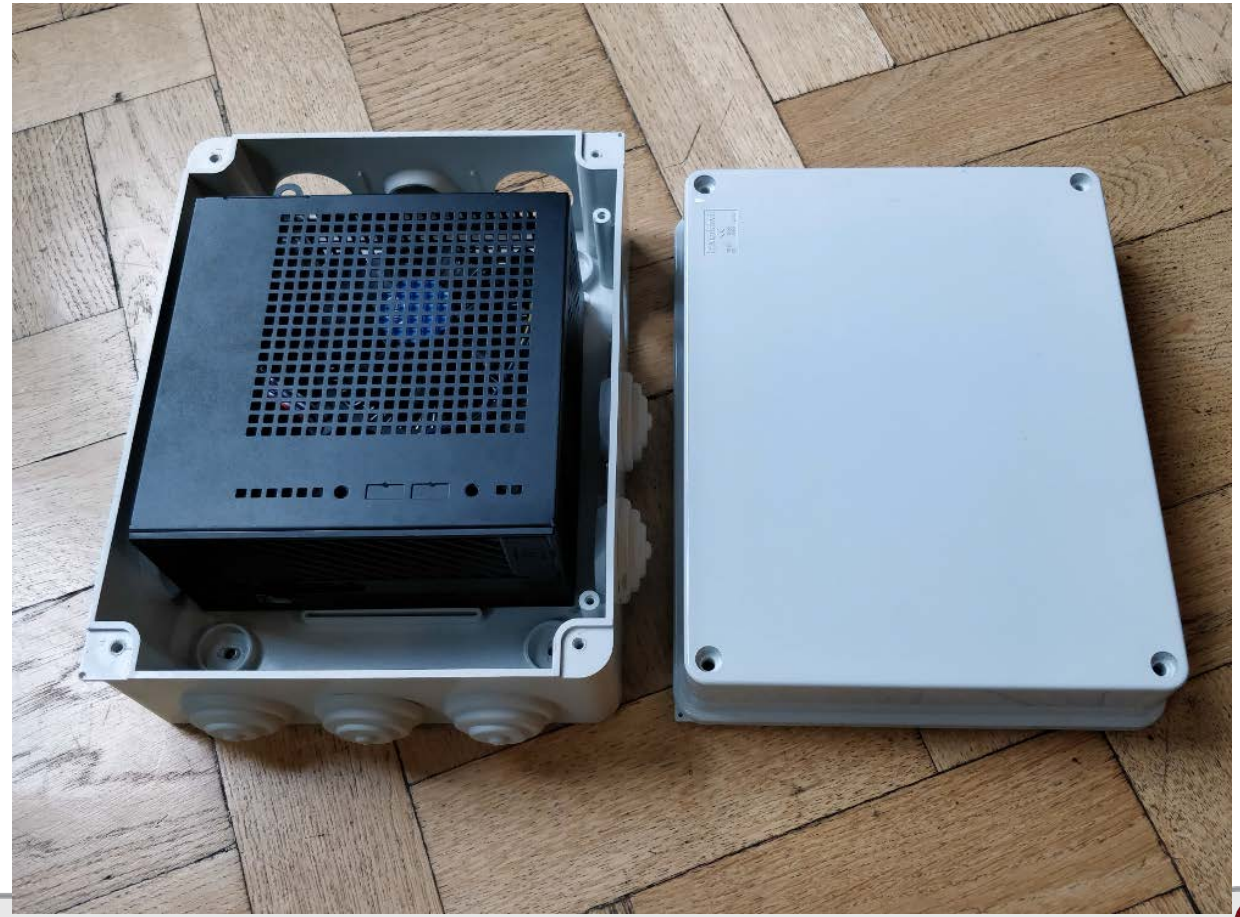
Sensorgehäuse

Version 2 – Metall, fixe Sensorneigung

Aufnahmesystem

Computer

2 kompakte PCs geschützt durch Plastikgehäuse



Aufnahmesystem

PCs

Montage im Gang (Schutz vor Zugriff)

Aufnahmen

Statistiken

15 Bilder pro Sekunde, Auflösung 512x424 Pixel

Zelle

- Mehr als 2300 Stunden (97 Tage) aufgenommen
- < 1% der Daten unbrauchbar (Sensor verdeckt)

Nassraum

- Mehr als 1400 Stunden (60 Tage) aufgenommen
- < 2% der Daten unbauchbar (Sensor verdreht)

Aufnahmen

Analyse

Manuelle Analyse der Aufnahmen aus der Zelle

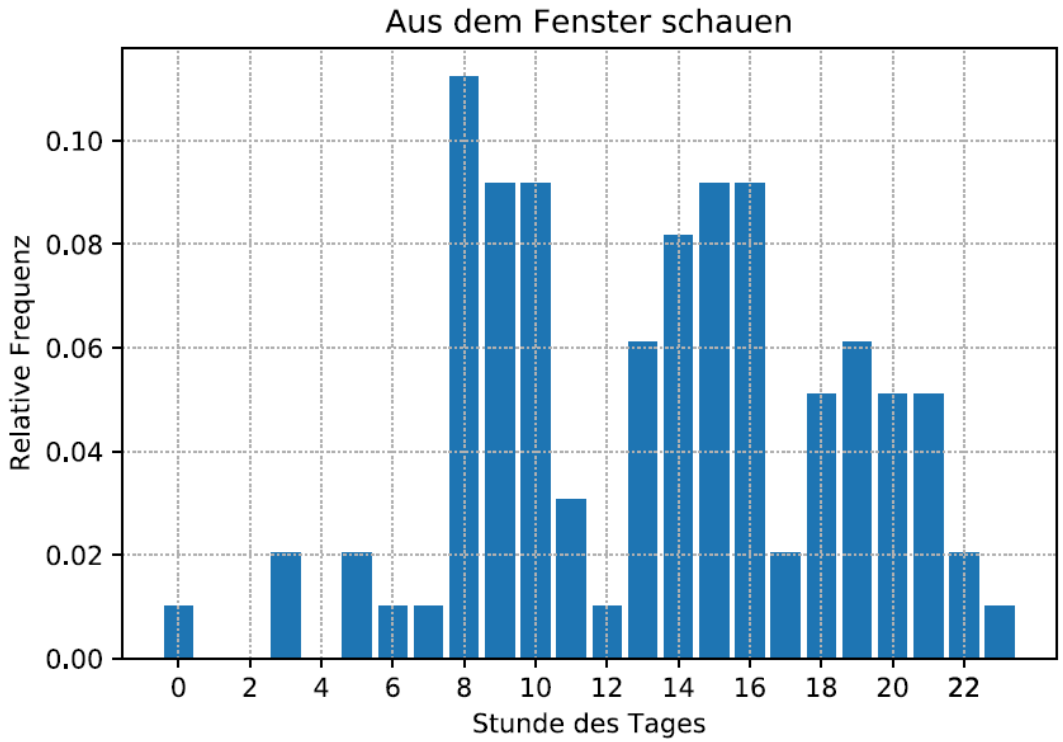
Anzahl Ereignisse pro Tag (Mittelwerte)

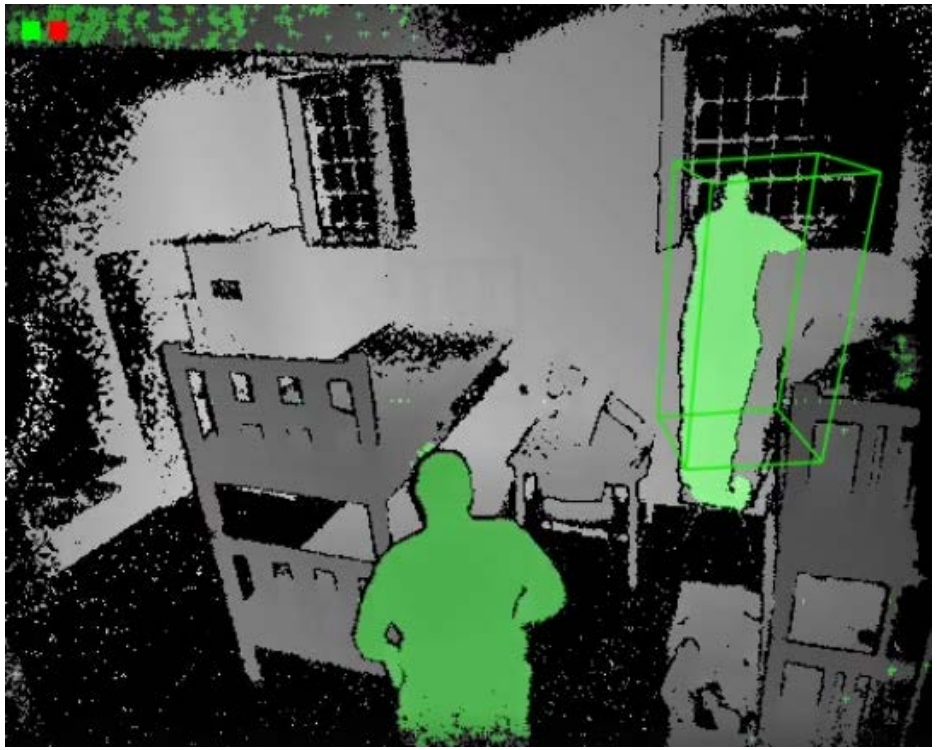
- Person schaut aus Fenster (10:9)
- Person sitzt oder steht auf Fensterbank (0:4)
- Person liegt auf dem Boden (0:1)

Aufnahmen

Analyse

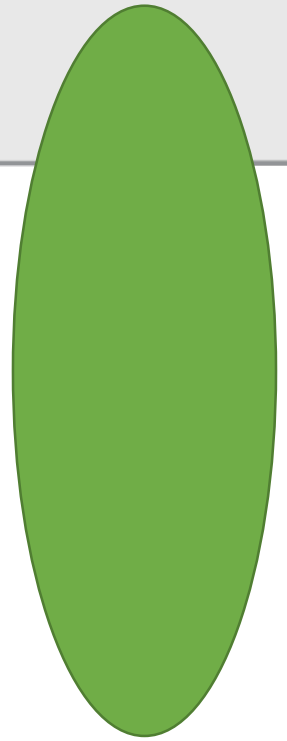
- Relative Häufigkeiten je nach Tageszeit

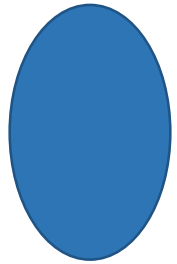




Agenda

- Introduction and Motivation
- Examples
- MIT Moral Machine
- It s the Data
  - Behavior modelling (fall detection)
  - cancer research
  - Detection of suicidal activities
- Views from the developer





Views from the developer I

During design and development process

Fairness

1. Like to have
2. Having in mind, but not actively followed
3. Not to compare with unbalanced / balanced data
4. Knowing that the solution is not fair
5. It is for us (European, male, high income, …), we are on the safe side

Views from the developer II

During design and development process

Bias

1. Fearing/knowing that there is bias
2. Not standardized process to avoid bias
3. Unavoidable
4. Less accepted then fairness
5. “we build technology for our society”, we are on the safe side

Views from the developer III

During design and development process

Transparency

1. There is no transparency, especially for non experts
2. We personally do a lot to understand the way technology works
3. It is ok, that design is not transparent, keeps our salaries high

Fairness in machine intelligence

"A developer does not care about fairness or transparency of his/her algorithms"

Main goals:

Robustness, accuracy – correct replies from the machine, User Interface – easy to use, scalability, getting high amount of data etc.

December 2018

(The topic is simply not on the list of his/her main achievements)

TU Wien
Computer Vision Lab

Favoritenstr. 9/193-1, 4. th floor

A-1040 Vienna, AUSTRIA

Phone: +43-1-58801-19364

Fax: +43-1-58801-18399

www.cvl.tuwien.ac.at

Vielen Dank für Ihre Aufmerksamkeit !