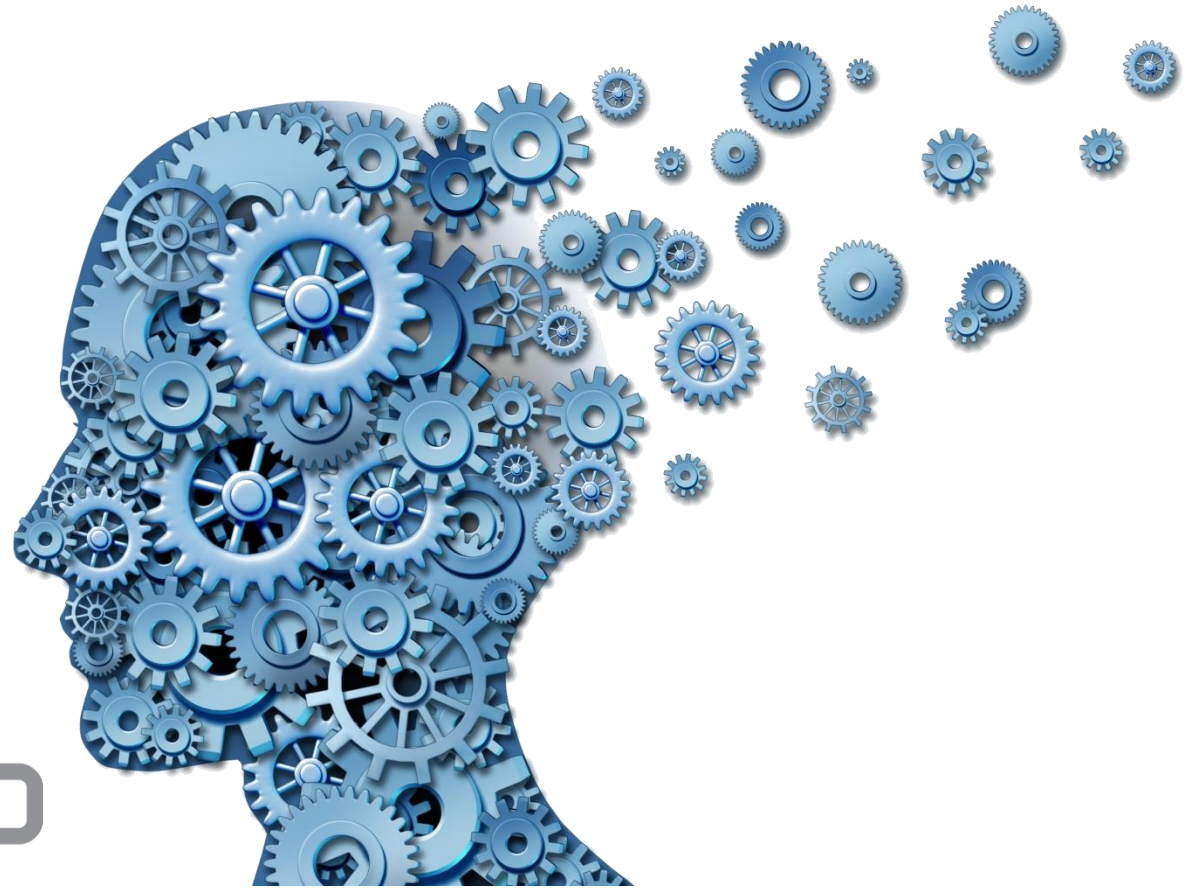
Zürcher Hochschule
für Angewandte Wissenschaften

Applied Data Science Research

A ZHAW Datalab perspective on the driver of digitization

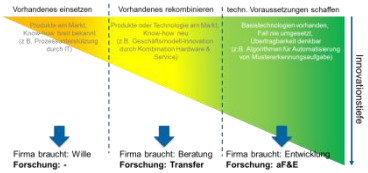
Kick-off Phd Network in Data Science
Zürich, April 13, 2018

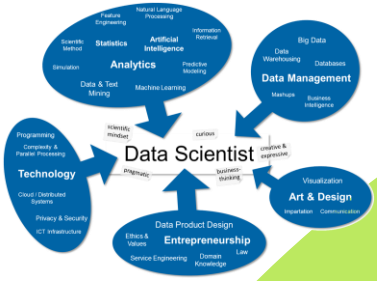
Thilo Stadelmann



datalab
www.zhaw.ch/datalab

Die nächsten 20 Minuten





Stand der Forschung

Rolle des PhD Programs

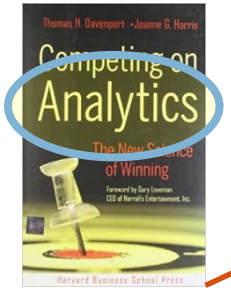
Data Science als Motor

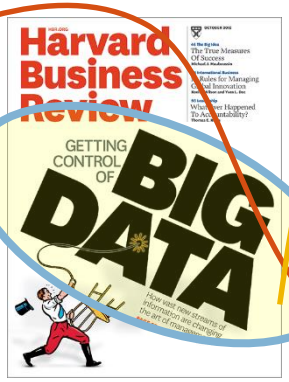
Ausblick

Zürcher Hochschule
für Angewandte Wissenschaften

Viele Begriffe, ein Trend: Digitalisierung

Schlagwörter und inhaltliche Treiber





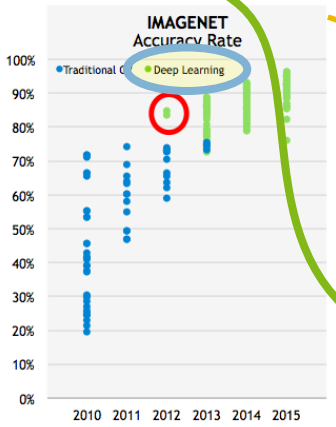
DIGITALisierung ZURICH 2025

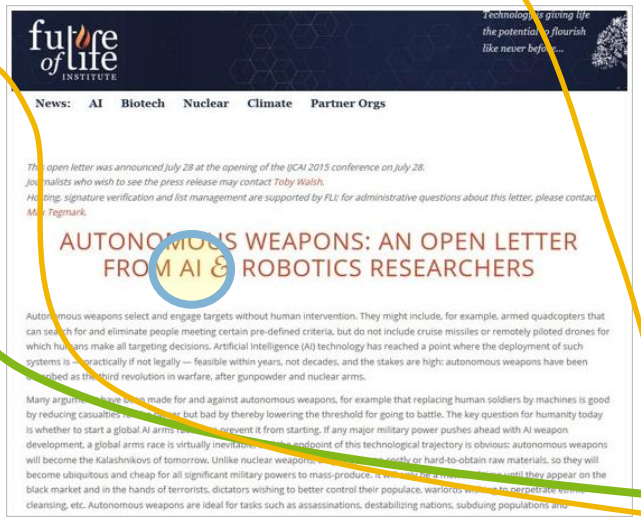
INDUSTRY 4.0 2025 INDUSTRIE INDUSTRIA

SWISS FINTECH INNOVATIONS

…

2007 2012 2016

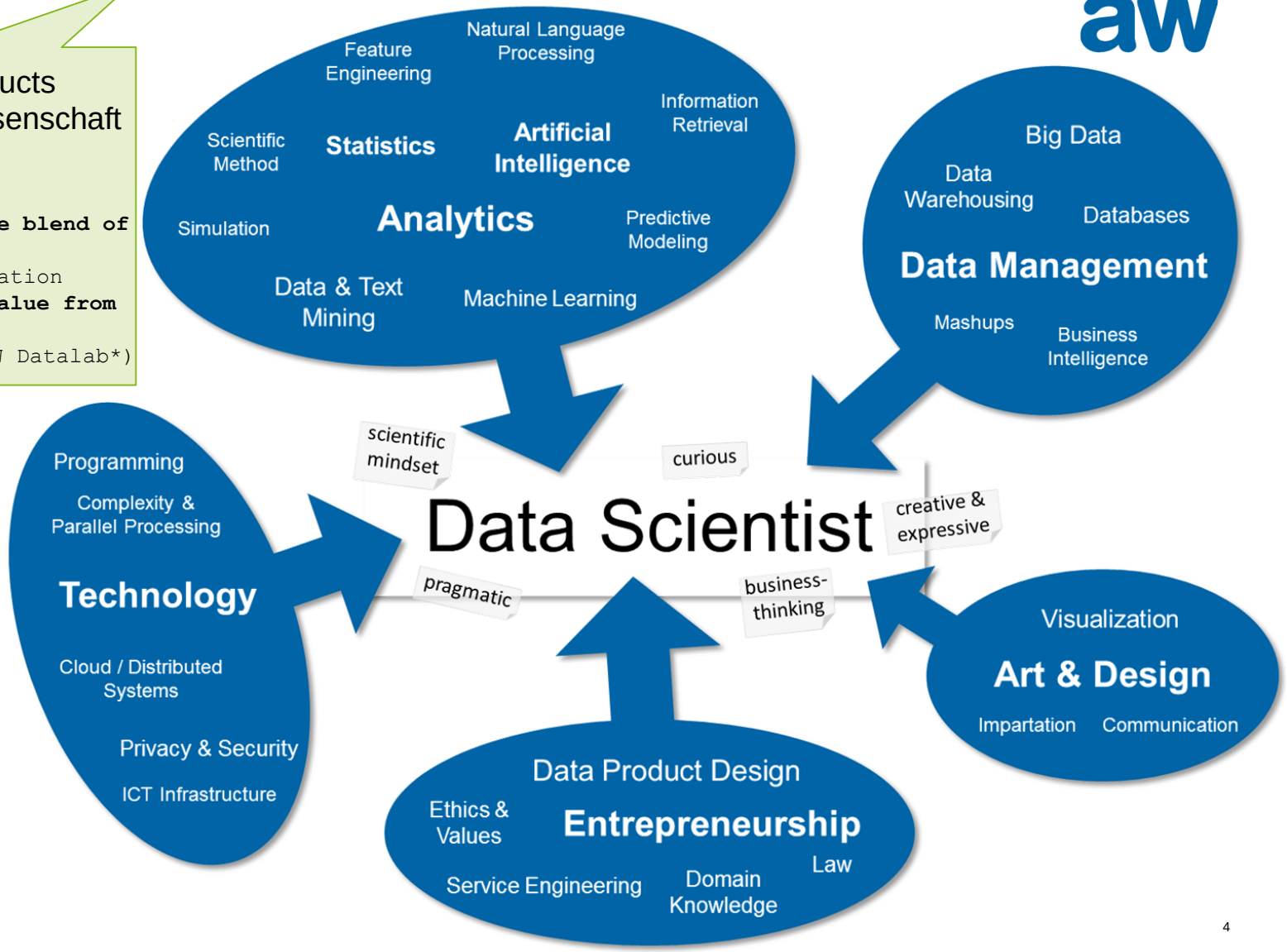




Was ist Data Science?

Ermöglicht Data Products
➔ **Angewandte** Wissenschaft
➔ Interdisziplinär

Data Science := "**Unique blend of skills** from analytics, engineering & communication aiming at **generating value from** the **data** itself […]"
(ZHAW Datalab*)



*) Stadelmann, Stockinger, Braschler, Cieliebak, Baudinot, Dürr and Ruckstuhl (2013). *Applied Data Science in Europe* . ECSS 2013, Amsterdam.

Teil 2

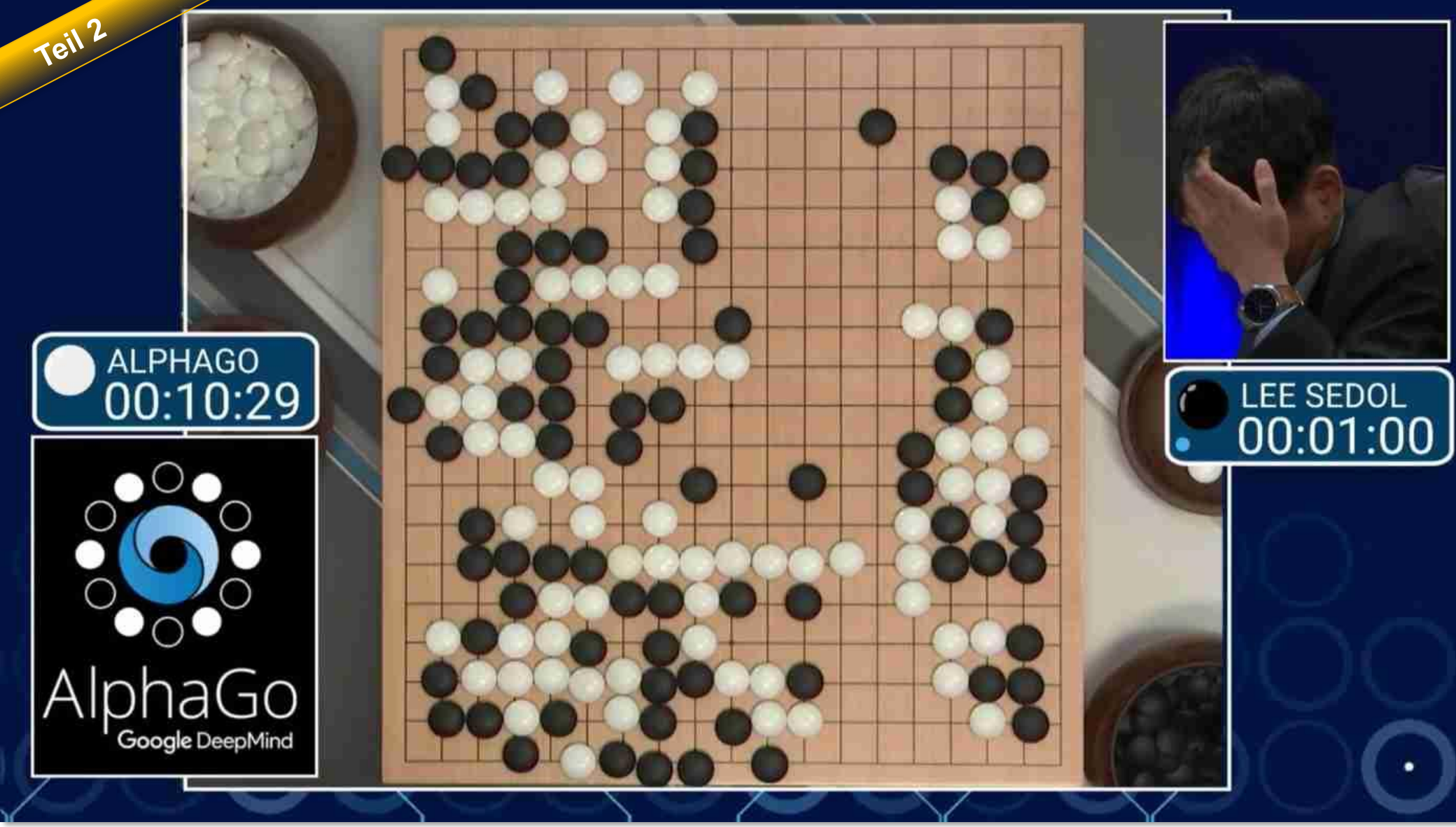
Zurich University of Applied Sciences

Der Stand der Forschung

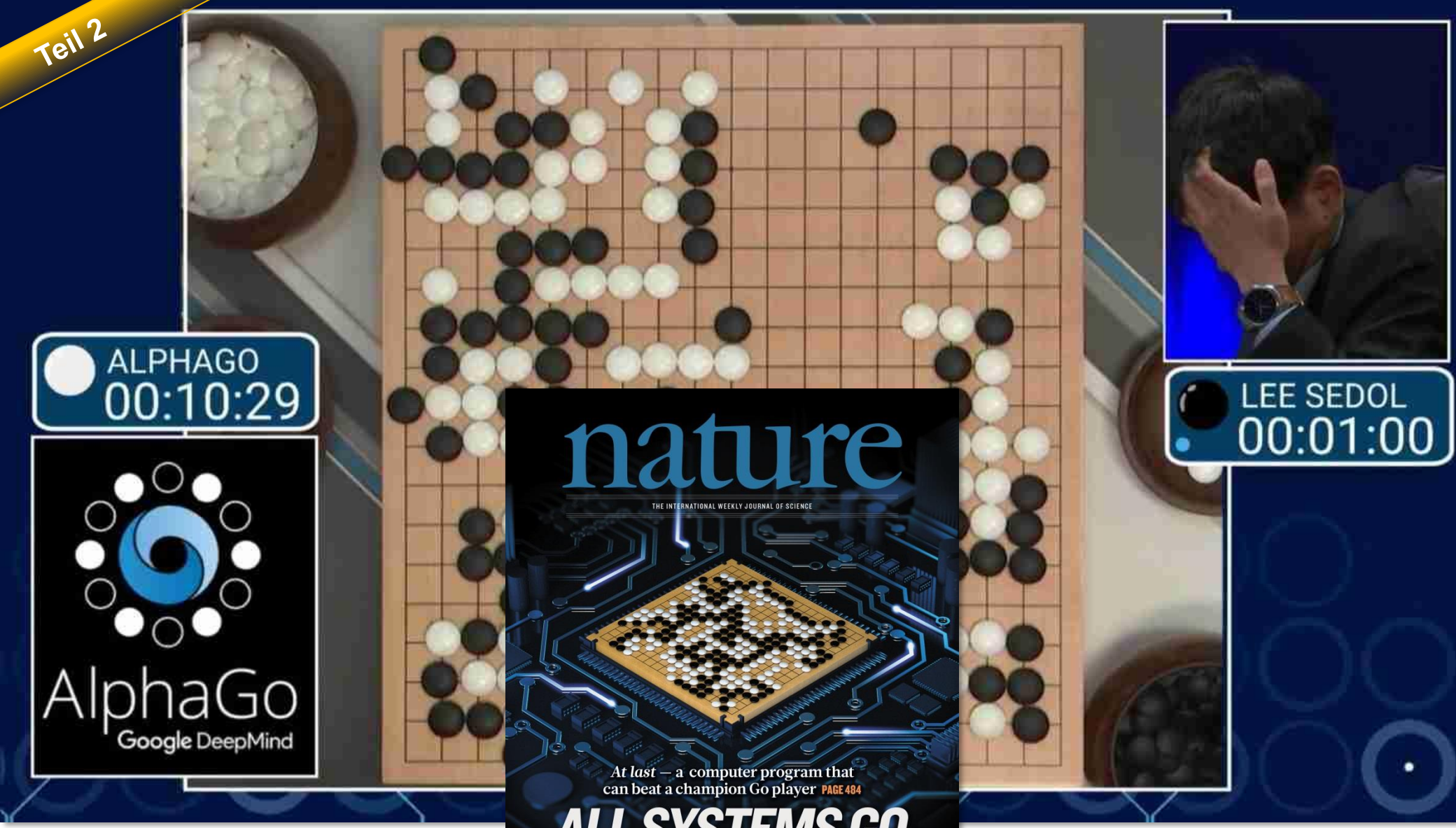
Durchbrüche allein im Bereich Deep Learning auf Big Data

Die **letzten 18 Monate** lieferten eine beeindruckende Liste **bedeutsame**r **Durchbrüche** in der Automatisierung **wahrnehmungsbezogener Aufgaben**.

➔ siehe die nächsten 2 Folien (weitere im Anhang)

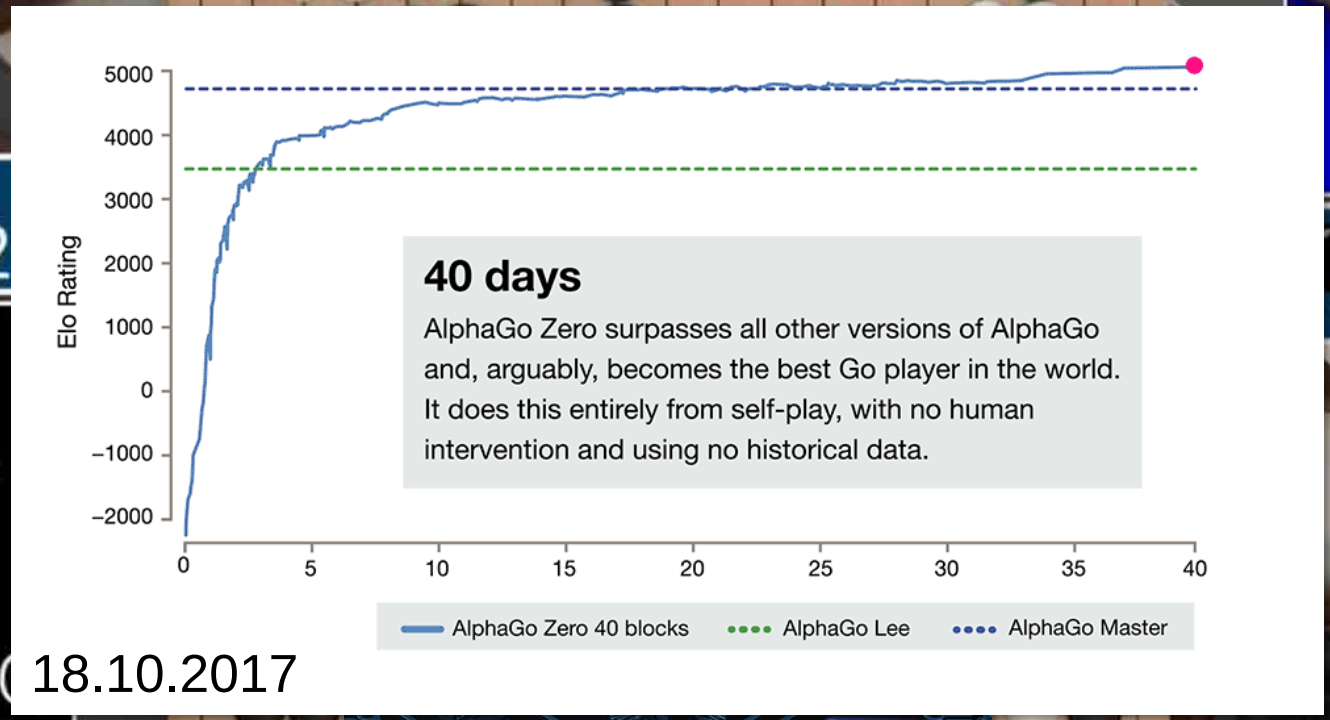

Teil 2
ALPHAGO
00:10:29
AlphaGo
Google DeepMind
LEE SEDOL
00:01:00

Teil 2


ALPHAGO
00:10:29
AlphaGo
Google DeepMind
LEE SEDOL
00:01:00

nature
THE INTERNATIONAL WEEKLY JOURNAL OF SCIENCE
At last — a computer program that can beat a champion Go player PAGE 484
ALL SYSTEMS GO
CONSERVATION
SONGBIRDS À LA CARTE
Illegal harvest of millions of Mediterranean birds
PAGE 452
RESEARCH ETHICS
SAFEGUARD TRANSPARENCY
Don't let openness backfire on individuals
PAGE 459
POPULAR SCIENCE
WHEN GENES GOT 'SELFISH'
Dawkins's calling card forty years on
PAGE 462
NATURE.COM/NATURE
28 January 2016 £10
Vol. 529, No. 7587

Teil 2



18.10.2017

Teil 2

…und viele weitere!

Zurich University of Applied Sciences

zh aw

Brandon Amos About Blog

Image Completion with Deep Learning in TensorFlow

August 9, 2016

- Introduction
- Step 1: Interpreting images as samples from a probability distribution
  - How would you fill in the missing information?
  - But where does statistics fit in? These are images.
  - So how can we complete images?
- Step 2: Quickly generating fake images
  - Learning to generate new samples from an unknown probability distribution
  - [ML-Heavy] Generative Adversarial Net (GAN) building blocks
  - Using $G(z)$ to produce fake images
  - [ML-Heavy] Training DCGANs
  - Existing GAN
  - [ML-Heavy]
  - Running DC
- Step 3: Finding the
  - Image compl
  - [ML-Heavy]
  - [ML-Heavy]
  - Completing y
- Conclusion
- Partial bibliography
- Bonus: Incomplete

Introduction

Content-aware fill is a p completion and inpaintin do content-aware fill, im "Semantic Image Inpaint shows how to use deep l some deeper portions fo section can be skipped i from images of faces. I h completion.tensorflow.

We'll approach image co

1. We'll first interpret
2. This interpretation
3. Then we'll find the

Andrej Karpathy blog About Hacker's guide to Neural Networks

The Unreasonable Effectiveness of Recurrent Neural Networks

May 21, 2015

There's something magical about Recurrent Neural Networks (RNNs). I still remember when I trained my first recurrent network for Image Captioning. Within a few dozen minutes of training my first baby model (with rather arbitrarily-chosen hyperparameters) started to generate very nice looking descriptions of images that were on the edge of making sense. Sometimes the ratio of how simple your model is to the quality of the results you get out of it blows past your expectations, and this was one of those times. What made this result so shocking at the time was that the common wisdom was that RNNs were supposed to be difficult to train (with more experience I've in fact reached the opposite conclusion). Fast forward about a year: I'm training RNNs all the time and I've witnessed their power and robustness many times, and yet their magical outputs still find ways of amusing me. This post is about sharing some of that magic with you.

> *We'll train RNNs to generate text character by character and ponder the question "how is that even possible?"*

By the way, together with this post I am also releasing code on Github that allows you to train character-level language models based on multi-layer LSTMs. You give it a large chunk of text and it will learn to generate text like it one character at a time. You can also use it to reproduce my experiments below. But we're getting ahead of ourselves; What are RNNs anyway?

Recurrent Neural Networks

Sequences. Depending on your background you might be wondering: *What makes Recurrent Networks so special?* A glaring limitation of Vanilla Neural Networks (and also Convolutional Networks) is that their API is too constrained: they accept a fixed-sized vector as input (e.g. an image) and produce a fixed-sized vector as output (e.g. probabilities of different classes). Not only that: These models perform this mapping using a fixed amount of computational steps (e.g. the number of layers in the model). The core reason that recurrent nets are more exciting is that they allow us to operate over sequences of vectors: Sequences in the input, the output, or in the most general case both. A few examples may make this more concrete:

```
VIOLA:
Why, Salisbury must find his flesh and thought
That which I am not aps, not a man and in fire,
To show the reining of the raven and the wars
To grace my hand reproach within, and not a fair are hand,
That Caesar and my goodly father's world;
When I was heaven of presence and our fleets,
We spare with hours, but cut thy council I am great,
Murdered and by thy master's ready there
My power to give thee but so much as hell:
Some service in the noble bondman here,
Would show him to her wine.

KING LEAR:
O, if you were a feeble sight, the courtesy of your law,
Your sight and several breath, will wear the gods
With his heads, and my hands are wonder'd at the deeds,
So drop upon your lordship's head, and your opinion
Shall be against your honour.
```

the right, a recurrent network generates images of digits by learning to sequentially add color to a canvas (Gregor et al.):

the morning paper

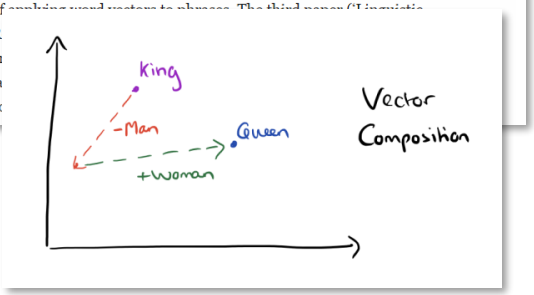
The amazing power of word vectors

APRIL 21, 2016

For today's post, I've drawn material not just from one paper, but from five! The subject matter is 'word2vec' – the work of Mikolov et al. at Google on efficient vector representations of words (and what you can do with them). The papers are:

- **Efficient Estimation of Word Representations in Vector Space** – Mikolov et al. 2013
- **Distributed Representations of Words and Phrases and their Compositionality** – Mikolov et al. 2013
- **Linguistic Regularities in Continuous Space Word Representations** – Mikolov et al. 2013
- **word2vec Parameter Learning Explained** – Rong 2014
- **word2vec Explained: Deriving Mikolov et al's Negative Sampling Word-Embedding Method** – Goldberg and Levy 2014

From the first of these papers ('Efficient estimation…') we get a description of the *Continuous Bag-of-Words* and *Continuous Skip-gram* models for learning word vectors (we'll talk about what a word vector is in a moment…). From the second paper we get more illustrations of the power of word vectors, some additional information on optimisations for the skip-gram model (hierarchical softmax and negative sampling), and a discussion



Teil 2

…und viele weitere!

Zurich University of Applied Sciences

zh aw

Brandon Amos About Blog

Image Completion with Deep Learning in TensorFlow

August 9, 2016

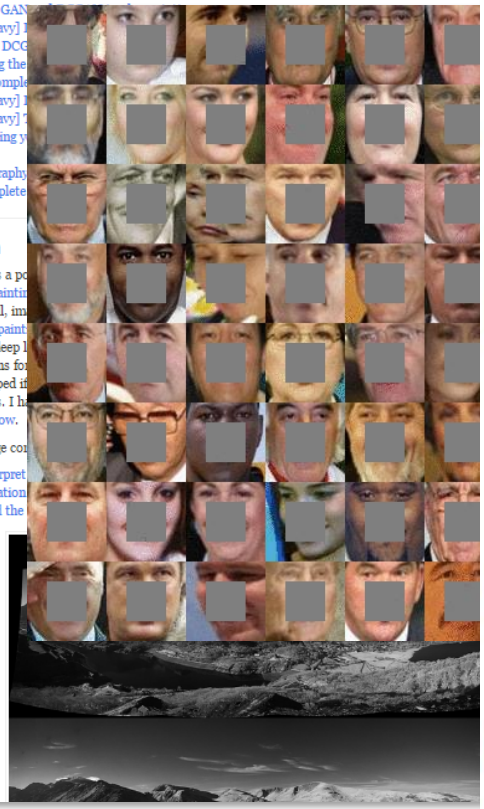
- Introduction
- Step 1: Interpreting images as samples from a probability distribution
  - How would you fill in the missing information?
  - But where does statistics fit in? These are images.
  - So how can we complete images?
- Step 2: Quickly generating fake images
  - Learning to generate new samples from an unknown probability distribution
  - [ML-Heavy] Generative Adversarial Net (GAN) building blocks
  - Using $G(z)$ to produce fake images
  - [ML-Heavy] Training DCGANs
  - Existing GAN
  - [ML-Heavy]
  - Running DC
- Step 3: Finding the
  - Image comple
  - [ML-Heavy]
  - [ML-Heavy]
  - Completing y
- Conclusion
- Partial bibliography
- Bonus: Incomplete

Introduction

Content-aware fill is a po
completion and inpaintin
do content-aware fill, im
"Semantic Image Inpaint
shows how to use deep l
some deeper portions fo
section can be skipped i
from images of faces. I h
completion.tensorflow.

We'll approach image co

1. We'll first interpret
2. This interpretation
3. Then we'll find the



Andrej Karpathy blog About Hacker's guide to Neural Networks

The Unreasonable Effectiveness of Recurrent Neural Networks

GEEK.COM

TECH

Nvidia AI Generates Fake Faces Based On Real Celebs

BY STEPHANIE MLOT 10.31.2017 :: 10:00AM EST

32 SHARES

I'm getting a distinctly mid-90s "The Rachel" vibe from the woman in the top left corner (via Nvidia)

STAY ON TARGET

AI Shelley Pens Truly Creepy Horror Stories–And You Can Help

Neural Network Serves Up Truly Frightening Halloween Costume Ideas

Celebrity scandals are about to get a lot more complicated.

Nvidia has **developed** a way of producing photo-quality, AI-generated human profiles—by using famous faces.

hand,

law,

ds,

the morning paper

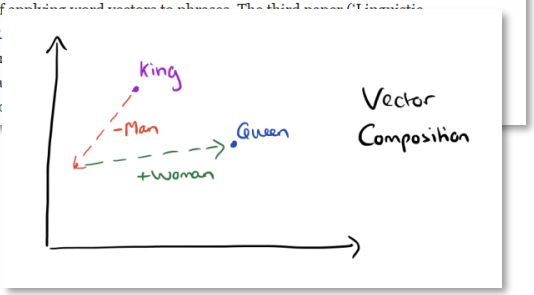
The amazing power of word vectors

APRIL 21, 2016

For today's post, I've drawn material not just from one paper, but from five! The subject matter is 'word2vec' – the work of Mikolov et al. at Google on efficient vector representations of words (and what you can do with them). The papers are:

- **Efficient Estimation of Word Representations in Vector Space** – Mikolov et al. 2013
- **Distributed Representations of Words and Phrases and their Compositionality** – Mikolov et al. 2013
- **Linguistic Regularities in Continuous Space Word Representations** – Mikolov et al. 2013
- **word2vec Parameter Learning Explained** – Rong 2014
- **word2vec Explained: Deriving Mikolov et al's Negative Sampling Word-Embedding Method** – Goldberg and Levy 2014

From the first of these papers ('Efficient estimation…') we get a description of the *Continuous Bag-of-Words* and *Continuous Skip-gram* models for learning word vectors (we'll talk about what a word vector is in a moment…). From the second paper we get more illustrations of the power of word vectors, some additional information on optimisations for the skip-gram model (hierarchical softmax and negative sampling), and a discussion



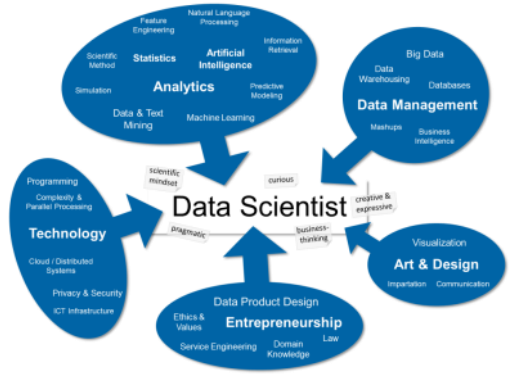
Teil 3

ZHAW Datalab: Est. 2013

Zurich University of Applied Sciences

Forerunner

- **One of the first** interdisciplinary data science initiatives in Europe
- One of the first interdisciplinary centers at ZHAW



Foundation

- **People**: ca. **70 researchers** from **5 institutes** / **3 departments** opted in
- Vision: **Nationally leading** and **internationally recognized** center of excellence
- Mission: **Generate projects** through critical mass and mutual relationships
- Competency: **Data product design** with structured and unstructured data

Success factors

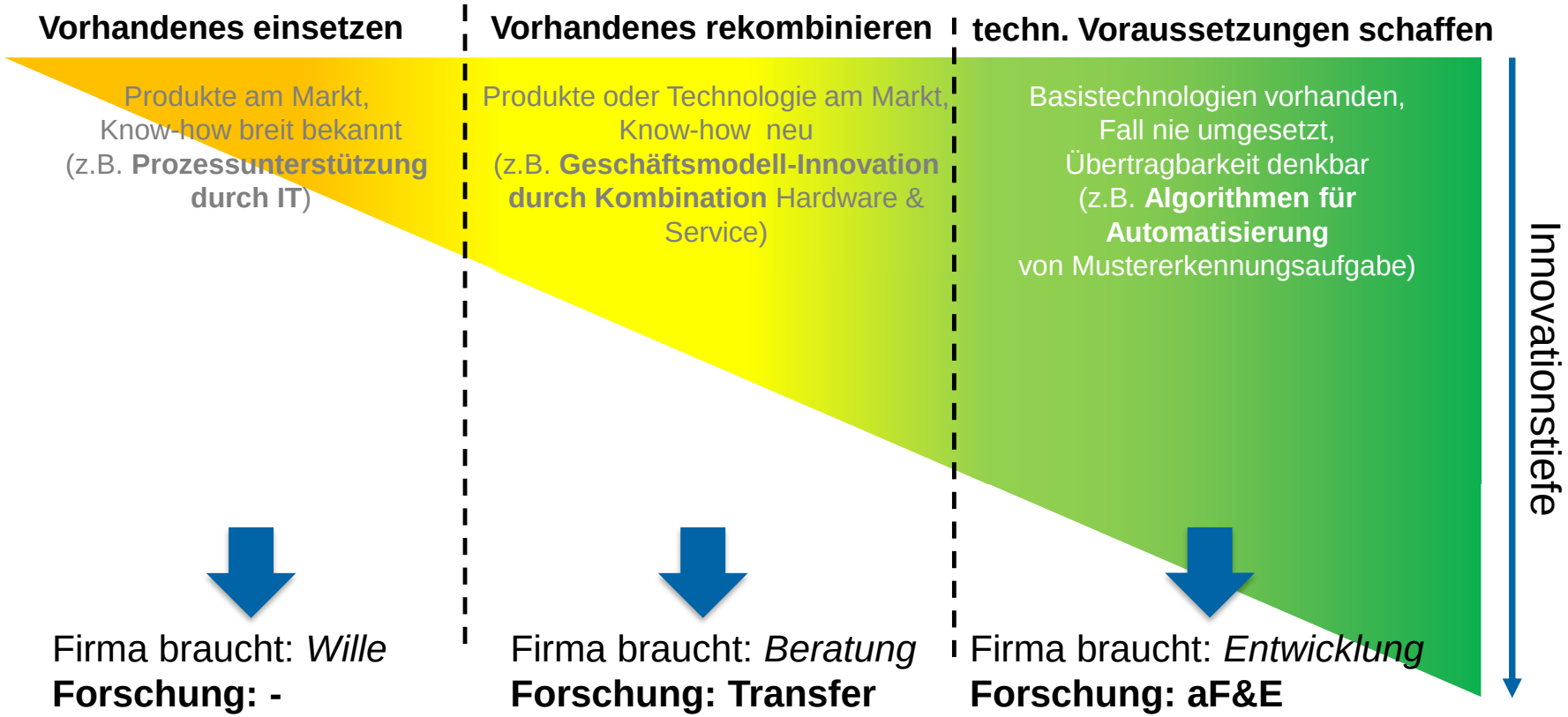
- **Lean** organization and operation → geared towards projects
- Years of successful **pre-Datalab collaboration**

Zürcher Hochschule
für Angewandte Wissenschaften

Unsere Rolle als Forscher an Fachhochschulen

Innovation in Zeiten der Digitalisierung → siehe Konferenz Digitale Schweiz

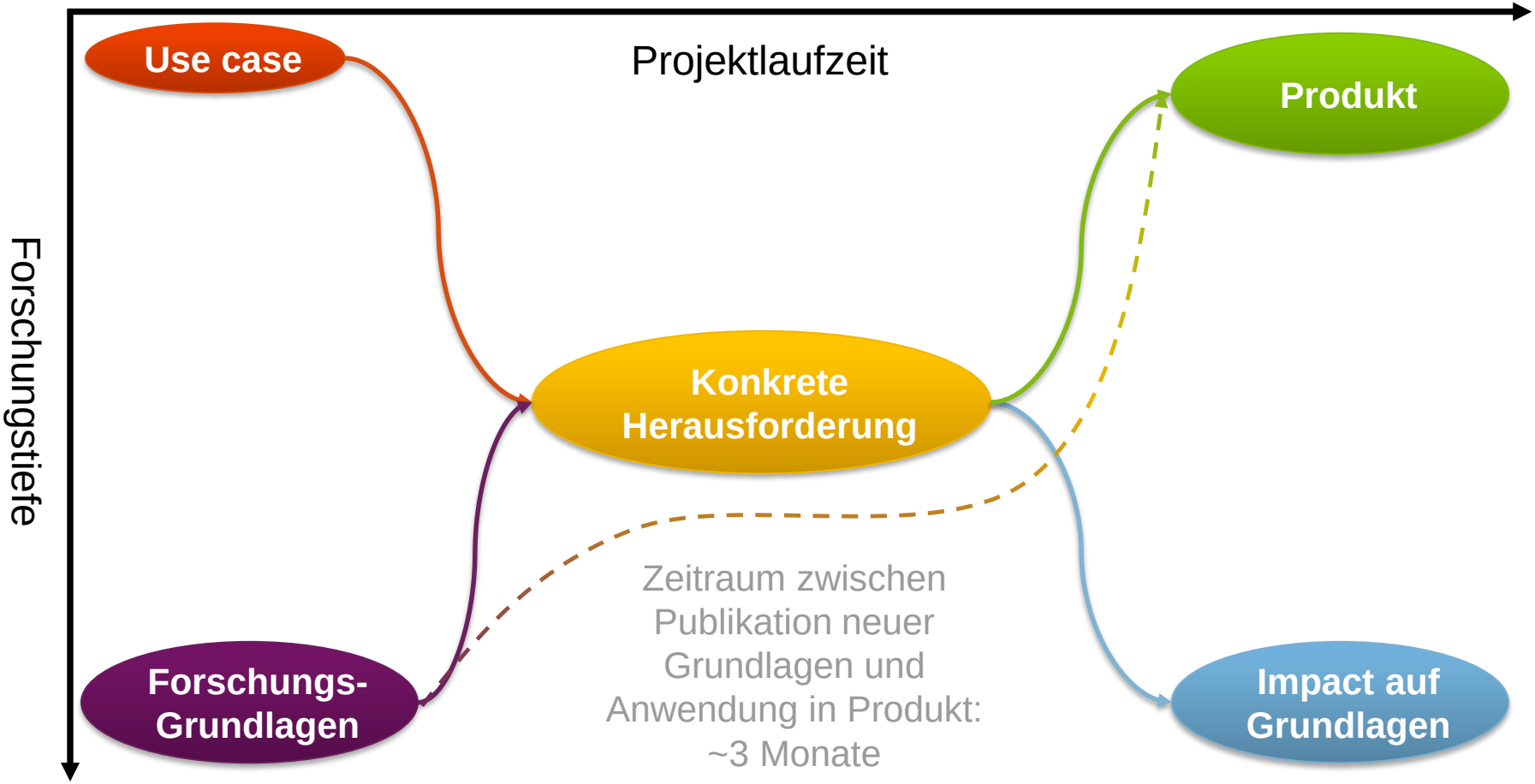
3 Arten von betrieblicher Innovation in der Digitalisierung



Zürcher Hochschule
für Angewandte Wissenschaften

Parallelität von Grundlagen- und aF&E

Innovation in Zeiten der Digitalisierung, contd.



Teil 3

Zürcher Hochschule
für Angewandte Wissenschaften

Angewandte F&E im ZHAW Datalab

Fünf Beispiele

- **Produktionsautomatisierung** für KMU mit *Deep Neural Networks*
- **Intuitive Suche** in Datenbanken für Bioinformatiker mit *Big Data Technologie*
- **Behandlungsplanung** für Aneurysmen mit *Maschinellem Lernen*
- **Energieoptimale Gebäudesteuerung** mit *Simulation* und *Model Predictive Control*
- **Automatische Agenten** in Computerspielen mit *Reinforcement Learning*

- (zwei weitere: siehe Anhang)

Teil 3

ARGUS MEDIA BASED INTELLIGENCE

PANOPTES – Automated Article Segmentation of Newspaper Pages for "Real Time Print Media Monitoring“

M. Cieliebak & T. Stadelmann, ZHAW

zhaw School of Engineering

datalab www.zhaw.ch/datalab

Overview

Partners

Who are we

ARGUS der Presse AG
- Switzerland's leading media monitoring and information provider
- Experience of more than 100 years

ZHAW Datalab
- Interdisciplinary research group at Zurich University of Applied Sciences
- Combining the knowledge of different fields related to machine learning

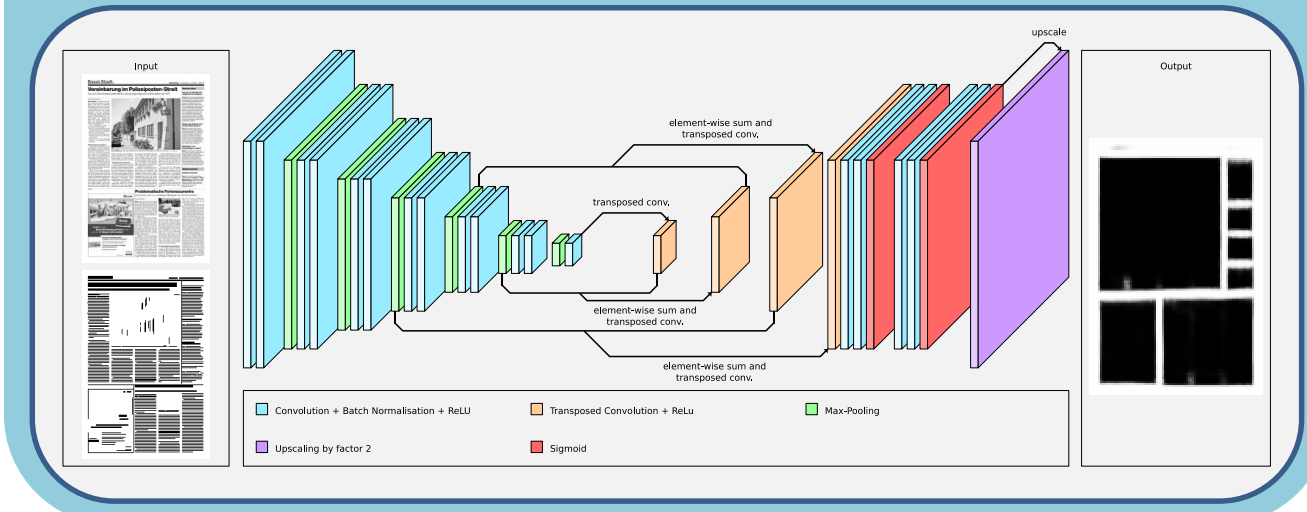
The Project

What do we do

Goal
- Real Time Print Media Monitoring
  - Extraction of relevant articles from newspaper pages
  - Delivering articles to customers

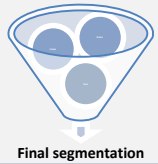
Problem
- Fully automated article segmentation
- Identification of article elements (e.g. title, subtitle, etc.)

Most Successful Approach [3]



Combination

Combination of rules, visual and textual features



Final segmentation

Result

References

[1] D. C. Ciresan, A. Giusti, L. M. Gambardella, and J. Schmidhuber. *Deep neural networks segment neuronal membranes in electron microscopy images.* In *NIPS*, pages 2852–2860, 2012.
[2] T. Mikolov, K. Chen, G. Corrado, and J. Dean. *Efficient Estimation of Word Representations in Vector Space.* In Proceedings of Workshop at *ICLR*, 2013.
[3] B. Meyer, T. Stadelmann, J. Stampfli, M. Arnold, M. Cieliebak. *Fully Convolutional Neural Networks for Newspaper Article Segmentation*. In Proceedings of ICDAR, Kyoto, Japan, 2018.

This project was funded by CTI under project number 17719.1 PFES-ES

Schweizerische Eidgenossenschaft
Confédération suisse
Confederazione Svizzera
Confederaziun svizra

Swiss Confederation

Federal Department of Economic Affairs, Education and Research EAER

Commission for Technology and Innovation CTI
Innovation Promotion Agency

swiss group for artificial intelligence and cognitive science

SGAICO

Teil 3

Zürcher Hochschule
für Angewandte Wissenschaften

Bio-SODA: Enabling Complex, Semantic Queries to Bioinformatics Databases through Intuitive Searching over Data

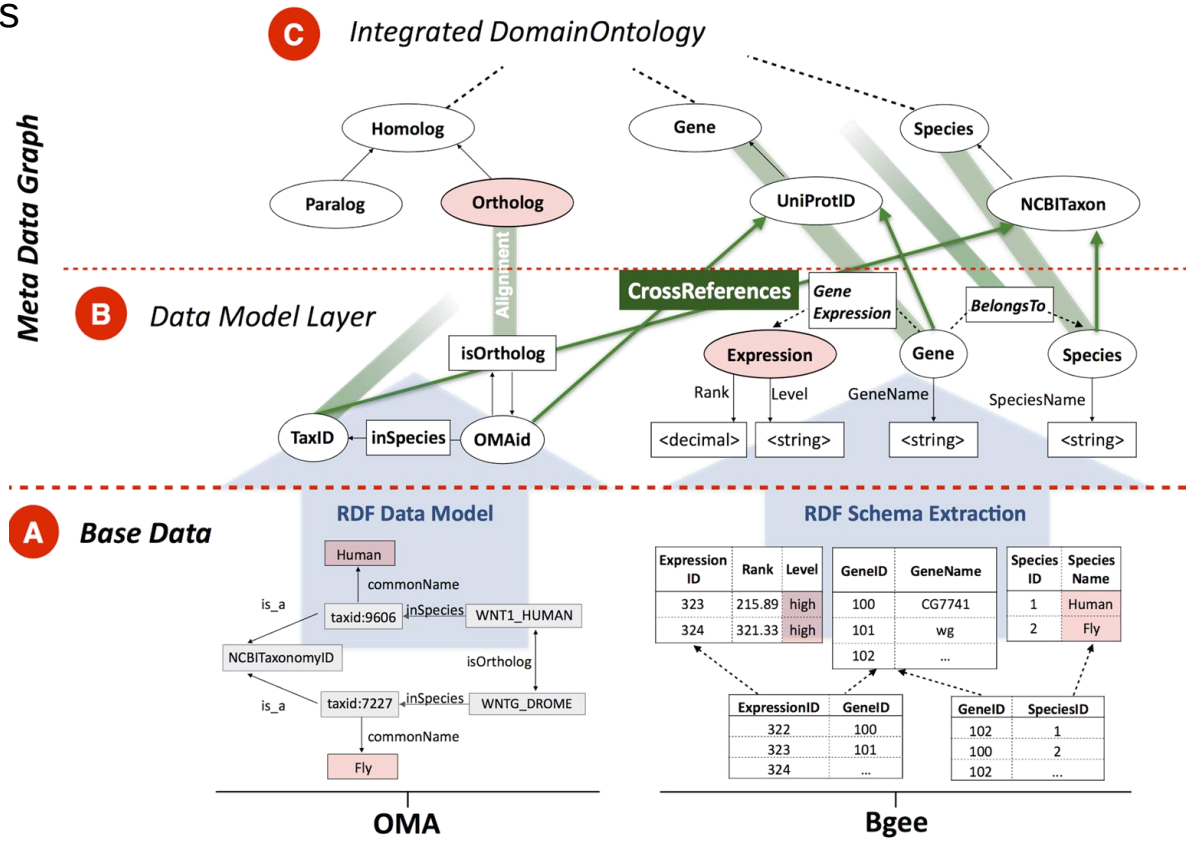
Intuitive exploration

- ✓ without knowing SPARQL, SQL, etc
- ✓ without knowing database schemas
- ✓ large datasets

Impact

- ➢ large bioinformatics user bases
- ➢ future federation of life sciences

Lead: Kurt Stockinger, ZHAW



75 NFP Big Data Nationales Forschungsprogramm

FNSNF Fonds national suisse Schweizerischer Nationalfonds Fondo nazionale svizzero Swiss National Science Foundation

Zurich University of Applied Sciences zhaw

UNIL | Université de Lausanne

SIB Swiss Institute of Bioinformatics

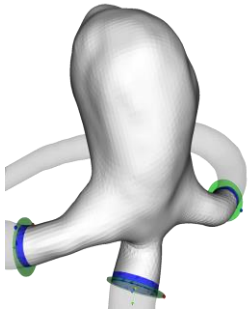
Teil 3

Zürcher Hochschule
für Angewandte Wissenschaften

AneuX: Ist die Form signifikant für die Gefährdung eines Aneurysmas?

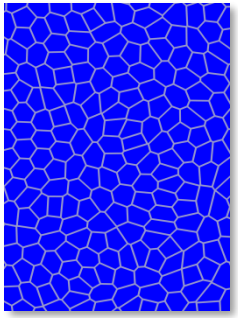
Aneurysm im Röntgenbild (XA)

Isoliertes Aneurysma Zur Formanalyse

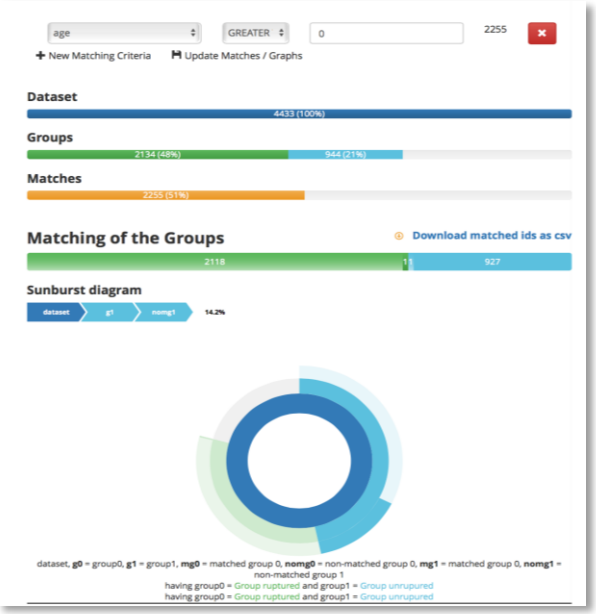


Zellen der Gefässwand

Modell der Gefässwand



Webtool für statistische Analyse



SystemsX.ch funding: 2M CHF, Begutachtung SNSF

- Morphologische Analyse von Aneurysmen mit Machine Learning
- Biologisch motiviertes Simulationsmodell für Zellwandveränderung
- Aufbau eines Krankheitsmodells für die Behandlungsplanung
- Aufbau einer Datenbank von Aneurysmen
- Erstellung von Werkzeuge zur Analyse der klinischen Daten und Bilddaten

Partner (Co-Antragsteller Sven Hirsch, ZHAW):

AneuX
Shape as Biomarker for Aneurysm Disease

HUG Hôpitaux Universitaires de Genève

Universität Zürich[UZH]

UNIVERSITÉ DE GENÈVE

cabmm

IT'IS FOUNDATION

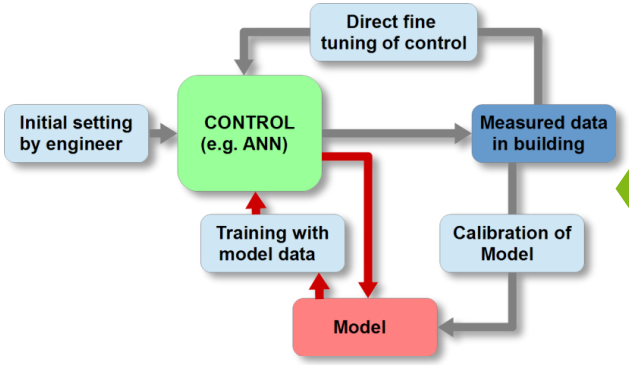
Swiss Neuro Foundation

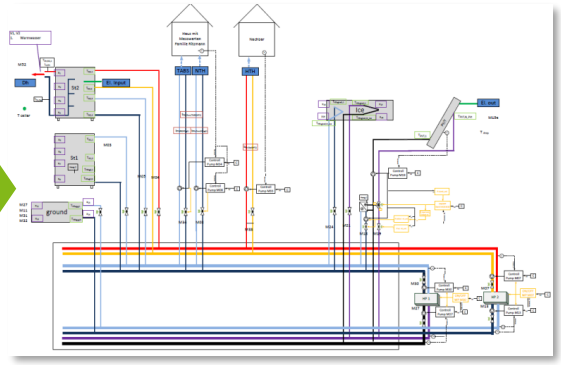
ETH zürich

Teil 3

Zürcher Hochschule
für Angewandte Wissenschaften

Hydrobus: Simulation-based Optimization





Offizielles Organ des Schaffhauser Bauernverbandes und seiner Fachsektionen

SCHAFFHAUSER BAUER 2

DONNERSTAG, 30. JUNI 2016, NUMMER 25

PILOTPROJEKT

Kreativität und Erfindergeist sind seine Markenzeichen

NACHGEFRAGT

«Faszinierend!»

The challenge

- Not enough training data for AI in socio-technological systems

The project

- **Self-adaption** of control to time-varying demands **in a multi-apartment building** using **simulations**
- Combined entropy and **energy optimization** of HVAC-system **based on Model Predictive Control**
- Integrates **renewable energy technology**, social dynamics and scenario-based weather prediction

The upside

- Enables a **Swiss SME** to harvest results from **modern mathematics, data science and AI**
- Gives **science** the opportunity to **test modern approaches on real-world problems**

RINO Electronics AG

Schweizerische Eidgenossenschaft
Confédération suisse
Confederazione Svizzera
Confederaziun svizra

Swiss Confederation

Innosuisse – Swiss Innovation Agency

Zürcher Hochschule
für Angewandte Wissenschaften

Schlussfolgerungen

Welche Rahmenbedingungen benötigen wir, um erfolgreich zu bleiben?

Drei Thesen

- Digitale Innovationen laufen in **extrem kurzen** Zyklen ab
- **Einfachere** digitale **Innovationen** bestehen in der neuartigen **Kombination** von vorhandenen **Technologien** mit einem geeigneten Prozess- und **Businessmodell**
- **Komplexere** digitale Innovationen verlangen eine **Gleichzeitigkeit** von **Grundlagen**forschung, **angewandter Forschung** und **Umsetzung**

Zwei Schlussfolgerungen

➔ Die besten **Forschungsideen beweisen sich am Markt** in der Anwendung

➔ Grundlagen und anwendungsorientierte **Forschung** verlaufen **verzahnt parallel** anstatt sequentiell