

# Umfragekodierung mittels künstlicher Intelligenz

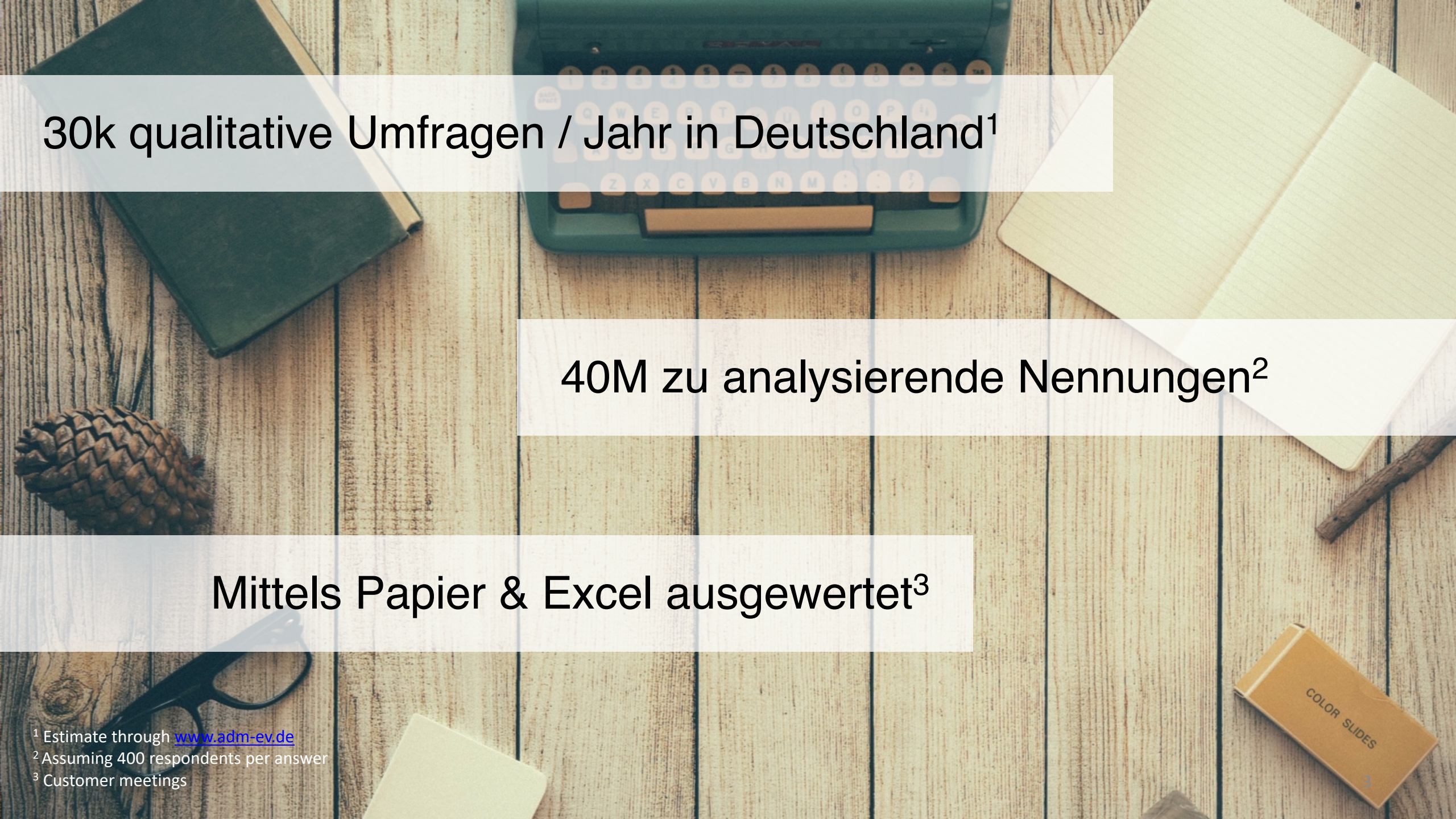
Maurice Gonzenbach, Caplena, Zürich, 22.3.2018



# Agenda

- 1. Problemstellung: Kodieren**
- 2. Ansätze**
- 3. Technischer Exkurs: KI und Kontext**
- 4. Resultate & Anwendungen**
- 5. Ausblick**





30k qualitative Umfragen / Jahr in Deutschland<sup>1</sup>

40M zu analysierende Nennungen<sup>2</sup>

Mittels Papier & Excel ausgewertet<sup>3</sup>

<sup>1</sup> Estimate through [www.adm-ev.de](http://www.adm-ev.de)

<sup>2</sup> Assuming 400 respondents per answer

<sup>3</sup> Customer meetings



# Analyse von offenen Fragen

## 1. Datenerhebung



## 2. Kodierung

sleek UI, impressive predictions



DESIGN

AI

ML model makes coding efficient



AI

EFFICIENT

amazing speedup and support

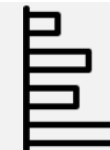


EFFICIENT

SUPPORT

## 3. Auswertung

DESIGN  
AI  
SUPPORT  
EFFICIENT



# Ansätze für automatische Textanalyse

## “Out of the Box”

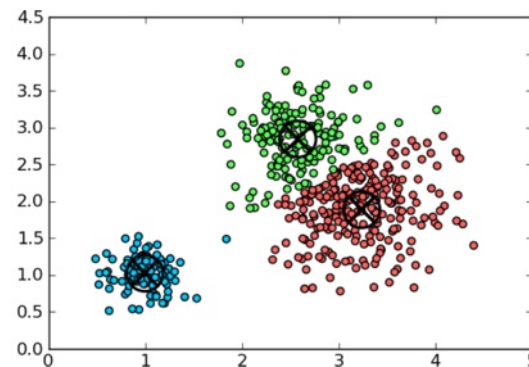
- Einfache APIs
- Vordefinierte Kategorien



Google Cloud Platform

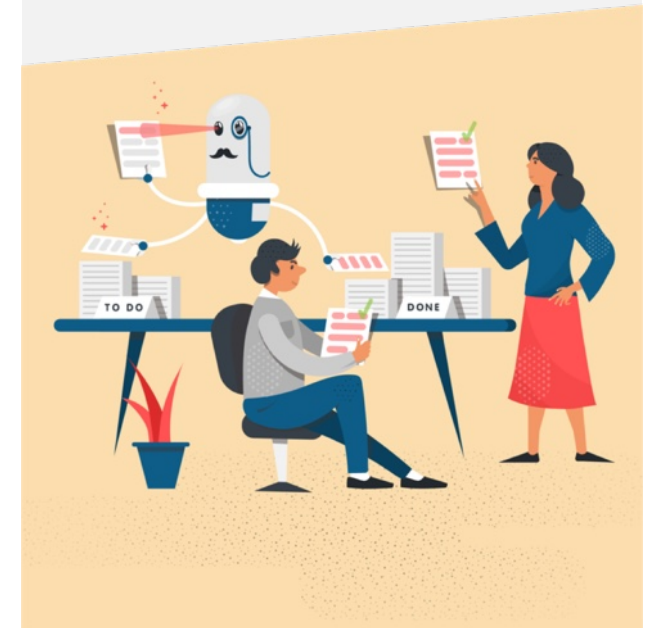
## Unüberwacht

- APIs / Open Source
- Grobe Gruppierung



## Überwacht

- Open Source / Codit.co
- Lernt von Beispielen

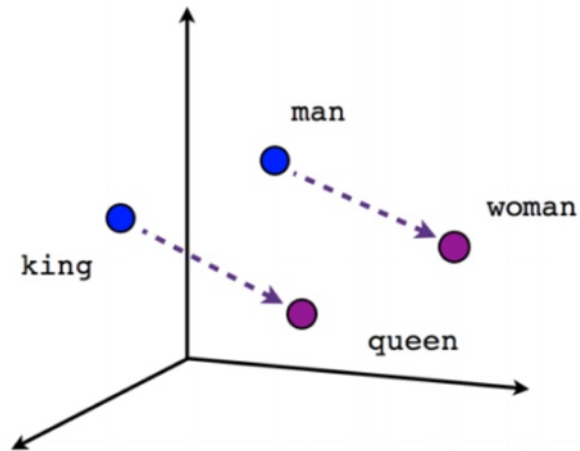


# Welches Bild steht auf dem Kopf?

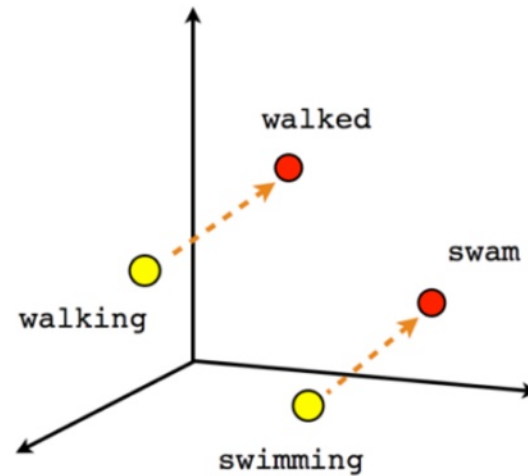


Kontext ist eine Herausforderung für Algorithmen

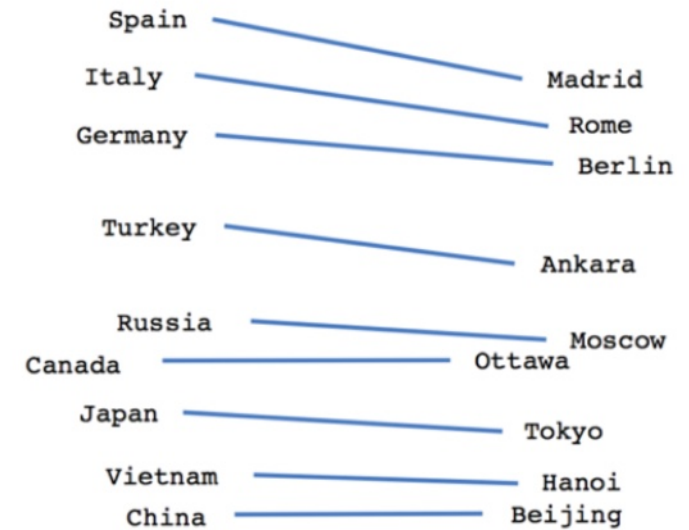
# Wie Computer mit Worten rechnen



Male-Female



Verb tense



Country-Capital

# Kontext ist entscheidend (1)

DATEN

## Filmkritik

- ~ 7.9M Amazon Filmkritiken
- 1.5B Wörter, 600k Einzigartige

## Nachrichten

- Verschiedene Nachrichtenportale, 2013-2017
- 4.5B Wörter, 1.2M Einzigartige

ÄHNLICHKEITEN

**WATSON**  
**Holmes, Sherlock, Emma, Mycroft...**

**WATSON**  
**Smith, Clarke, Hodge, McGilchrist...**

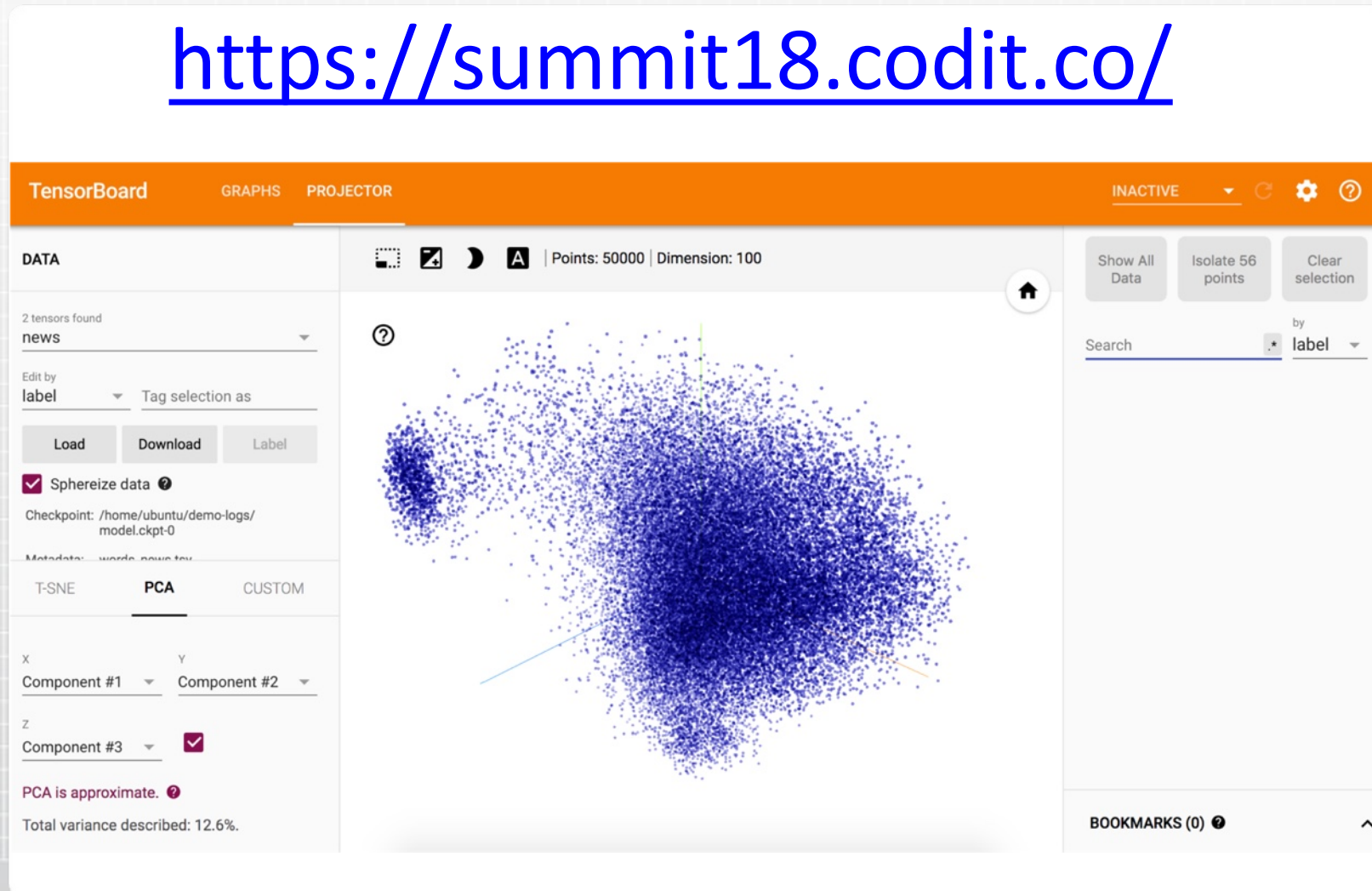


# Kontext ist entscheidend (2)

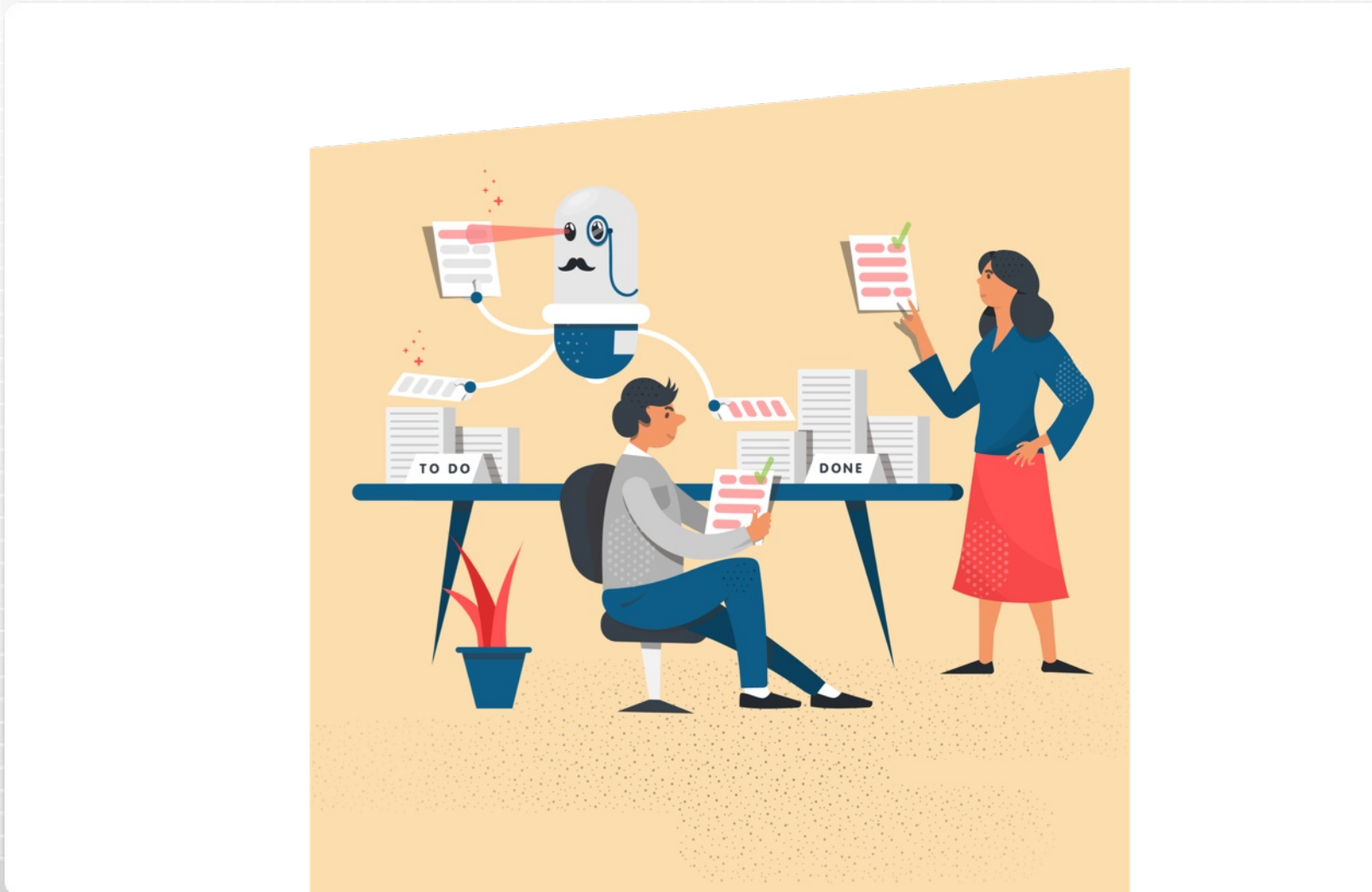
DATEN	<b>Filmkritik</b>	<b>Nachrichten</b>
ÄHNLICHKEITEN	<b>ABSORBING engrossing, compelling, riveting...</b>	<b>ABSORBING reabsorbing, high-concentration</b>
ANALOGIEN	<b>LUKE to HERO like DARTH to VILLAIN</b>	<b>US to FOOTBALL like ENGLAND to RUGBY</b>

# Entdecken Sie es selbst

<https://summit18.codit.co/>

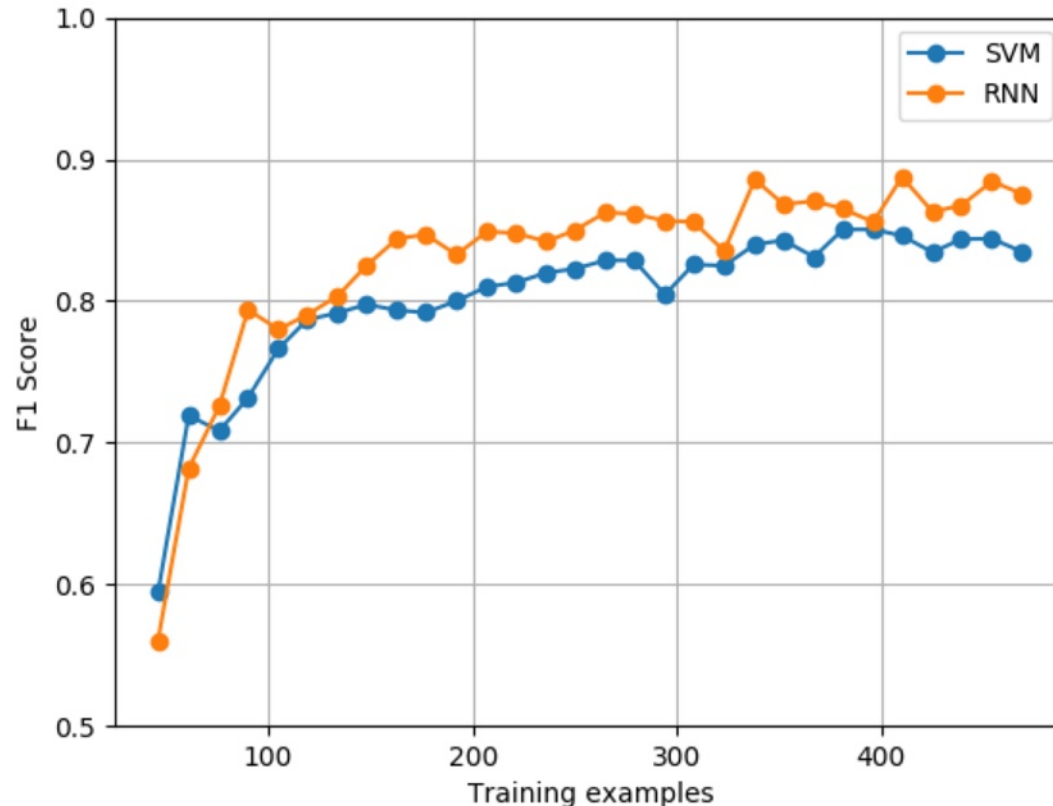


# Menschen als Lehrer & Supervisor





# Auf Umfragen spezialisierte KI

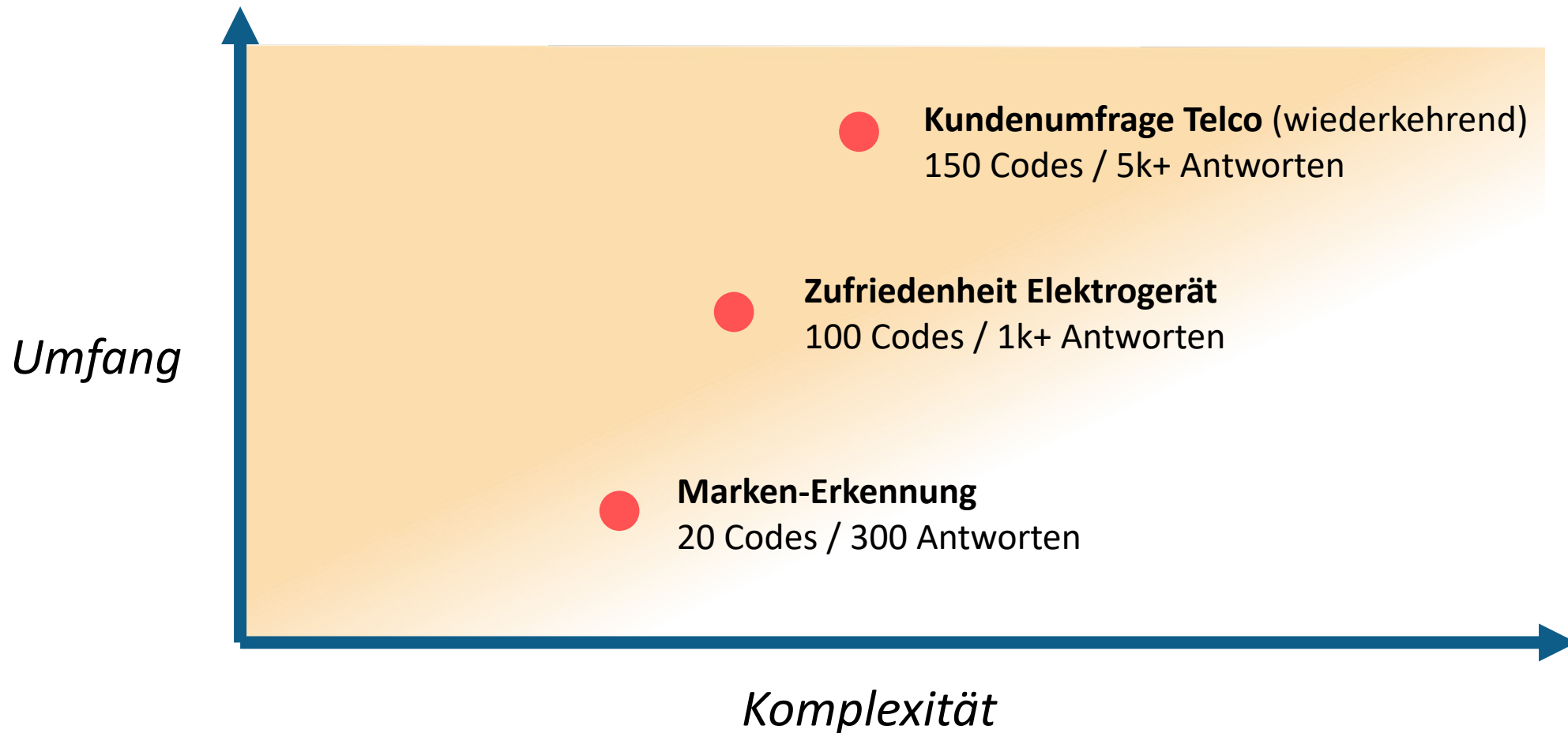


*Umfrage über Marken-Erkennung mit 15 Codes*

Lernt schnell durch

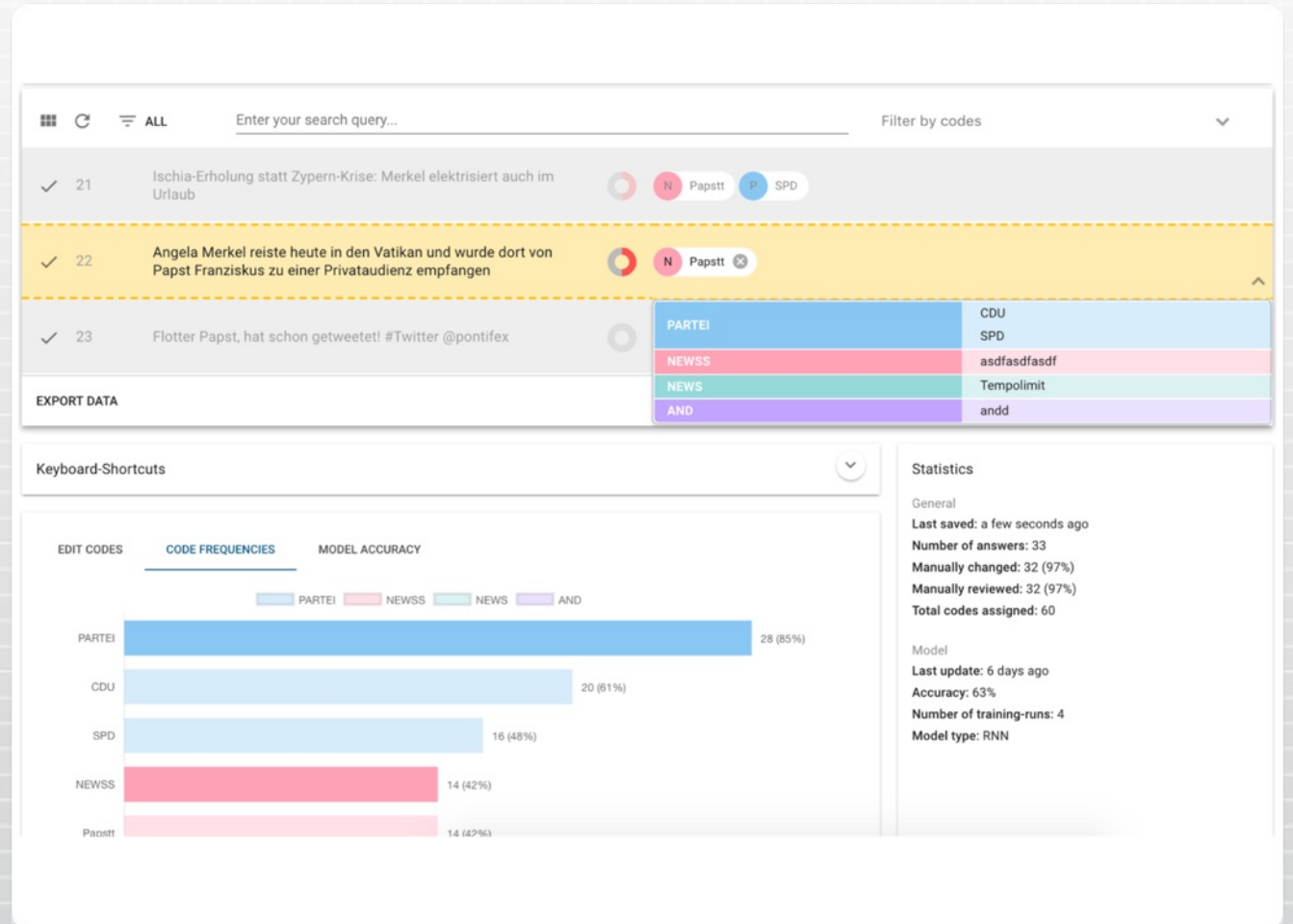
- Aufbau von Sprachverständnis mittels Millionen von Dokumenten
- **Transfer-Lernen** auf ähnlichen Umfragen

# Wo bringt maschinelle Kodierung am Meisten?



# Zusammenarbeit zwischen Mensch & Algorithmus

- Simple Benutzeroberfläche
- KI schlägt Codes für Antworten vor
- Mit jeder manuell verkodeten Antwort wird die KI schlauer
- Anwender bleibt Entscheidungsträger



The screenshot displays a web application for manual coding of survey responses. The interface includes a search bar, a list of survey items with suggested codes, a keyboard shortcuts section, and a statistics section.

**Survey Items and Suggested Codes:**

Item ID	Text	Suggested Codes
21	Ischia-Erholung statt Zypern-Krise: Merkel elektrisiert auch im Urlaub	Papstt, SPD
22	Angela Merkel reiste heute in den Vatikan und wurde dort von Papst Franziskus zu einer Privataudienz empfangen	Papstt
23	Flotter Papst, hat schon getweetet! #Twitter @pontifex	PARTEI, CDU, SPD, NEWS, Tempolimit, AND

**Keyboard-Shortcuts:**

Code	Value
PARTEI	CDU
NEWS	SPD
NEWS	asdfsdfasdfsdf
NEWS	Tempolimit
AND	andd

**Statistics:**

General

- Last saved: a few seconds ago
- Number of answers: 33
- Manually changed: 32 (97%)
- Manually reviewed: 32 (97%)
- Total codes assigned: 60

Model

- Last update: 6 days ago
- Accuracy: 63%
- Number of training-runs: 4
- Model type: RNN

**Code Frequencies:**

Code	Count	Percentage
PARTEI	28	85%
CDU	20	61%
SPD	16	48%
NEWS	14	42%
Papstt	14	42%



# Wie unsere Kunden Cudit.co benutzen

## Marktforschung

- Verarbeitung diverser Umfragen während der Mensch in Kontrolle bleibt
- Effizienzgewinne, insbesondere bei grösseren oder wiederholten Umfragen

**Schweizer Marktforschungsunternehmen**

## B2C

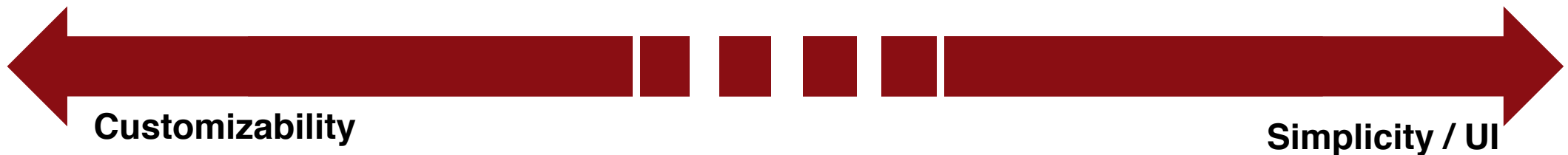
- Analyse von Kunden-Feedback
- Schnellere Entscheidungswege durch automatische Analyse von neuen Antworten

**Deutsche Telco**

# Wir wollen durch Fokus & Einfachheit überzeugen



MACHINE BOX

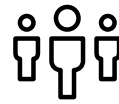


# Unser Weg bis anhin



## Stats

*80k + Antworten analysiert über unsere Plattform*



## Lancierung

*Februar 2018*



## Unsere Kunden

„[cudit.co](https://cudit.co) macht Kodieren extrem leicht. Schneller Einstieg, simple Bedienung und die hilfreiche KI machen das Tool momentan zur besten Option für mich.“

Marc Haupt, freiberuflicher Marktforscher



# Branchentrends & Weiterentwicklung

## Marktforschung

- Klassische Umfragen werden bleiben
- Differenzierung durch Kombination von Technologie & Qualität

## Codit.co

- Dezentralisierung der Kodierleistung
- Aufbau von «Modellbibliothek»
- Marktplatz zum Austausch von Modellen

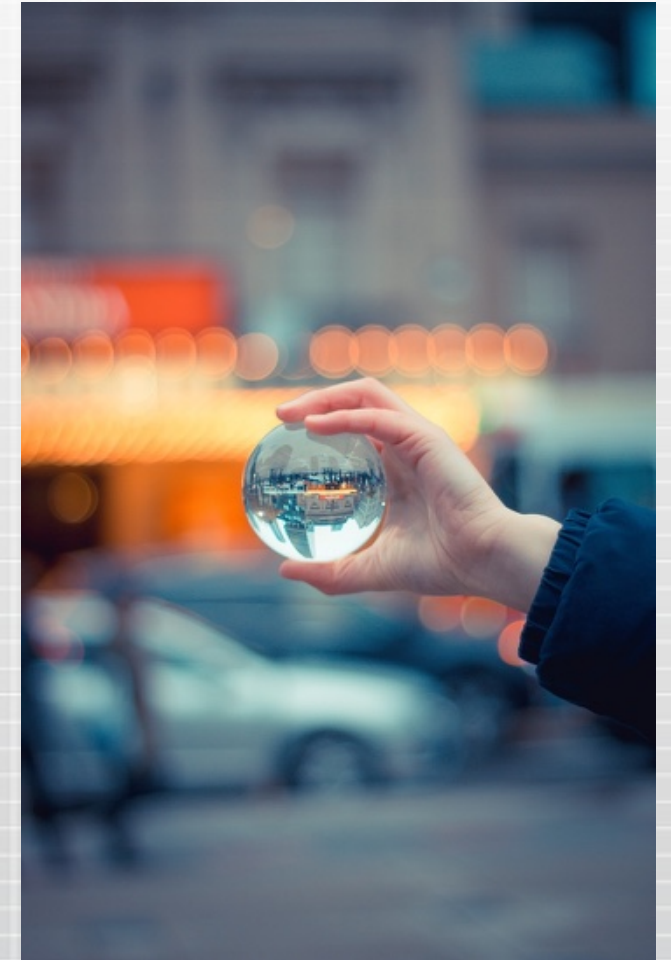


Photo: Bekir Dönmez via Unsplash

?

Probieren Sie es aus! <https://codit.co>  
Interaktive Visualisierung der Technologie:  
<https://summit18.codit.co>

# CAPLENA

## Kontakt



### **Maurice Gonzenbach**

[maurice@caplena.com](mailto:maurice@caplena.com)

Tel. +41 79 455 52 17

[www.caplena.com](http://www.caplena.com)



### **Pascal de Buren**

[pascal@caplena.com](mailto:pascal@caplena.com)

Tel. +41 79 518 91 75

[www.caplena.com](http://www.caplena.com)

