„LIEBER KÜNSTLICH INTELLIGENT ALS NATÜRLICH DUMM“

DR. SEBASTIAN FISCHER – AL/ML @ T-LABS

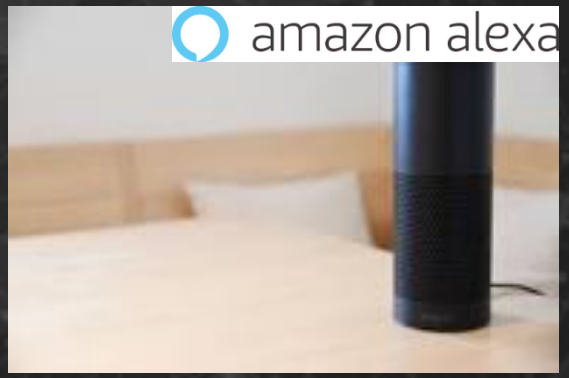
T · ·

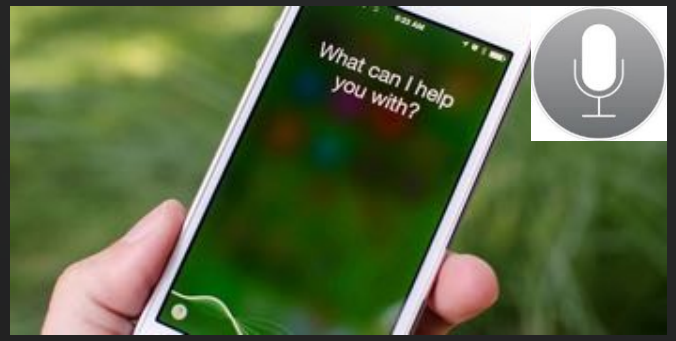
LIFE IS FOR SHARING.

Telekom
Innovation
Laboratories

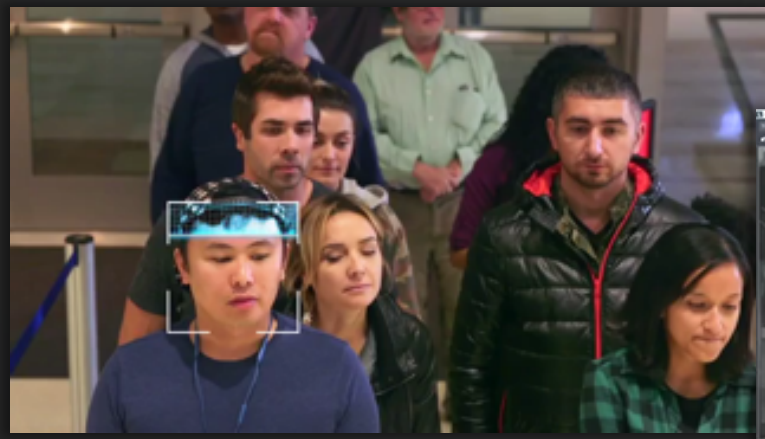
#BLOCKCHAIN
#INTELLIGENCE
#EXPERIENCES

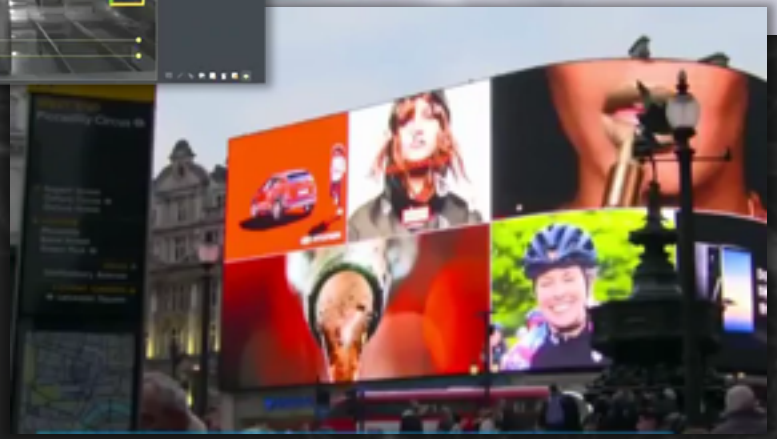
AI EVERYWHERE!





NEXT GENERATION AI





AI FOR TELCOS

AI APPLICATION AREAS

VIRTUAL ASSISTANTS

- "Human touch"-like services in customer journey
- 24/7 problem solving
- Automated & improved CRM incl. self service

AGENT & CRM SUPPORT

- Sales agent support
- Customer churn prediction & prevention
- Work force optimization

INTERNAL PROCESSES

- Security & fraud management
- Financial forecast automation
- Product & pricing adjustments
- Revenue assurance & debt collection

NETWORK

- (Automated) network planning & optimization
- Zero-touch operations
- Predictive maintenance

PRODUCTS

- Product recommendation engines
- Personalization (e.g. of TV content)
- Capacity/availability Improvements

AI CORE THEMES

INTERACTING WITH PEOPLE

SUPPORTING PEOPLE

SOLVING COMPLEX PROBLEMS

BECOMING PART OF PRODUCTS


DANGER
THIN ICE

AI UNDER ATTACK

Cybersecurity heads the lists of AI-related concerns

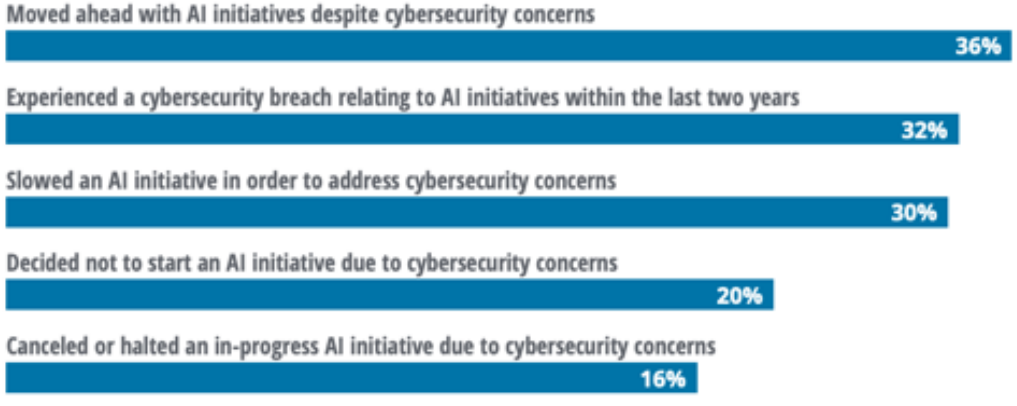
Potential AI risks of top concern to companies: Ranked 1–3, where 1 is greatest concern

	Ranked 1	Ranked 2	Ranked 3	Ranked top three
Cybersecurity vulnerabilities of AI	23%	15%	13%	51%
Making the wrong strategic decisions based on AI	16%	13%	14%	43%
Legal responsibility for decisions/ actions made by AI systems	11%	15%	13%	39%
Failure of AI system in a mission-critical or life-or-death context	13%	14%	12%	39%
Regulatory noncompliance risk	12%	15%	10%	37%
Erosion of customer trust from AI failures	11%	11%	11%	33%
Ethical risks of AI	10%	12%	10%	32%

Source: Deloitte State of AI in the Enterprise, 2nd Edition, 2018.

Cybersecurity threats are giving some companies pause

Effect of cybersecurity concerns on companies



Source: Deloitte State of AI in the Enterprise, 2nd Edition, 2018.

AI UNDER ATTACK

ATTACK

Models can be fooled through **malicious input by adversaries**

BIAS

Bad data used to train AI can contain implicit racial, gender, or ideological biases

PRIVACY

Risk of **breach of proprietary data** through attacks on AI models


00:06.73

STAY HUMAN

- We are **responsible**
- We **care**
- We put our **customers first**
- We are **transparent**
- We are **secure**
- We **set the grounds**
- We **keep control**
- We foster the **cooperative model**
- We **share** and **enlighten**

WE MUST TURN OUR ETHICAL AMBITIONS INTO VERIFIABLE ACTIONS.

LIFE IS FOR SHARING.

BE RESPONSIBLE

DATA IS THE CORE

FROM DATA TO ACTIONABLE INSIGHTS

DATA RELATED CHALLENGES & VALUE GENERATION POTENTIAL

- How to **deal with data** volume, velocity, veracity, variety (structured/ unstructured, real-time/non real-time)
  - to store
  - to move
  - to integrate
  - to search
  - to transform
- How to **make sense of data**?
  - to know what happened (hindsight/oversight)
  - to understand & explain why (insight)
  - to forecast (foresight)
- How to **translate insights to business value**?
- How to make insights actionable?

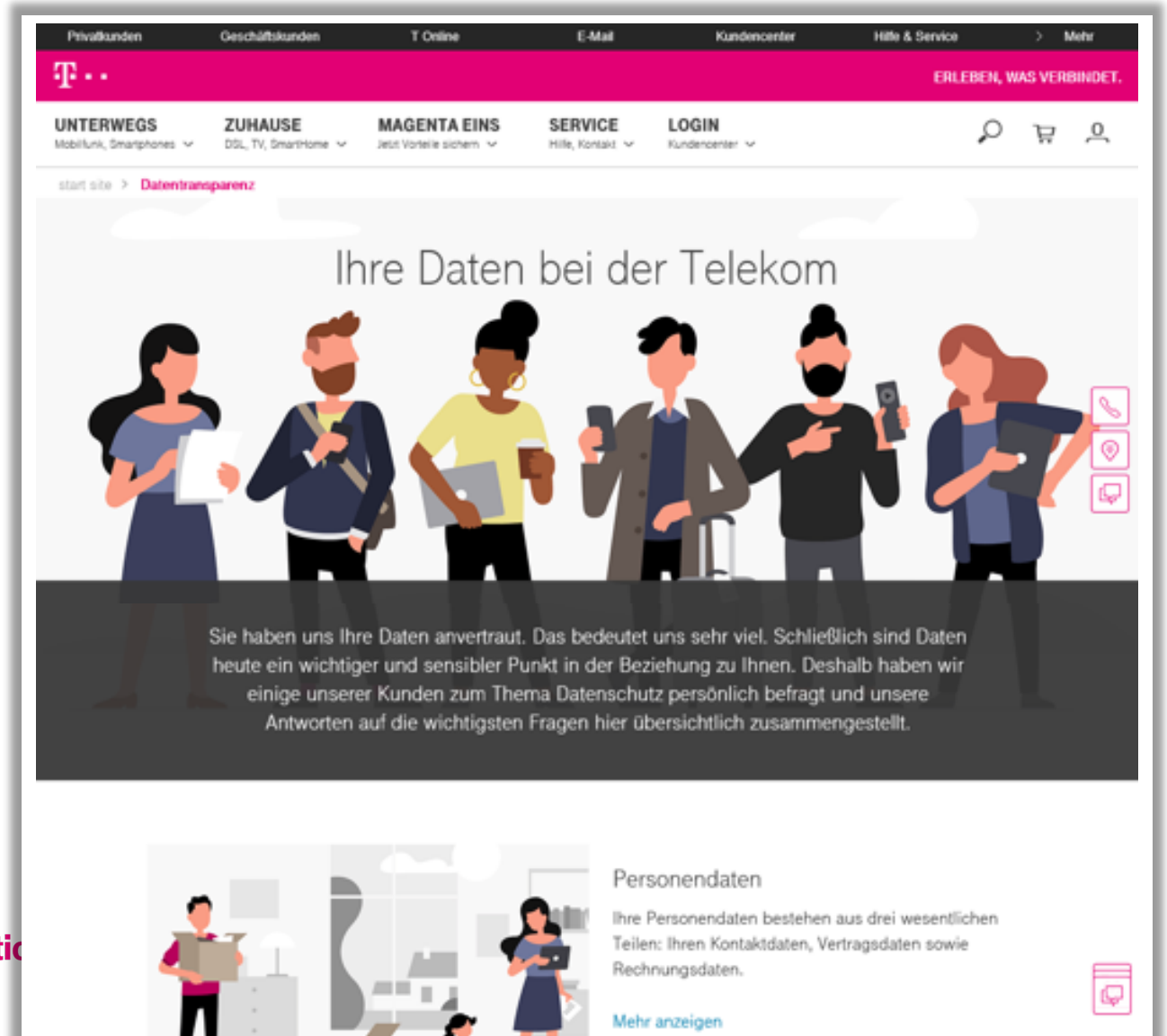
From Data to Insights: Challenges

Actual business value generation

DATA GOVERNANCE

- How is data ownership defined?
- How to manage data usage and privacy legislation requirements?
- How to maintain data quality and integrity?
- How to manage and control data access?
- How to ensure cross NatCo use case?

TRANSPARENCY LEADS TO TRUST



MARKET RESEARCH ON MULTIPLE LEVELS

> 20 STUDIES
3 SURVEYS
2.000 RESPONDENTS
10 SPRINTS
31 PILOT USERS

HIGH ACCEPTANCE WITH CUSTOMERS

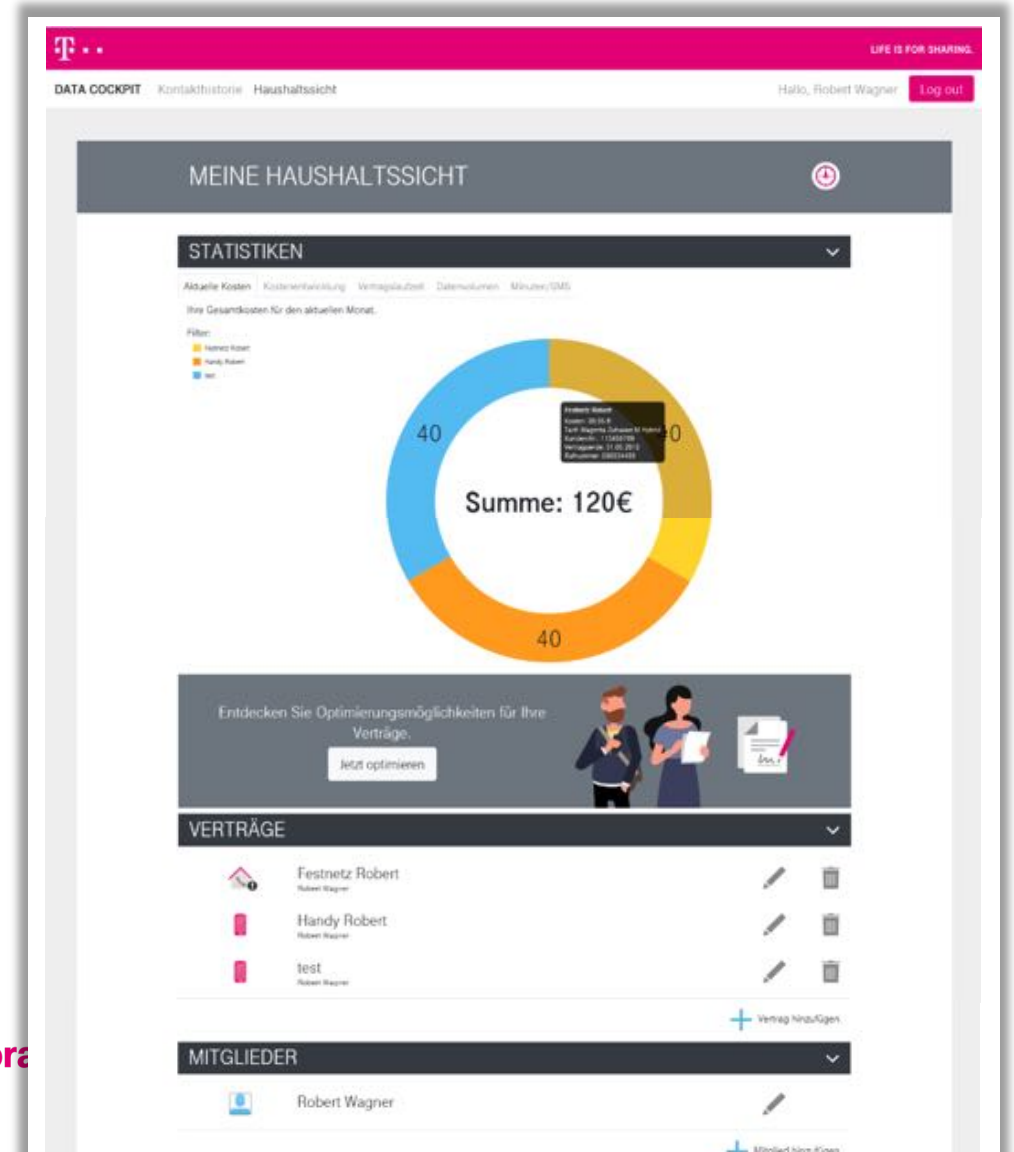
"Data Cockpit would foster my brand loyalty, being a reason to pay more for the service."

"Data Cockpit which informs on such a sensitive topic is highly relevant & it strengthens my trust."

"Data Cockpit is clear, concise and honest."

Source: UDI Workshop Sessions - Customer Sprint Club, Okt/Nov 2016, 31 participants

CUSTOMER BENEFITS LEAD TO WILLINGNESS TO SHARE



EXAMPLE: CUSTOMER SERVICE

OVERVIEW & EXAMPLES OF THREAT SCENARIOS
BIAS

Chatbot became racist

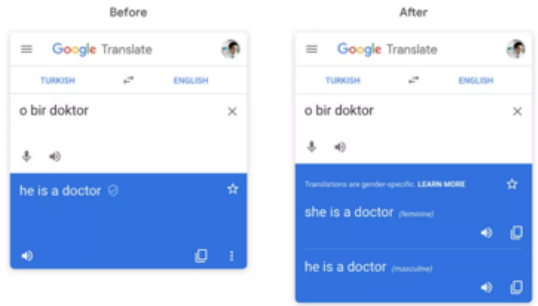
Microsoft

- Tay, an artificially intelligent chatbot with the personality of a flippant 19-year-old, was released in 2016
- The goal was to train the bot by letting users interact with it through social media channels
- Users soon figured out how to make Tay say awful and racist things and Microsoft took it offline in less than a day

Gender bias in translation

Google

- As Google Translate learns from content that is already on the web, it tends to reproduce gender-based assumptions in language
- The classic example in language is that a doctor is perceived as male and a nurse is female
- If these biases exist in a language then a translation model will learn it and amplify it



Facial recognition fail

Nikon

- A Nikon camera asked its Asian users if someone blinked in the photo – but no one did
- A algorithm can be trained to look for common features in faces, or more specifically, their shadows

Sources: xxx

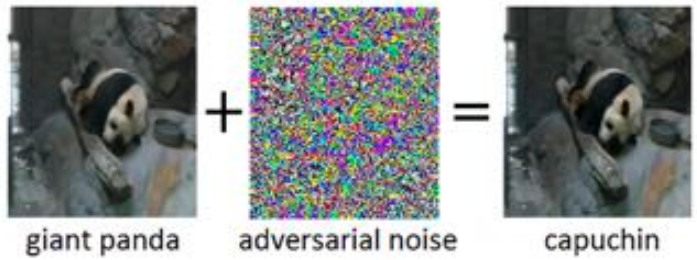
OVERVIEW & EXAMPLES OF THREAT SCENARIOS ATTACKS

CEO Fraud with fake voice

- Lyrebird's voice imitation software has made a fraud of 220,000 euros possible
- The managing director of a British energy company, believing his boss was on the phone, followed orders to wire money to an account in Hungary
- The AI software can learn the voice of a person within a few minutes and then imitate it

Image manipulation

- The addition of a small amount of adversarial noise to the image of a giant panda leads the DNN to misclassify this image as a capuchin
- The added noise in the adversarial example is imperceptible to a human
- Often, the target is misclassification or a specific incorrect prediction which would benefit an attacker



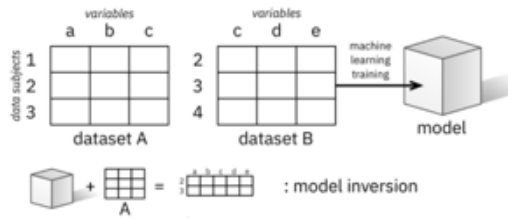
Poisoning auto-complete

- An adversary employs a Sybil attack to poison a web browser's auto-complete function
- It suggests the word "fraud" at the end of an auto-completed sentence with a target company name in it
- Sybil attacks use multiple 'sock puppet' accounts controlled by a single entity to violate the integrity of a system

OVERVIEW & EXAMPLES OF THREAT SCENARIOS PRIVACY

Inferring personal data

- Attackers have access to some personal data belonging to specific individuals included the training data
- They can infer further personal information about those same individuals by observing the inputs and outputs of the ML model
- The information attackers can learn goes beyond generic inferences about individuals with similar characteristics

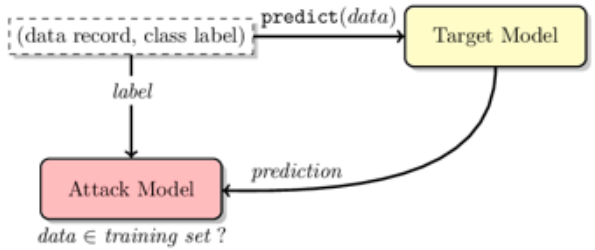


Reconstructing images of faces

- Attackers could reconstruct images of faces that a Facial Recognition Technology (FRT) system has been trained to recognise
- FRT systems are often designed to allow third parties to query the model
- When the model is given the image of a person whose face it recognises, the model returns its best guess as to the name of the person, and the associated confidence rate

Membership inference

- Membership inference attacks allow malicious actors to deduce whether a given individual was present in the training data of a ML model
- If hospital records are used to train a model which predicts when a patient will be discharged, attackers could use that model in combination with other data about a particular individual
- They can work out if the individuals were part of the training data.



Sources: xxx