„LIEBER KÜNSTLICH INTELLIGENT ALS NATÜRLICH DUMM“

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

LIFE IS FOR SHARING.
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
## Cybersecurity heads the lists of AI-related concerns

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

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


## Cybersecurity threats are giving some companies pause

Effect of cybersecurity concerns on companies

- Moved ahead with AI initiatives despite cybersecurity concerns: 36%
- Experienced a cybersecurity breach relating to AI initiatives within the last two years: 32%
- Slowed an AI initiative in order to address cybersecurity concerns: 30%
- Decided not to start an AI initiative due to cybersecurity concerns: 20%
- Canceled or halted an in-progress AI initiative due to cybersecurity concerns: 16%

Bad data used to train AI can contain implicit racial, gender, or ideological biases.
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.
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 search
  - to move
  - to transform
  - to integrate

- **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?

DATA SOURCES

- Collect
- Process
- Analyze
- Exploit

Valuable Insights
Decisions
Actions
TRANSPARENCY LEADS TO TRUST
MARKET RESEARCH ON MULTIPLE LEVELS

> 20 STUDIES
3 SURVEYS
2,000 RESPONDENTS
10 SPRINTS
31 PILOT USERS
“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

WE NEED A 360° PERSPECTIVE
OVERVIEW & EXAMPLES OF THREAT SCENARIOS

BIAS

Chatbot became racist

- 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

- 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

- 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
OVERVIEW & EXAMPLES OF THREAT SCENARIOS

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 in 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.