

Künstliche Intelligenz  
beeinflusst unser Leben:  
Was brauchen wir, um  
Entscheidungen von  
Algorithmen zu vertrauen?

Christian Westermann  
April 2019, Wien



“

Was wäre nötig, damit Sie einem selbstfahrenden Auto, einem selbststeuernden Flugzeug oder einem KI-unterstützten Diagnosegerät **uneingeschränkt** vertrauen würden?



## KI die **spüren** kann ...

Hören | Sehen | Sprechen | Fühlen



## KI die **denken** kann...

Verstehen | Assistieren | Wahrnehmen | Planen



## KI die **handeln** kann...

Körperlich | Kreativ | Kognitiv | Reaktiv



# Wenn es um KI geht, ist die Vergangenheit kein Maßstab für die Zukunft

## Schwache KI (Regelbasiert, Sprache)

Personalisierung: 76,897 mikro-genres



Regelbasierte Entscheidungsfindung



Robotics



## Schwache KI – mit Big Data (B-2-C, Suche, E-Commerce)

Deep learning – Bildverarbeitung



Schrift- und Spracherkennung

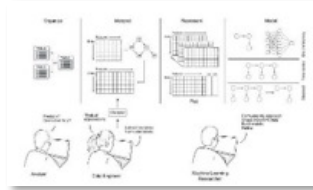


NLP und Lernprozess mit Big Data



## Demokratisierung & verkörperte KI

'Data Scientist in a Box'



Haushalts- und Serviceroboter



Selbstfahrende Autos

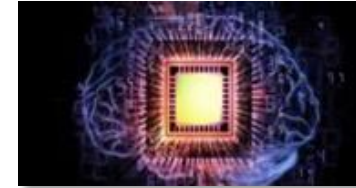


## Kollaborative KI auf neuer KI-Hardware

Kooperation Mensch-Maschine



Neuromorphe Datenverarbeitung



Schnittstelle Gehirn-Computer



## Künstliche Superintelligenz

Quantencomputer



Emotionale Roboter



90er  
Vergangenheit

00er

Gegenwart

Nächsten 5 Jahre

Nächsten 20 Jahre

Zukunft



# KI ist nicht perfekt

“

Menschen vertrauen zunächst  
Algorithmen – bis sie erleben,  
wie diese Fehler bei der  
Umsetzung ihrer Tätigkeiten  
machen.

# PwC's KI Risiko Framework

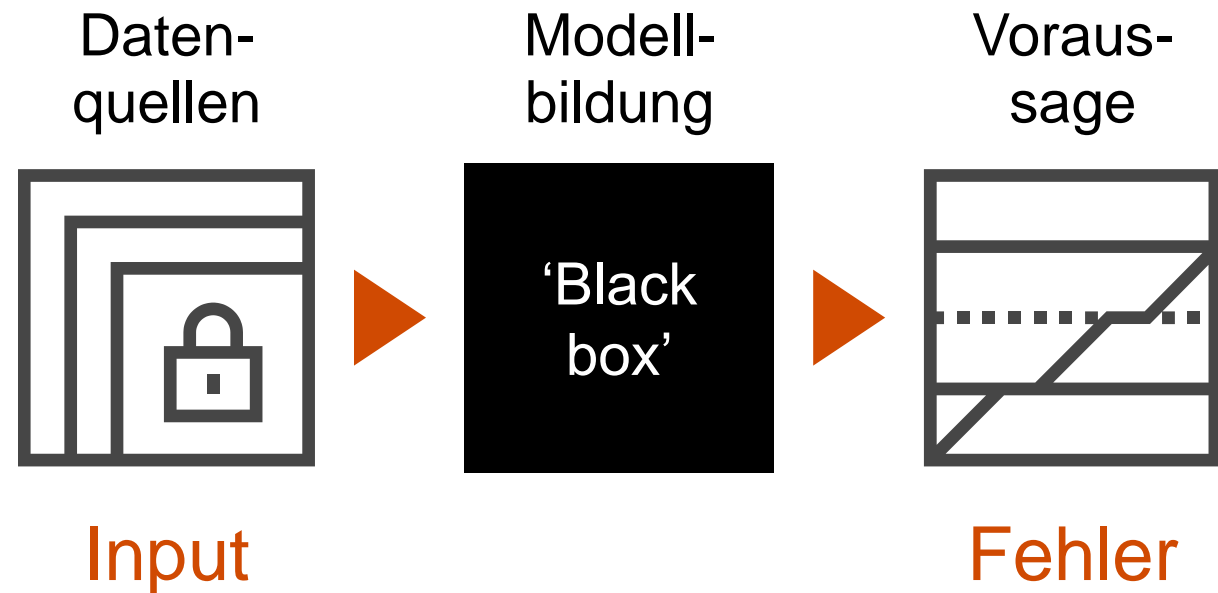


# KI Risiko #1



Performance

- Fehler
- Bias
- Mangelnde Interpretierbarkeit und Transparenz

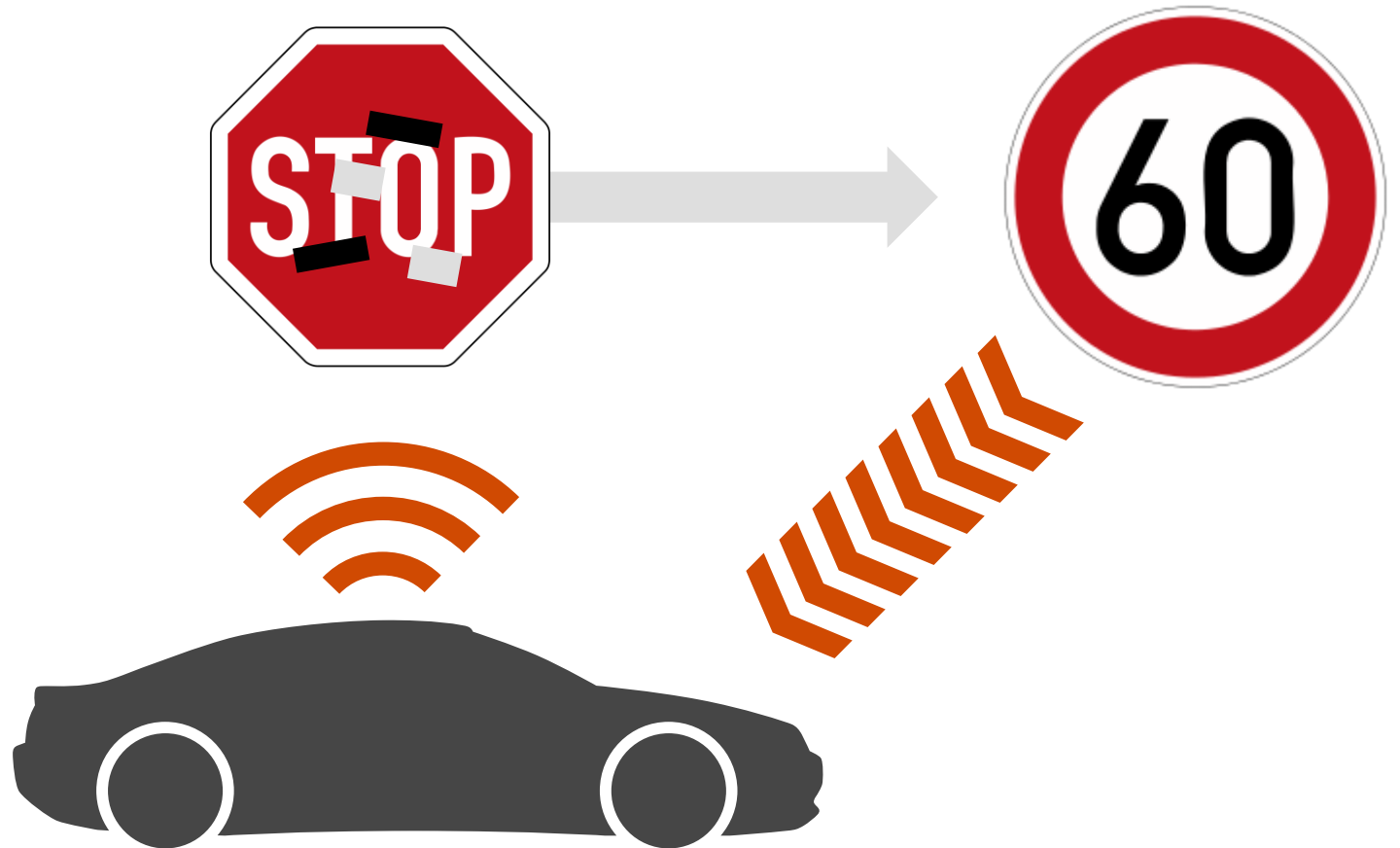


# KI Risiko #2



## Sicherheit

- Feindliche Angriffe
- Datensicherheit
- Open source





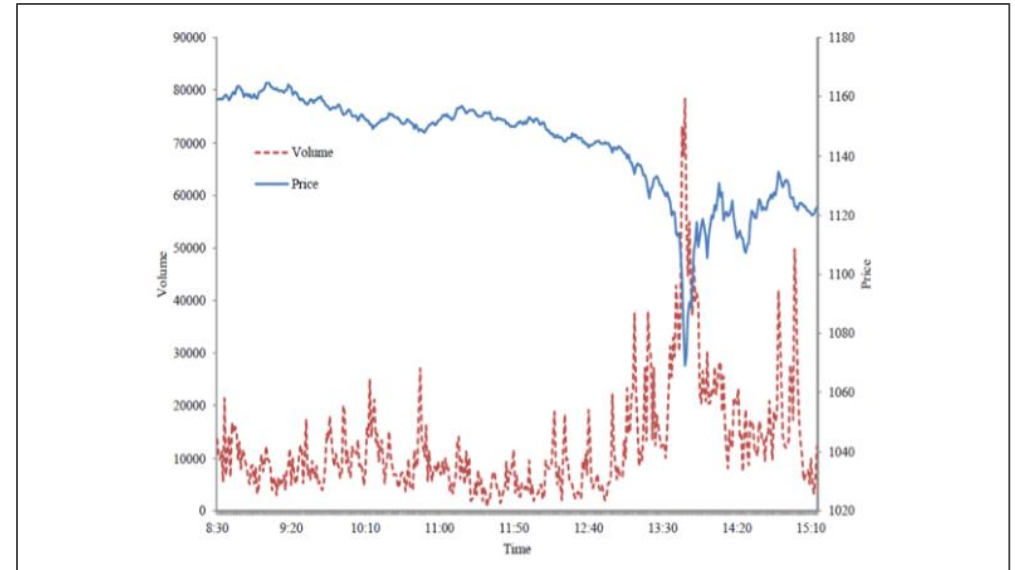
# KI Risiko #3



## Kontrolle

- KI ausser Kontrolle
- Zunehmende Komplexität
- Interdependenz

## 2010 Flash crash



## 2016 Tay/Microsoft



## 2018 Tesla



# KI Risiko #4

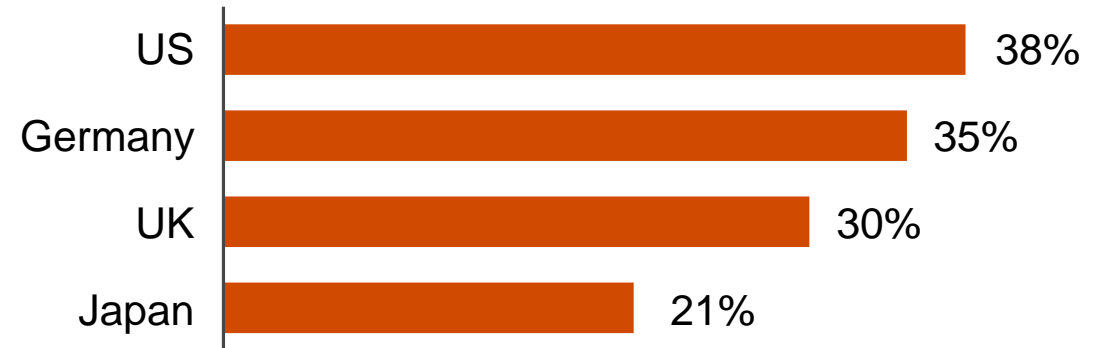


## Wirtschaftlich

- Verlust von Arbeitsplätzen
- The-Winner-takes-it-all Mentalität

## Gefährdete Arbeitsplätze

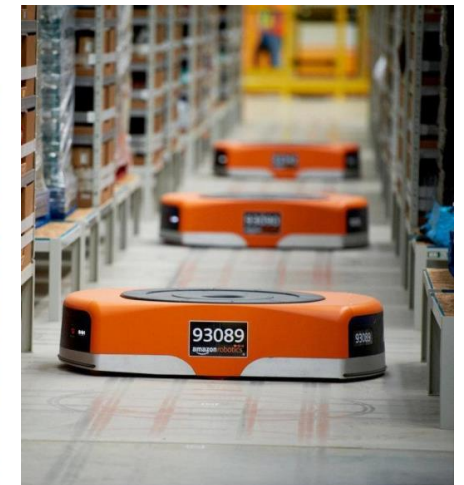
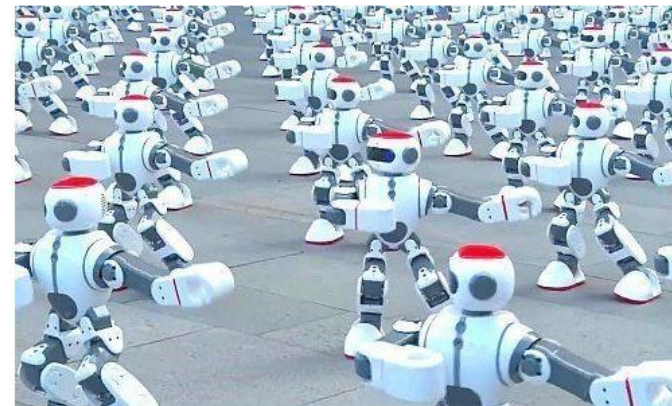
Percentage of jobs at high risk of automation by early 2030s  
Proportion of jobs



## Roboter Rüstungswettlauf

### UBS und CS: Wettrüsten der Roboter

Montag, 18. Februar 2019 07:23



# KI Risiko #5



## Gesellschaftlich

- Autonome Waffen
- Informationskluft

Autonome  
Waffen



Informations-  
kluft



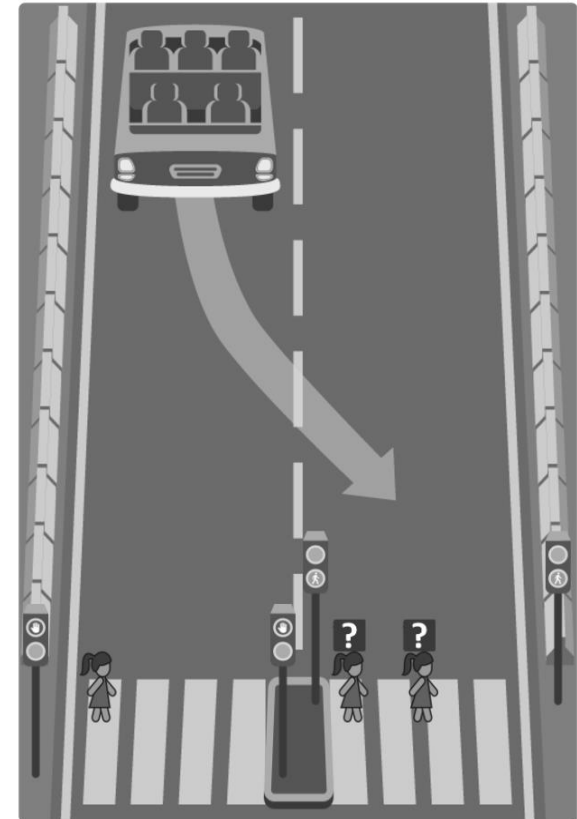
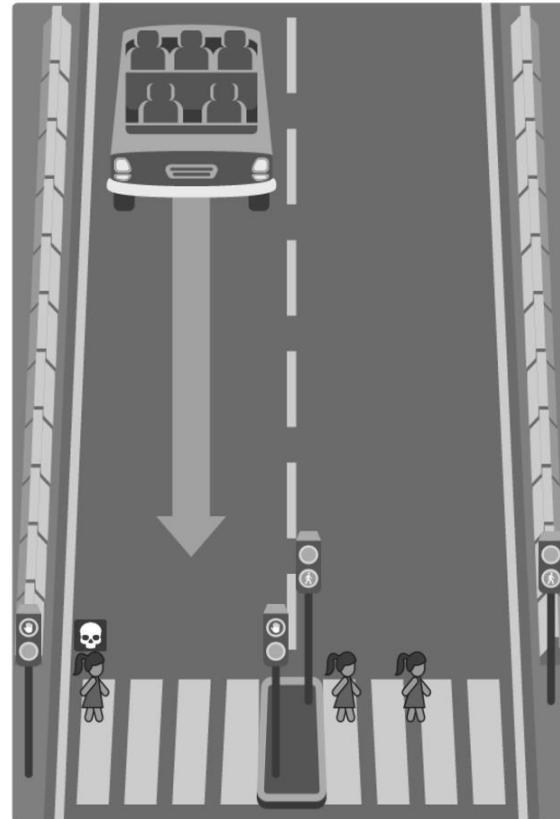
# KI Risiko #6



## Ethisch

- Wertemangel
- Werteanpassung

*"KI-Systeme sollten so konzipiert sein, dass ihre Ziele und Verhaltensweisen während des gesamten Betriebs mit menschlichen Werten übereinstimmen."*

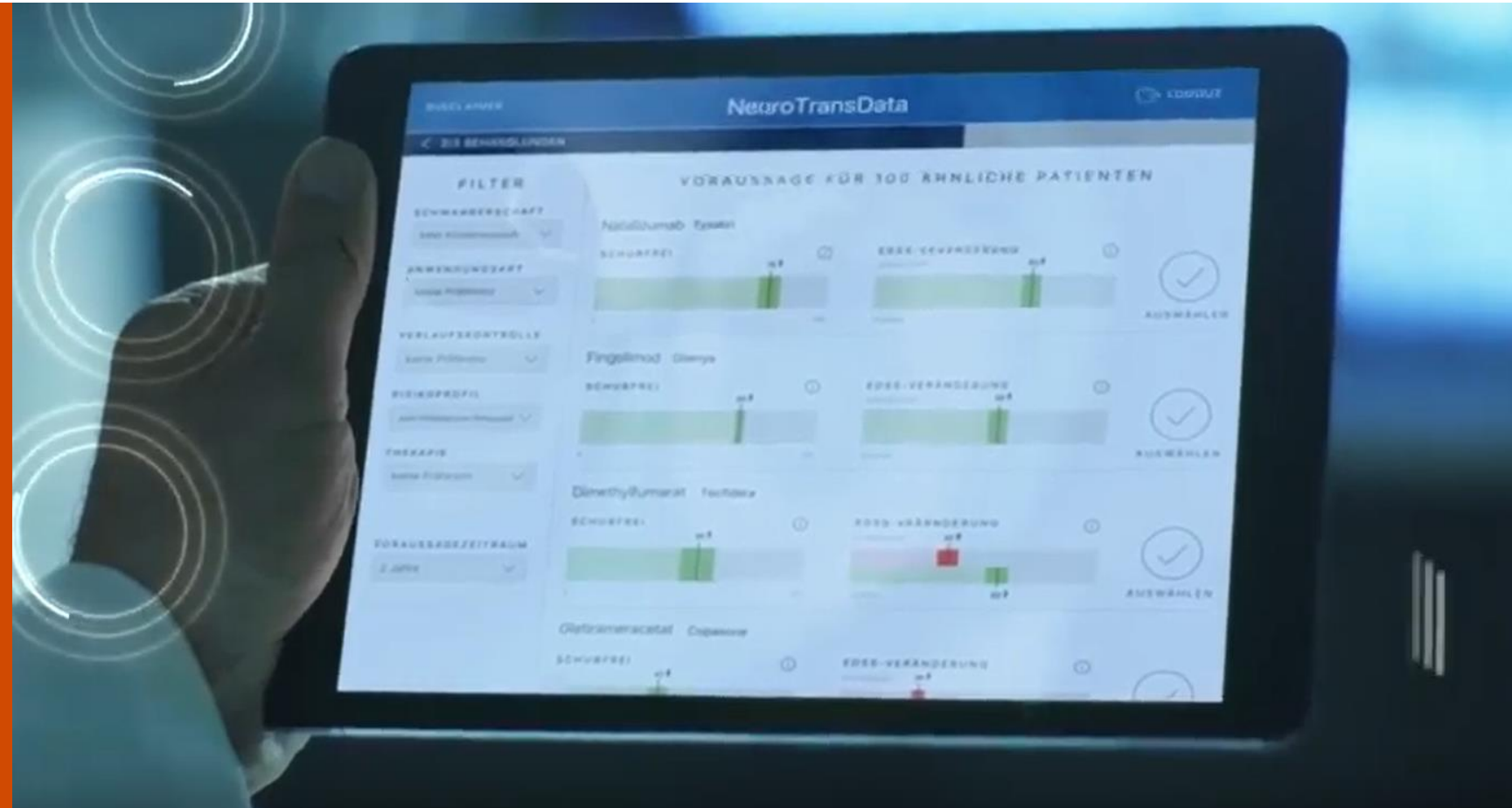


# Fallstudie– PHREND

Ein Entscheidungsunterstützungssystem für Ärzte, um Patienten mit personalisierter medizinischer Versorgung zu behandeln.

Verbessern **komplexe** oder **einfache** **Benutzeroberflächen** die Akzeptanz der KI?

Was sollte die KI erklären, wenn sich Ärzte in 80% der Entscheidungen auf ihre **eigenen Erfahrungen** stützen und verlassen?



# Fallstudie– Risikobewertung durch eine Blackbox

Ein prädiktiver Algorithmus zur Einstufung von Schifffahrtsrisiken

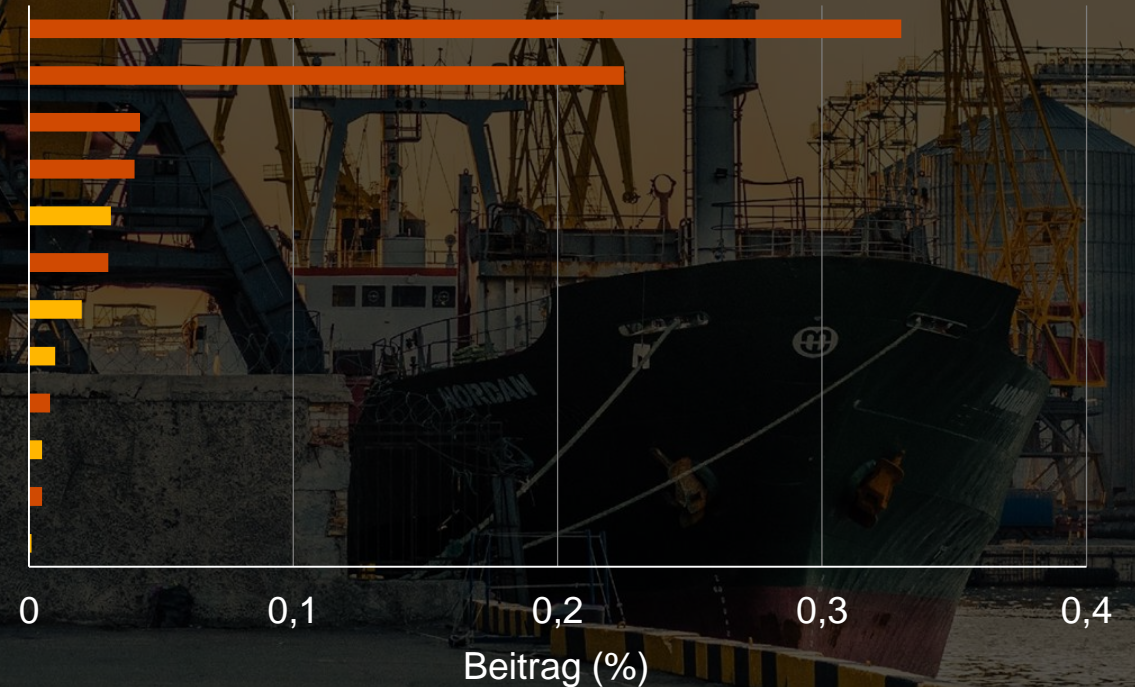
Wie können Stakeholder einem Risikobewertungsmodell vertrauen, wenn die Logik dahinter in einer Blackbox verborgen ist?

Was wäre nötig, um das Rating als fair zu beurteilen?

## Input-Variablen und deren Beitrag (wie angetroffen)

Input-Variablen

Frequently Traded Zone  
Ship Type  
Missing Class Risk  
Age  
CT Weighted  
Builder Risk  
Flag Risk  
Changes  
DWT Log  
Class Risk  
Distance Average Log  
DOC Risk Casualty



# Fallstudie– Risikobewertung durch eine Blackbox

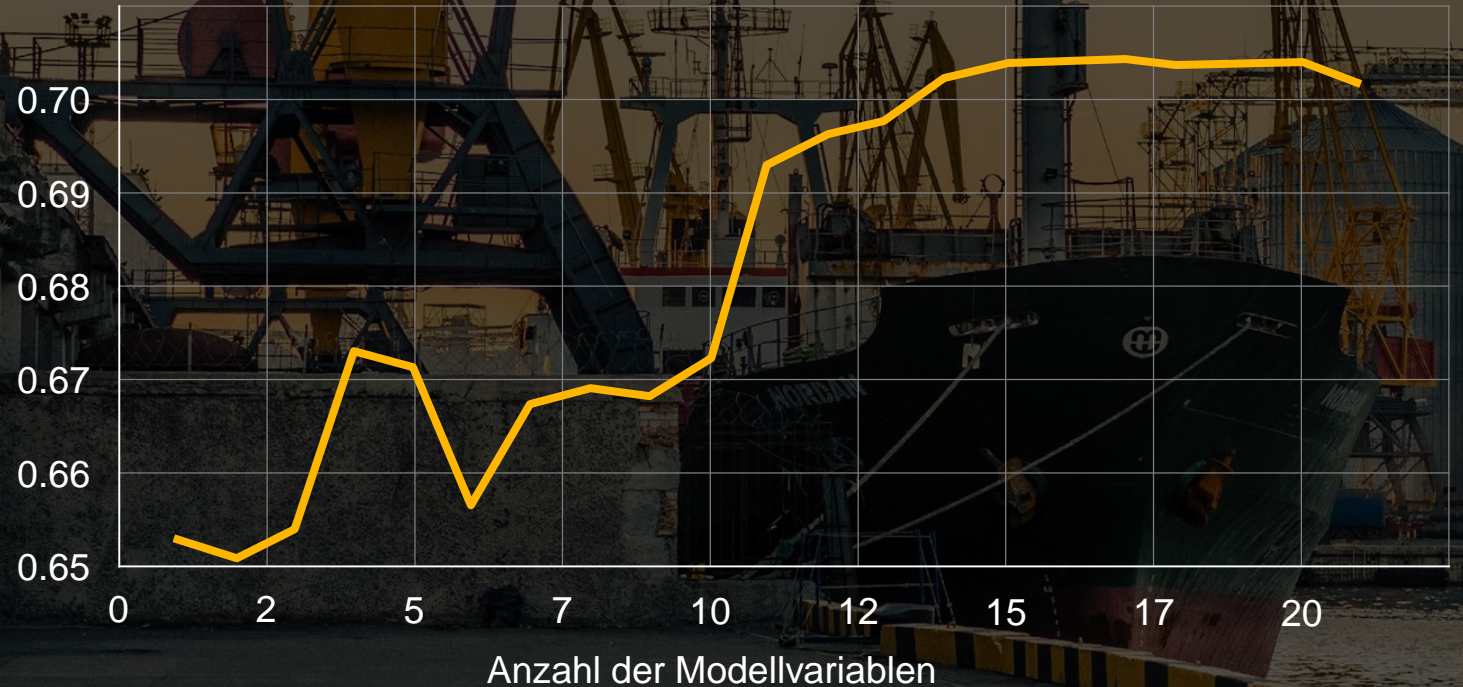
Ein prädiktiver Algorithmus zur Einstufung von Schifffahrtsrisiken

Wie können Stakeholder einem Risikobewertungsmodell vertrauen, wenn die Logik dahinter in einer Blackbox verborgen ist?

Was wäre nötig, um das Rating als fair zu beurteilen?

## Ergebnis der Vergleichsprüfung

Anzahl der korrekten Klassifizierungen



# Fallstudie– Risikobewertung durch eine Blackbox

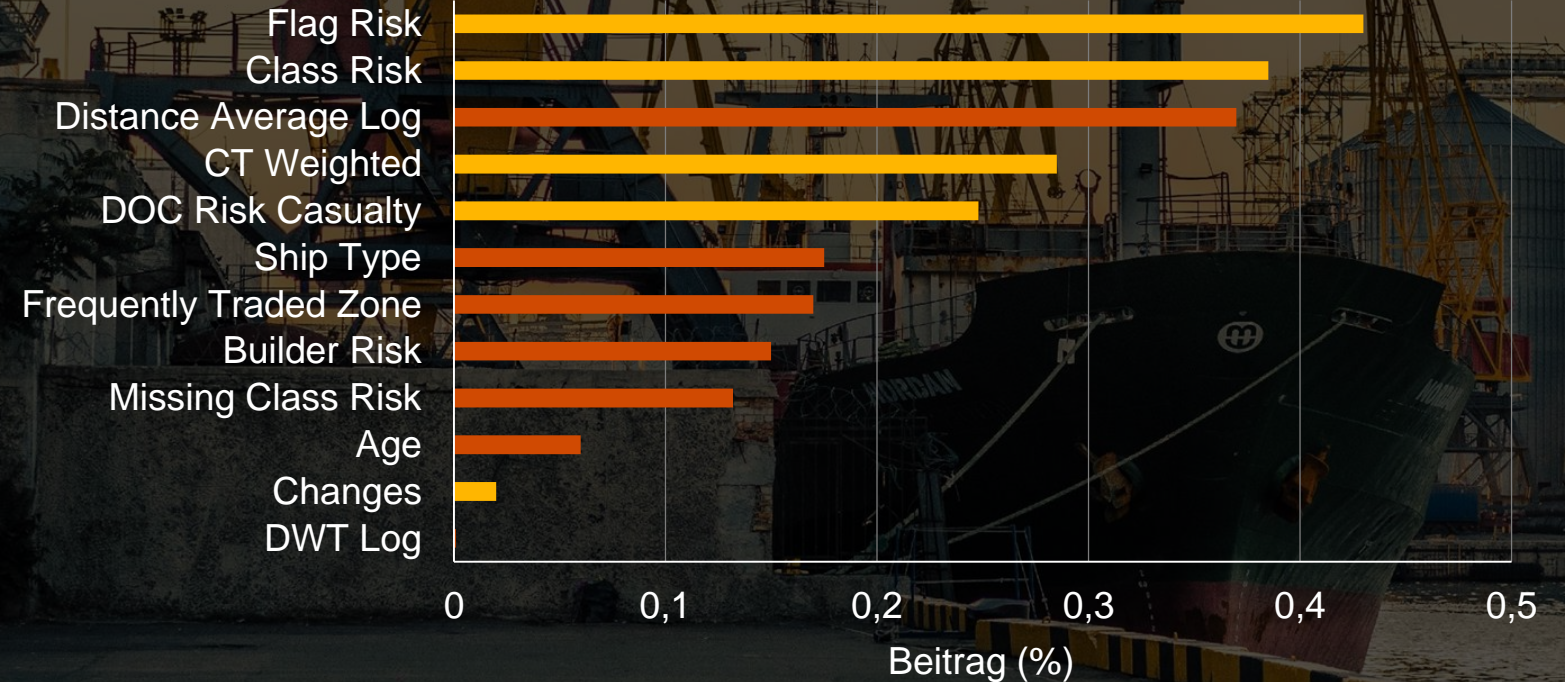
Ein prädiktiver Algorithmus zur Einstufung von Schifffahrtsrisiken

Wie können Stakeholder einem Risikobewertungsmodell vertrauen, wenn die Logik dahinter in einer Blackbox verborgen ist?

Was wäre nötig, um das Rating als fair zu beurteilen?

## Input-Variablen und ihr Beitrag (ausgewählt aus über 300 Alternativen)

Input-Variablen







Mensch und Maschine vs.  
Mensch oder Maschine

“

Die kombinierte  
Leistung von Mensch  
und Maschine ist besser  
als die der beiden  
alleine.

# Vielen Dank

[pwc.ch](https://www.pwc.ch)

© 2019 PwC. All rights reserved. Not for further distribution without the permission of PwC. “PwC” refers to the network of member firms of PricewaterhouseCoopers International Limited (PwCIL), or, as the context requires, individual member firms of the PwC network. Each member firm is a separate legal entity and does not act as agent of PwCIL or any other member firm. PwCIL does not provide any services to clients. PwCIL is not responsible or liable for the acts or omissions of any of its member firms nor can it control the exercise of their professional judgment or bind them in any way. No member firm is responsible or liable for the acts or omissions of any other member firm nor can it control the exercise of another member firm’s professional judgment or bind another member firm or PwCIL in any way.