Christian Spindler
06. September 2018
3rd COST Conference, ZHAW

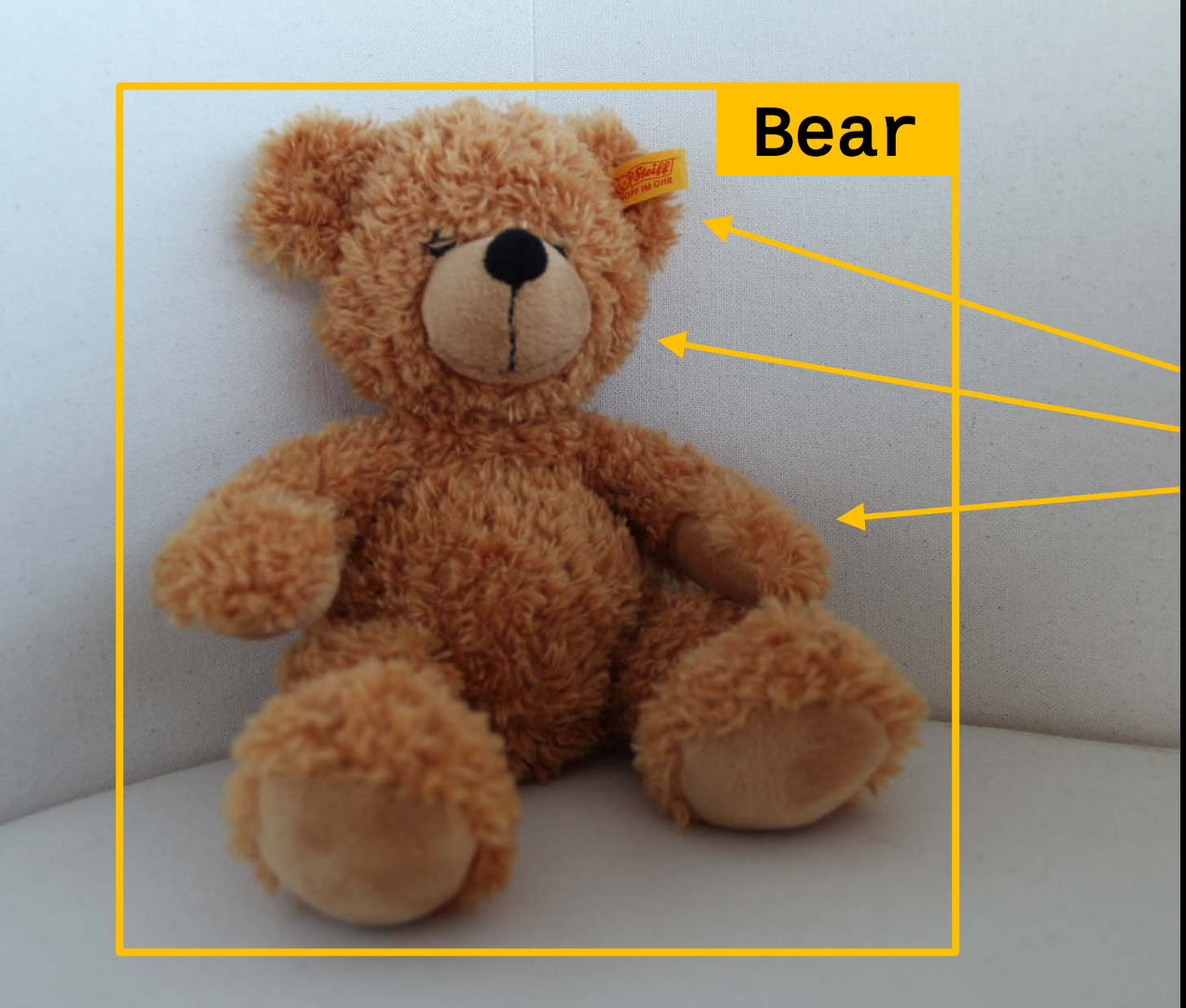
Trust in AI
Explainability and compliance

pwc

www.pwc.com/digital

Classifier:

90 %: bear
 5 %: ape
 2 %: cat



Classifier:

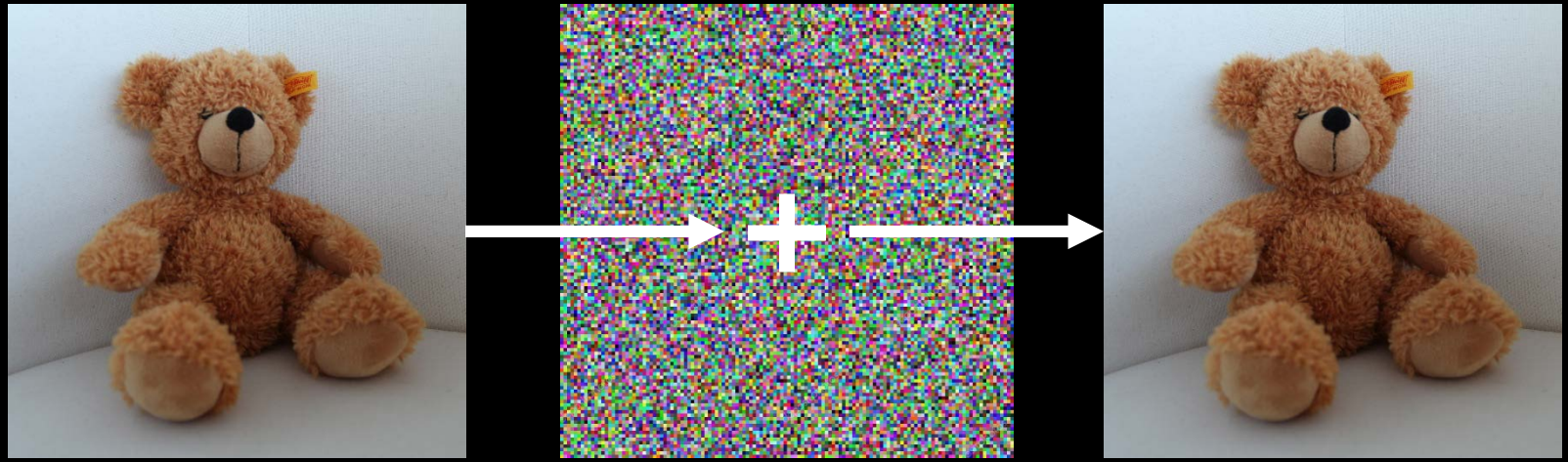
“Bear” because:

round ears
brown fur
arms

Classifier:

“Bear” because:

round ears
brown fur?
arms



99 % bear

99 % car

Universal adversarial perturbations, https://arxiv.org/abs/1610.08401

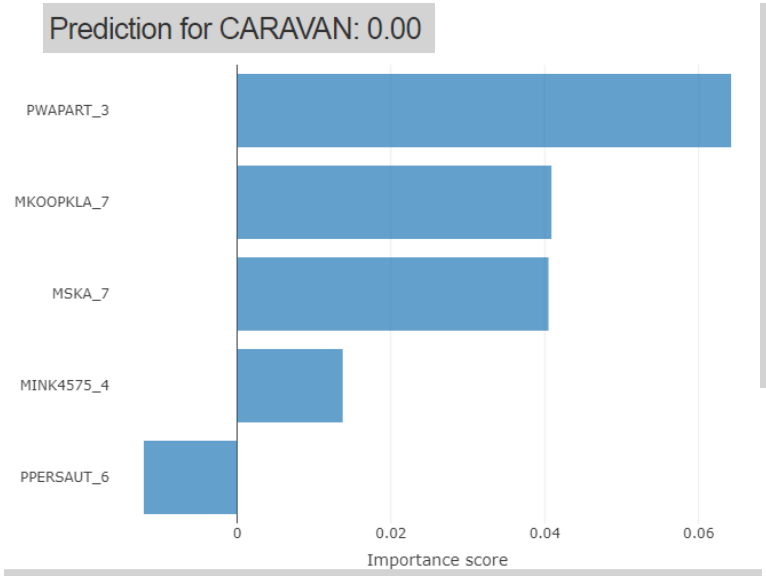
Insuring self-driving cars: Who takes responsibility in case of an accidents?

Focus today:
Building trust in the AI value chain

Good performance of explanation frameworks for structured data

Caravan insurance dataset

ABRAND	MAUT2	MAUT0	MAUT1	PWABEDF	MSKB1
1	0	1	8	0	1
1	1	2	7	0	2
1	0	2	7	0	5
1	0	0	9	0	2
1	2	1	6	0	0
0	3	3	5	0	2
0	0	1	8	0	1
0	4	2	4	0	1
0	2	3	5	0	1
1	1	2	6	0	2
0	2	1	6	0	0



Feature Explanations

Major contributions to the result:

Contribution private third party insurance is not 3

Purchasing power class is not 7

Social class A is not 7

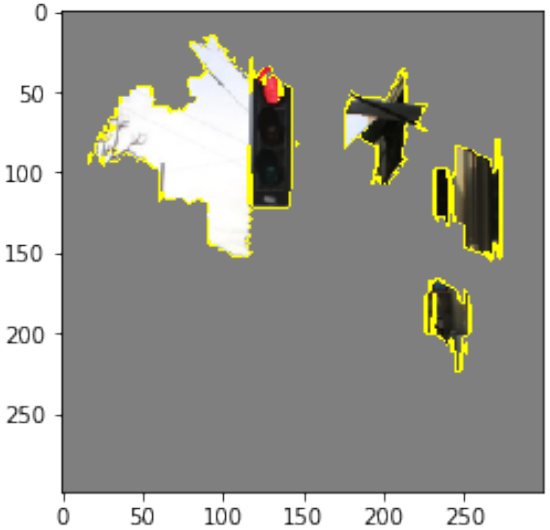
Income 45-75.000 is not 4

Contribution car policies is 6

The Insurance Company Benchmark, http://liacs.leidenuniv.nl/~puttenpwhvander/library/cc2000/

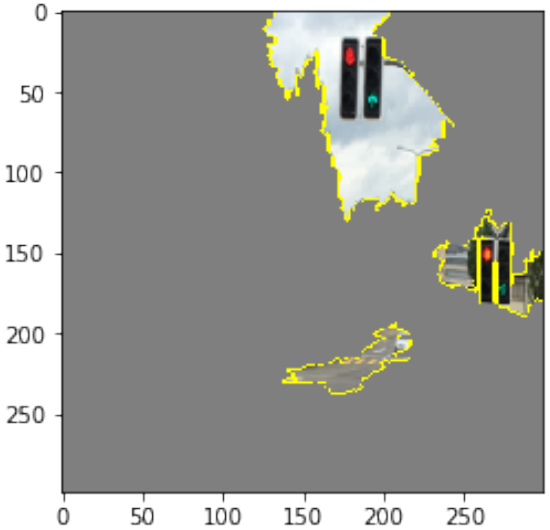
Mixed performance of benchmark vision nets on traffic relevant scenarios

```
'traffic_light':
0.998

'walking_stick':
0.00019

'maillot':
6.2e-5

'street_sign':
5.7e-5

'trolleybus':
3.0e-5
```



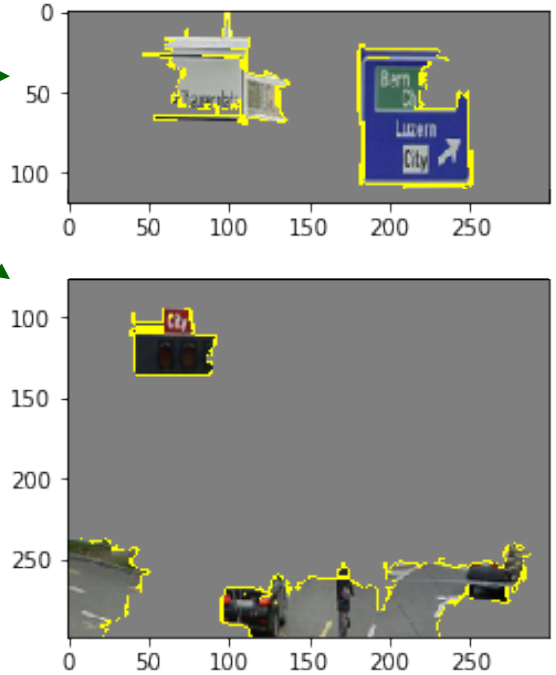
Mixed performance of benchmark vision nets on traffic relevant scenarios

```
'traffic_light':
0.999

'walking_stick':
9.4e-5

'maillot':
8.5e-5

'streetcar':
5.9e-5

'aircraft_carrier':
5.9e-5
```

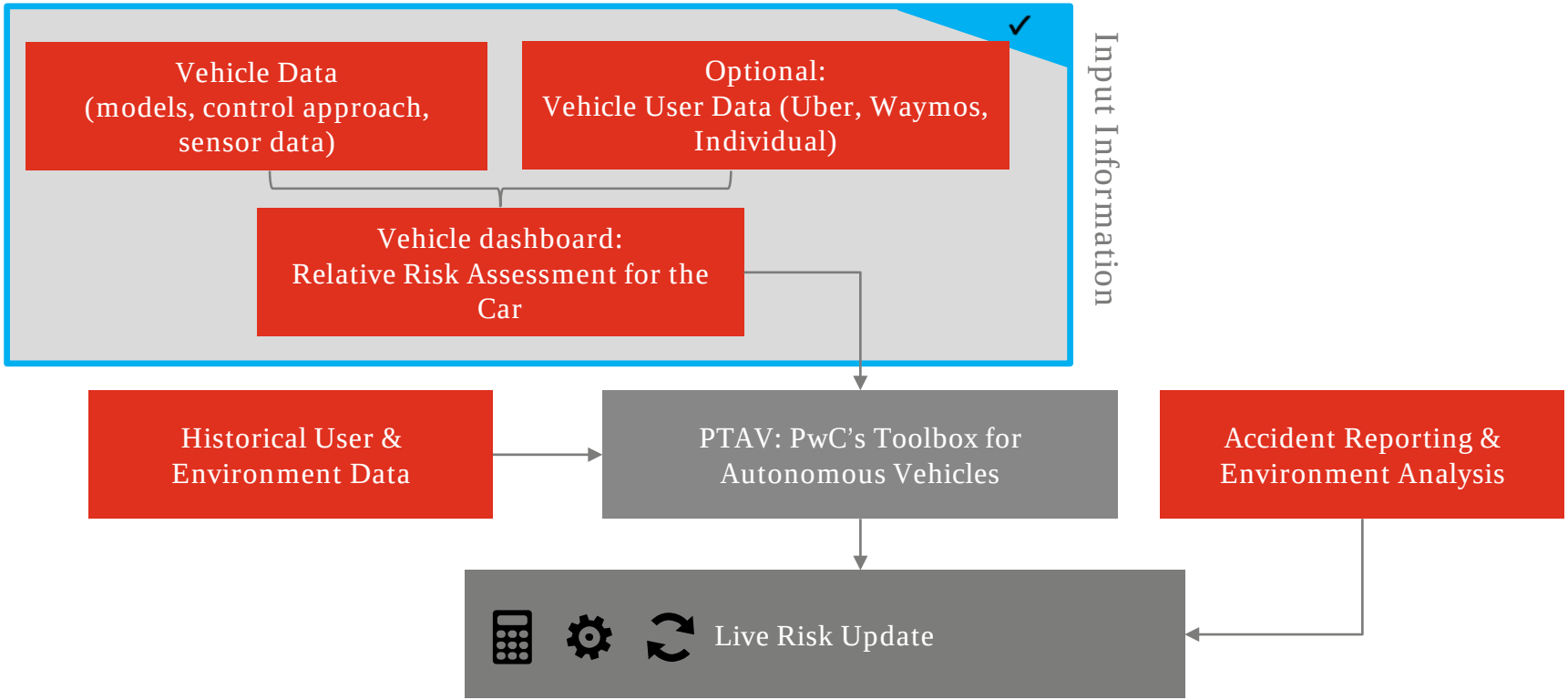


Mixed performance of benchmark vision nets on traffic relevant scenarios

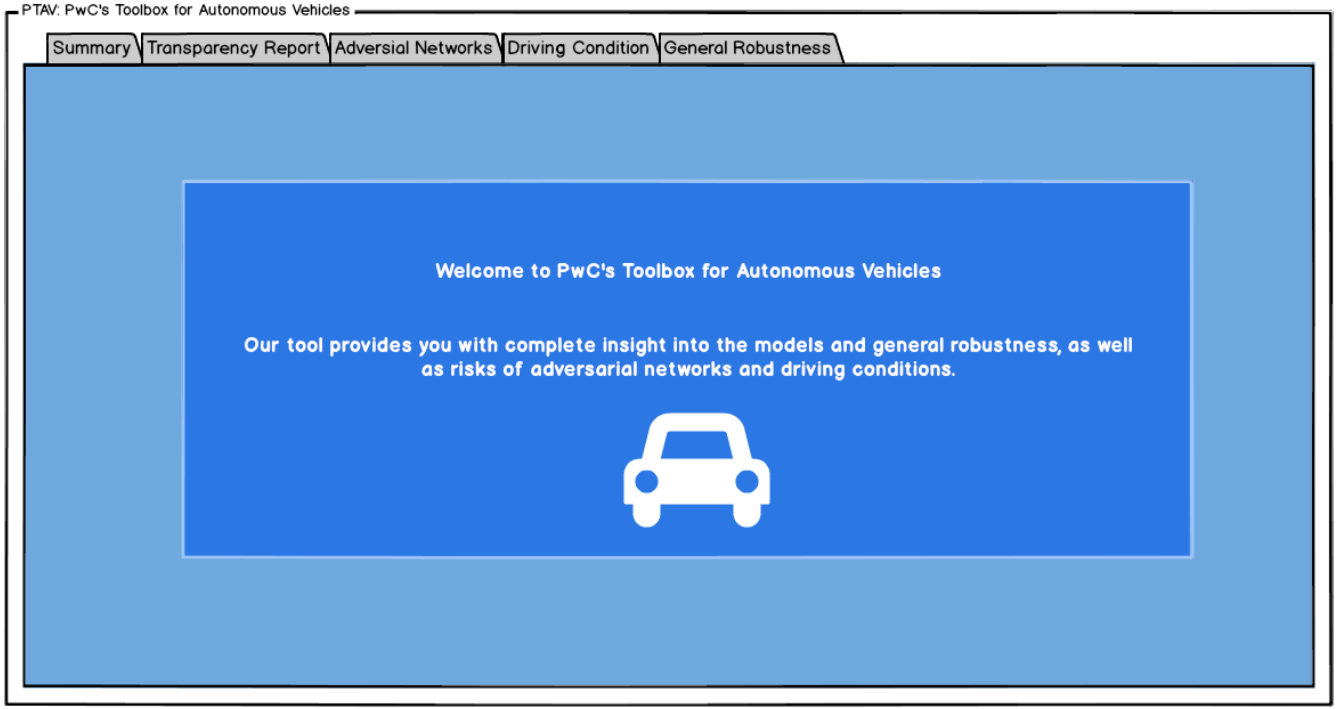

```
'street_sign':
0.80

'traffic_light':
0.17

'streetcar':
0.0019

'scoreboard':
0.0017

'racer':
0.00097
```



Insurance price (risk) estimator package



Insurance price (risk) estimator framework



PwC develops software and services for algorithmic transparency and accountability

AI Trust Builder

What:

Independent assessment of accuracy, suitability and robustness of advanced models, methods, filters and predictions from implemented AI/ML software solutions

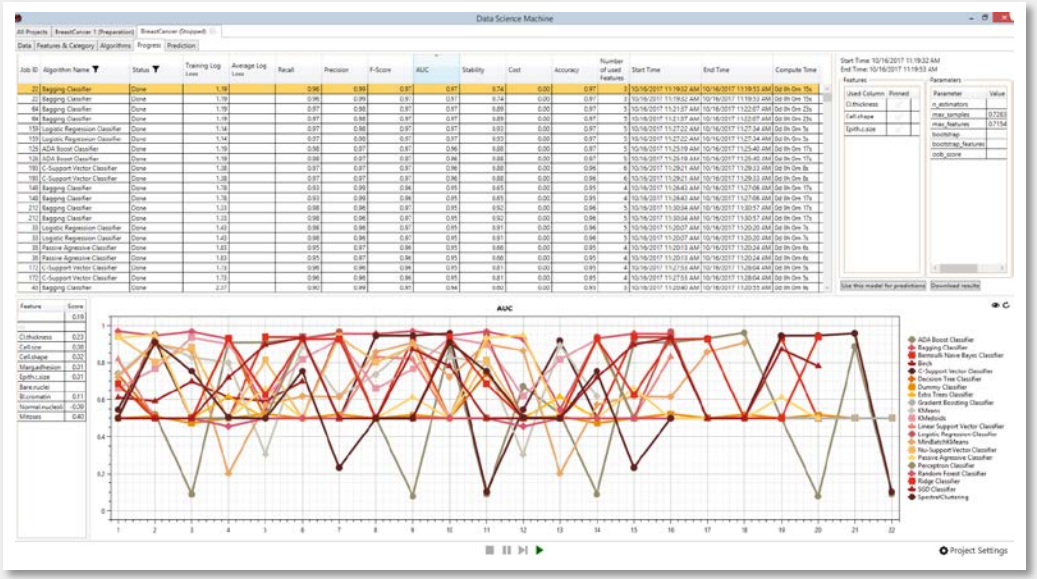
How:

- Application of automated machine-learning methods and algorithms against reference data sets
- Optimisation of best methods, filters and predictors and comparison against results from implemented AI/ML software solutions

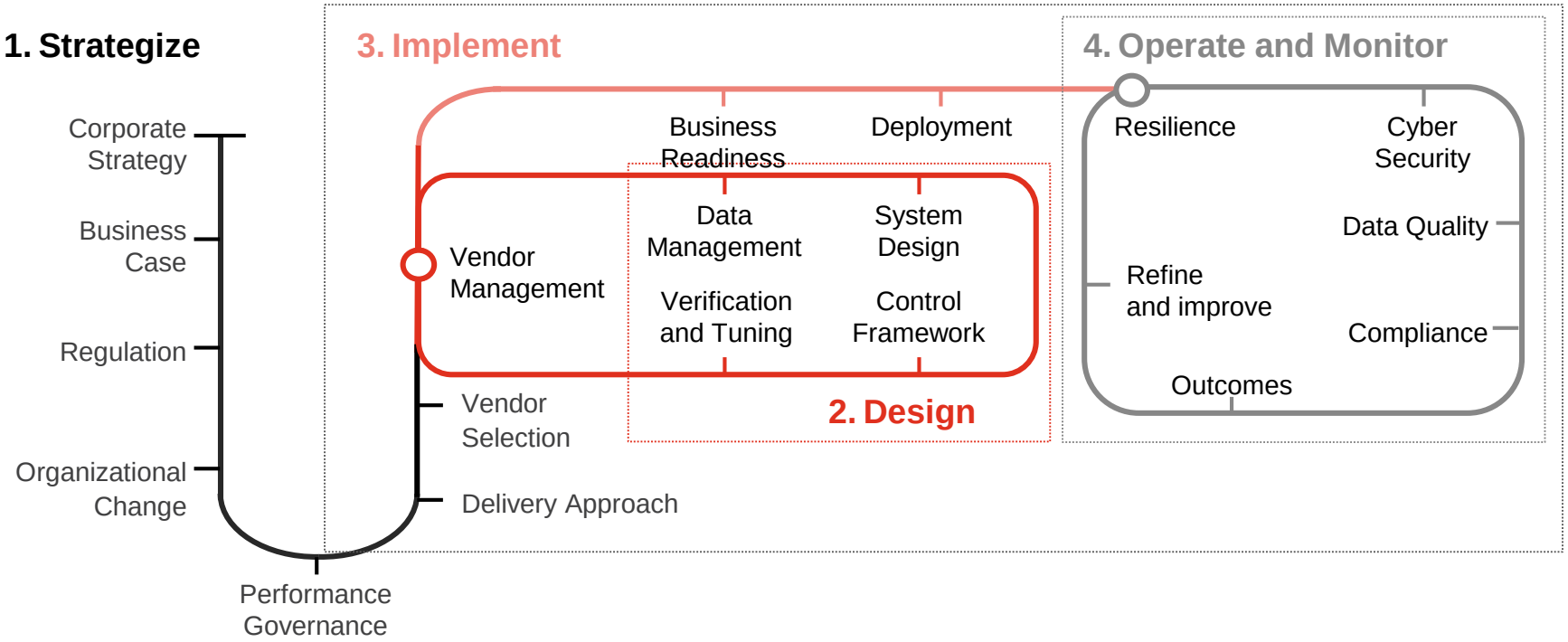
Deliverable:

- Quantified benchmark of implemented models, filters and predictors (accuracy, confidence)
- Qualified assessment of suitability and robustness of implemented models and methods

Analysing model performance



PwC's Responsible AI Framework



Creating value from responsible AI

Top line

Operational

Strategic

Build an AI strategy

By identifying trends and impacts through AI early on

Save cost

By knowing the levers in process optimization

Be compliant

By deploying non-discriminative algorithms and allow for explanation

Bottom line

Thank you!

Dr. Christian Spindler
IoT and ML lead
PwC Data & Analytics
christian.spindler@ch.pwc.com
+41 79 875 9243