Fraunhofer IKS

SAFE INTELLIGENCE

COGNITIVE SYSTEMS | ARTIFICIAL INTELLIGENCE & MACHINE LEARNING | AUTONOMOUS SYSTEMS | AUTONOMOUS DRIVING | INDUSTRY 4.0 | IOT

AI IN MEDTECH PRODUCTION

VISUAL INSPECTION FOR QUALITY ASSURANCE

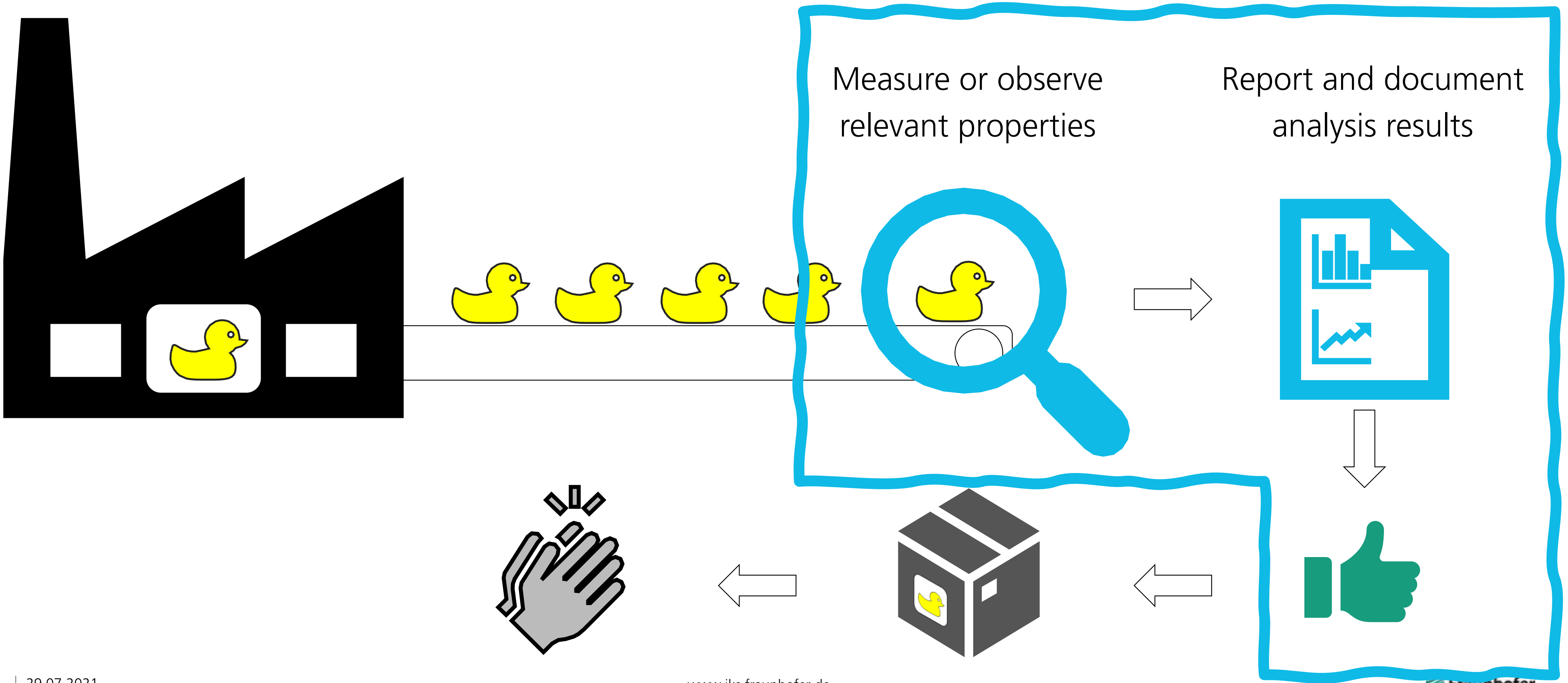
FRAUNHOFER IKS – COMPETENCY IN SAFE INTELLIGENCE

Fraunhofer Institute for Cognitive Systems IKS

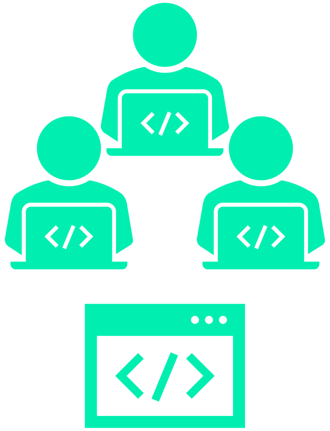
- Application-oriented research for reliable software technologies
- Expertise on cognitive systems, AI and modern software architectures in safety-critical applications
- Combining Safety and Intelligence: Safe Intelligence

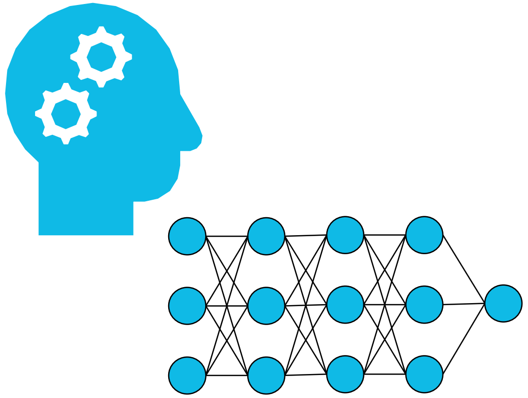
Quick facts
Founded in 2019
80 employees
3 research areas

(VISUAL) QUALITY ASSURANCE



HOW CAN QUALITY RELATED PROPERTIES BE CHECKED?





Manual Inspection

+ reliable, accepted
- slow, expensive

Software Algorithm

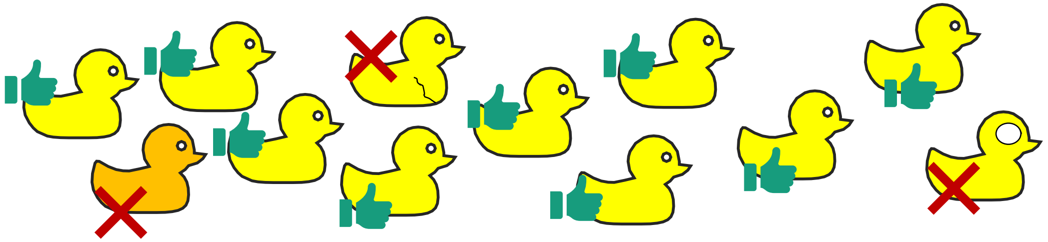
+ fast, testable
- difficult and expensive to develop

Machine Learning

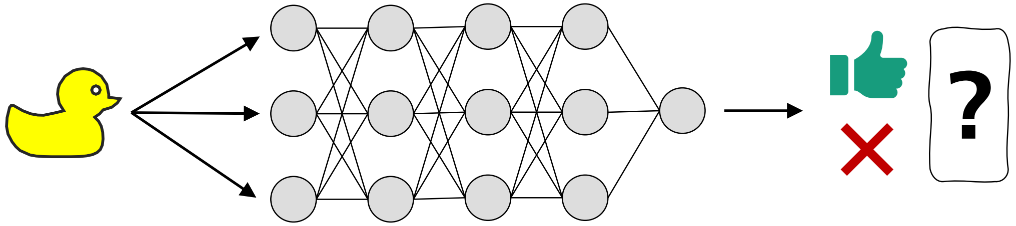
+ fast, easy to train, hip
- *nothing?*

MACHINE LEARNING: BASICS

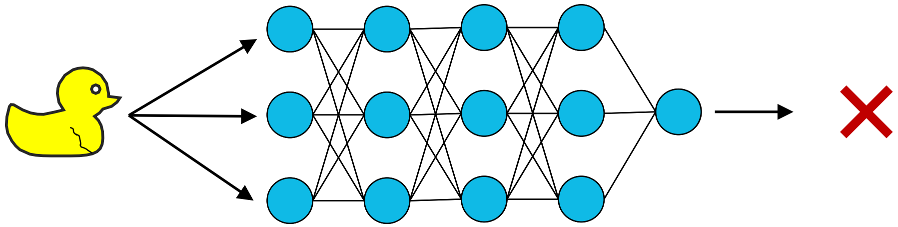
Examples

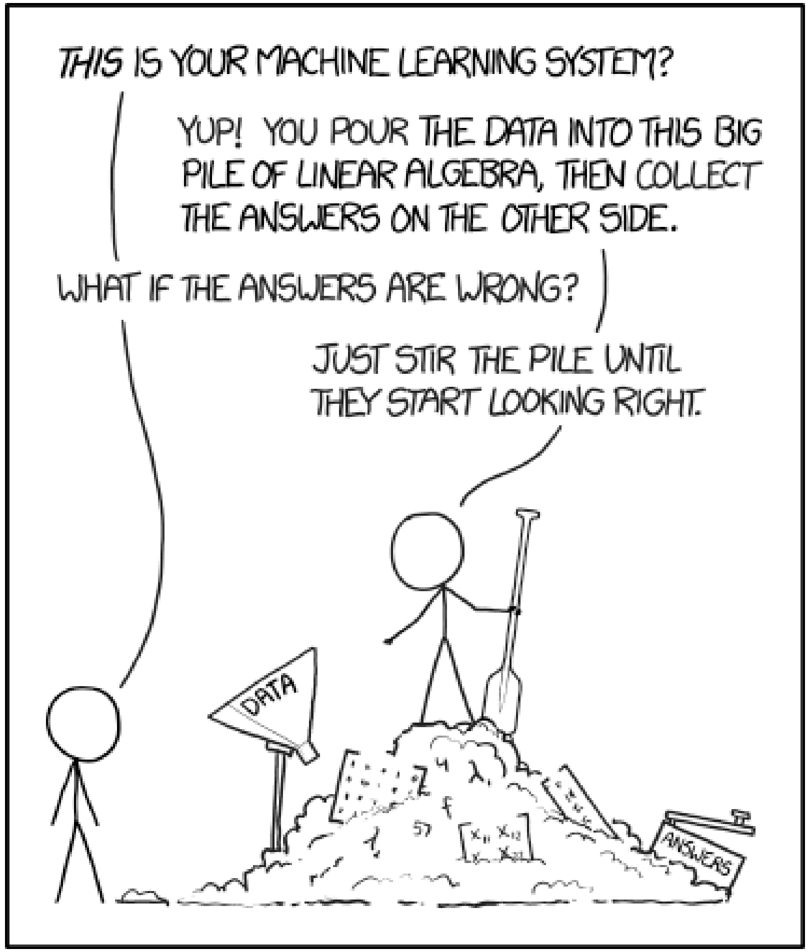


Model



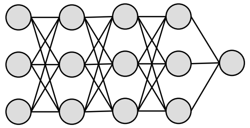
Training

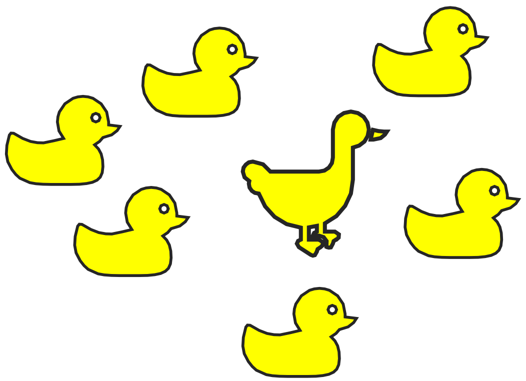


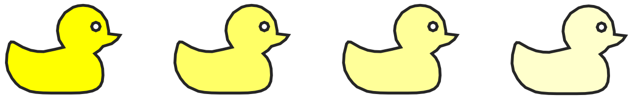


Source Image: https://xkcd.com/1838/

MACHINE LEARNING MODELS – WHAT CAN GO WRONG?







Overconfidence
Lack of Interpretability

Most modern DNNs are too confident in their own predictions, i.e. scores (like softmax) do not reflect certainty.
Complex models are opaque to the user and behave like black box functions.

Out-of-Distribution
Scalable Oversight

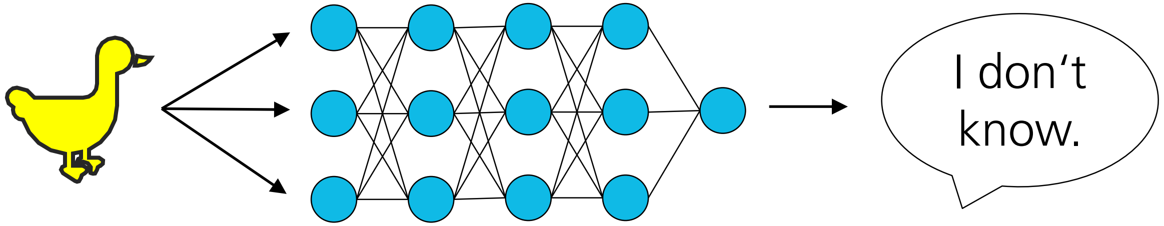
Unpredictable results for unknown and unspecified inputs.

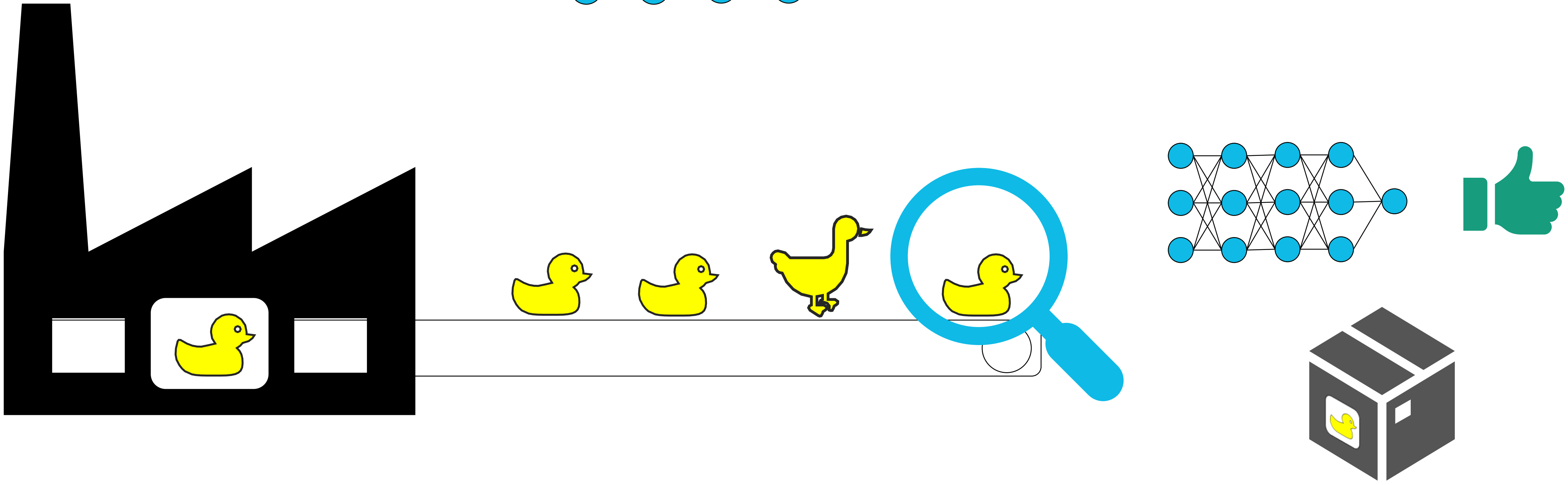
Critical situations occur rarely and are therefore inadequately trained.

Distributional Shift
Robustness

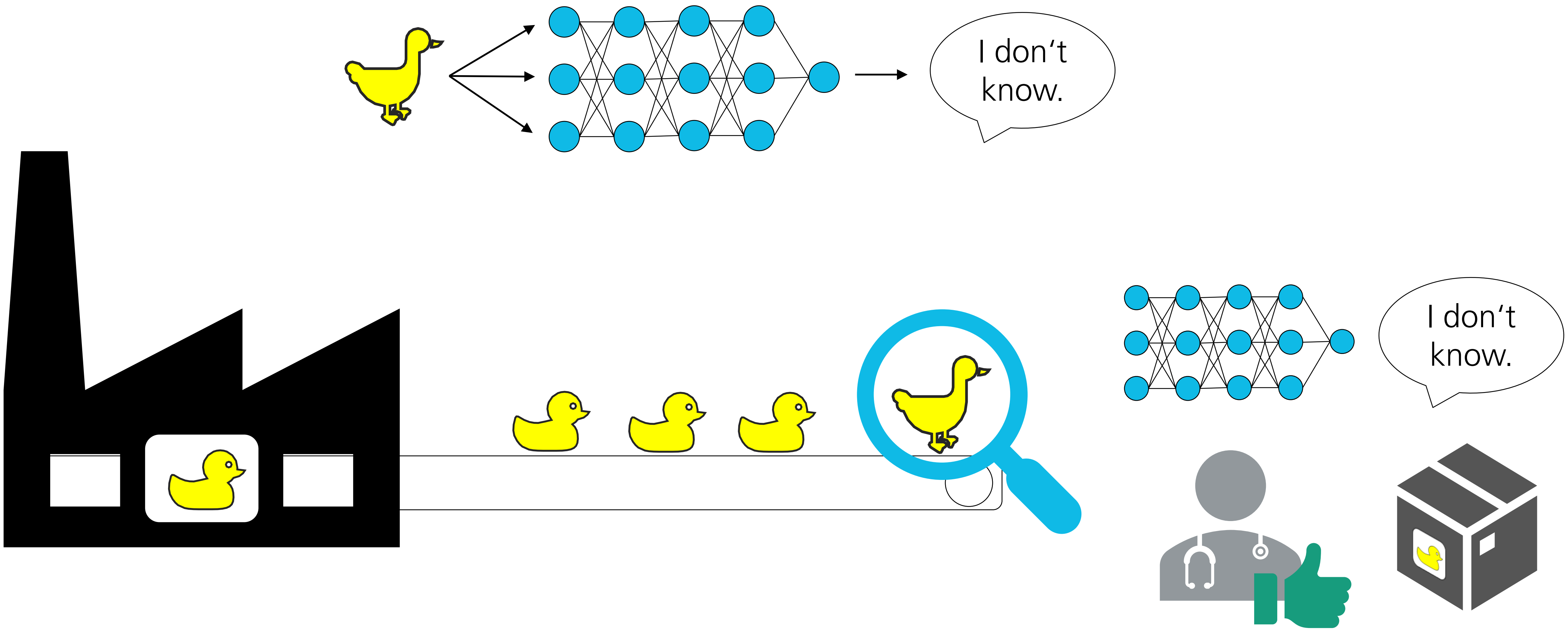
Performance sensitive to subtle changes in the environment over time
and to naturally occurring variations, e.g., due to sensor noise, lighting, etc.

UNCERTAINTY: KNOW WHEN YOU DO NOT KNOW

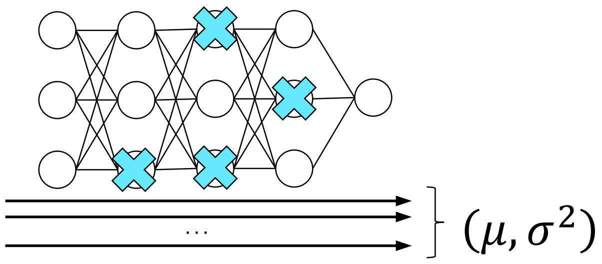




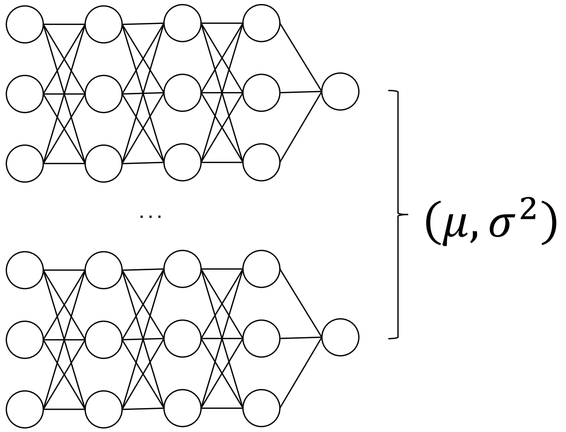
UNCERTAINTY: KNOW WHEN YOU DO NOT KNOW



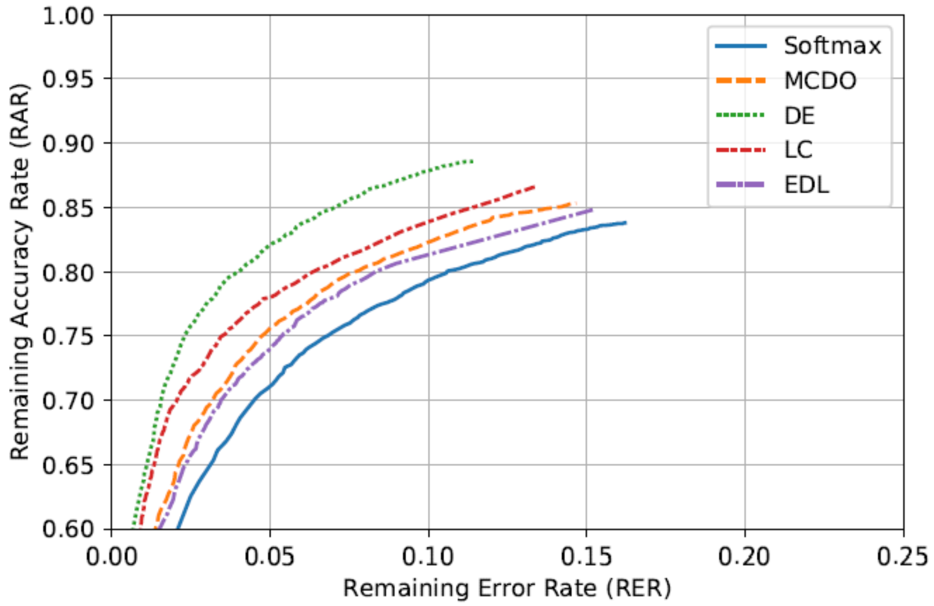
UNCERTAIN UNCERTAINTY



Monte Carlo Dropout (MCDO)
(Gal et al. 2016)



Deep Ensembles (DE)
(Lakshminarayanan et al. 2017)

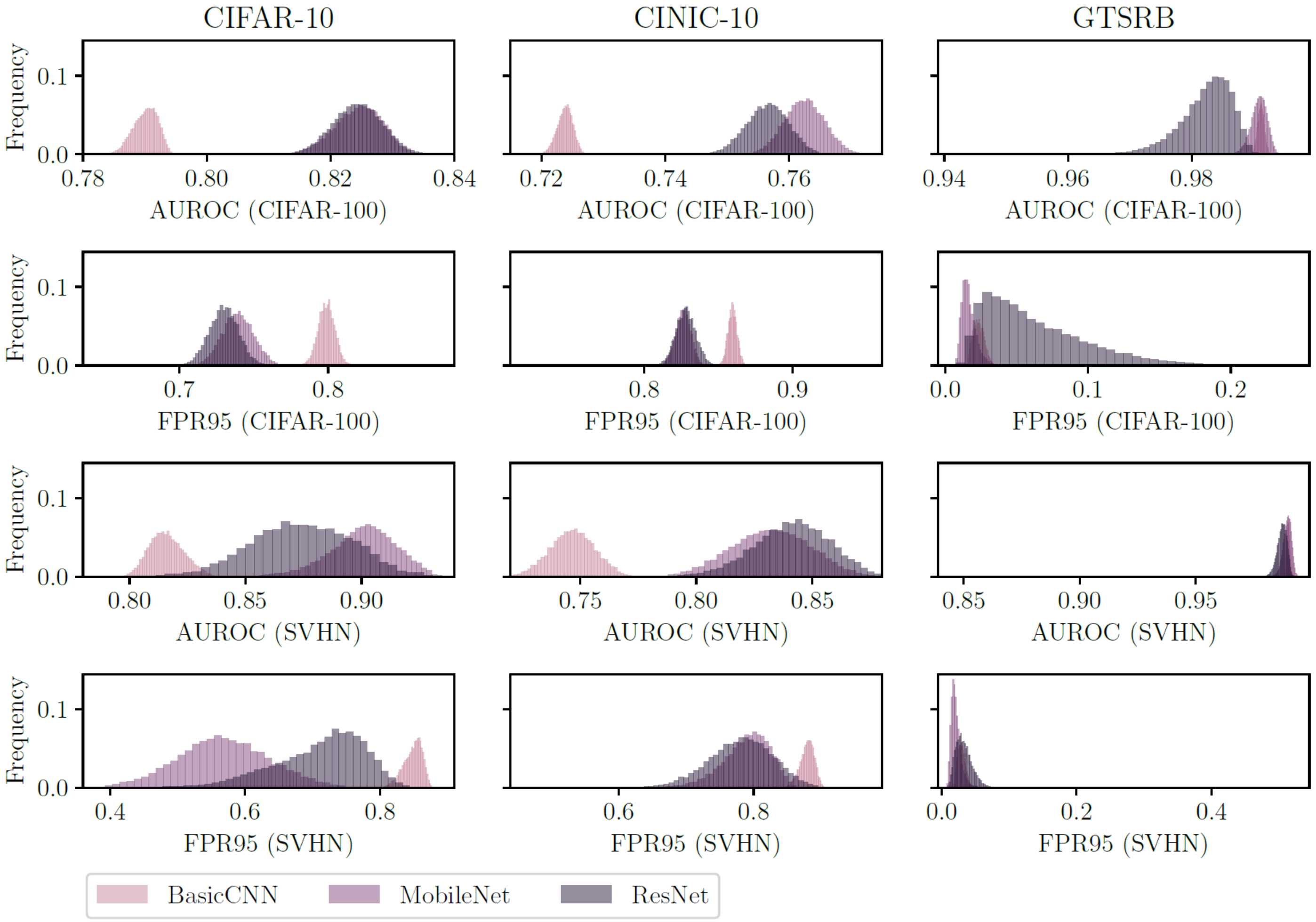


DEEP DIVE: ENSEMBLE DIVERSITY

		CIFAR-10			CINIC-10			GTSRB		
		BasicCNN	MobileNet	ResNet	BasicCNN	MobileNet	ResNet	BasicCNN	MobileNet	ResNet
None	Acc	81.75	85.98	**86.22**	70.64	77.09	77.56	98.78	98.39	98.20
		83.52	88.07	**88.35**	71.54	79.00	79.47	99.23	99.59	99.42
	Acc@1%	47.91	**54.50**	55.41	28.79	36.59	38.33	98.71	97.95	97.84
		53.05	**62.81**	63.23	30.76	42.64	43.36	99.23	99.59	99.42
	ECE	1.04	3.66	3.02	1.26	3.55	**2.11**	0.32	1.07	0.59
		2.81	6.87	6.41	2.92	6.31	**5.81**	0.94	3.26	2.89
Brightness	Acc	77.40	79.03	80.26	66.65	71.74	72.03	96.79	95.55	95.50
		79.53	83.85	83.66	68.02	74.91	75.14	97.71	97.60	97.55
	Acc@1%	37.78	39.60	41.29	23.05	29.43	29.25	94.40	91.84	91.35
		42.98	50.94	51.63	25.44	35.34	35.56	96.47	96.20	96.41
	ECE	1.01	3.40	2.86	1.75	4.17	2.74	0.47	2.04	0.56
		2.80	8.11	7.22	3.73	7.84	6.49	1.14	5.87	4.43
Contrast	Acc	61.23	64.02	**68.49**	50.69	61.00	64.02	97.21	94.81	95.54
		65.51	70.73	**76.62**	53.38	65.98	69.46	98.23	98.23	98.12
	Acc@1%	15.02	17.12	23.74	9.74	17.84	20.62	95.06	90.89	92.19
		21.35	27.10	35.18	12.13	22.88	27.26	97.53	97.66	97.52
	ECE	3.39	0.88	0.92	3.84	1.18	0.81	1.21	2.44	1.21
		7.97	4.48	5.45	6.49	5.68	5.88	2.32	6.34	6.32
Cutout	Acc	69.96	77.46	77.97	63.38	71.24	72.14	95.34	94.70	94.71
		73.08	81.72	82.64	64.90	74.07	75.56	96.07	96.52	96.38
	Acc@1%	25.94	**33.17**	36.61	20.41	28.38	29.37	91.27	89.68	89.64
		32.97	**49.21**	49.45	22.34	33.58	35.69	93.05	93.07	93.63
	ECE	1.56	2.28	2.11	0.49	3.39	1.59	0.44	1.37	0.68
		4.22	6.92	6.69	2.31	6.81	6.45	1.29	4.89	3.66

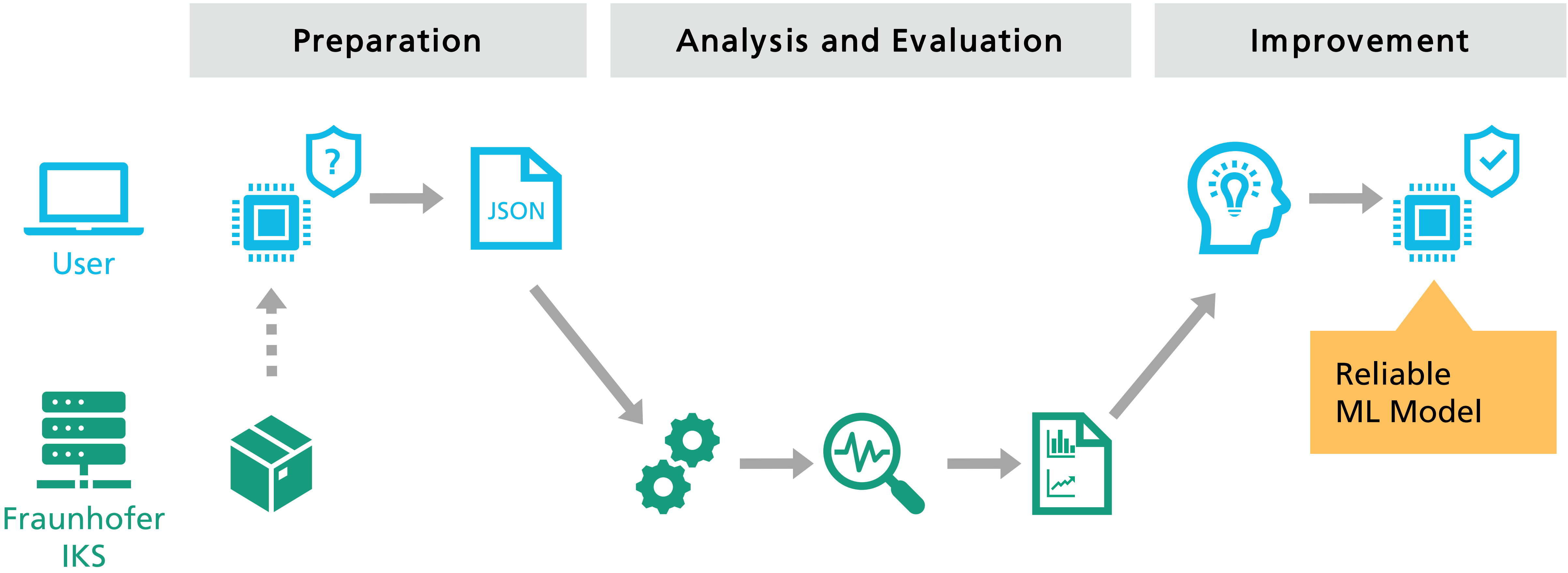
L. Heidemann et al. „Measuring Ensemble Diversity and Its Effects on Model Robustness“, AISafety 2021

DEEP DIVE: ENSEMBLE DIVERSITY (OOD)

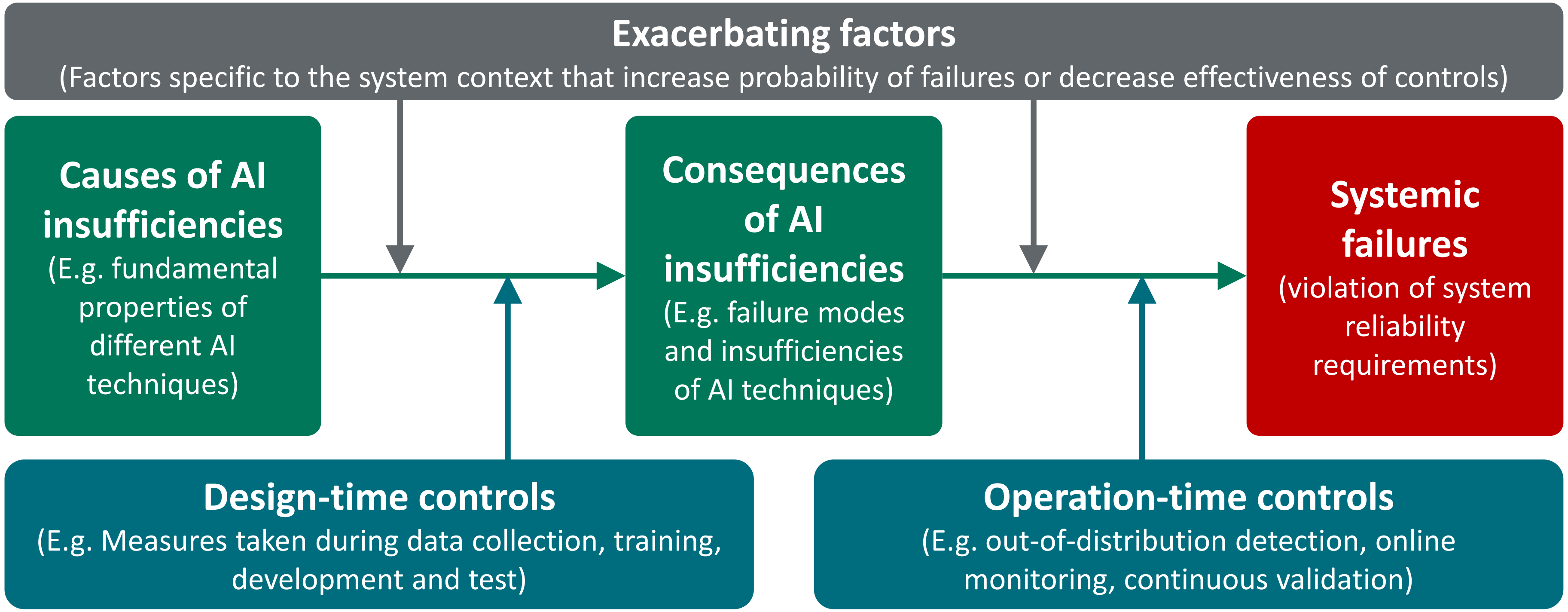


L. Heidemann et al. „Measuring Ensemble Diversity and Its Effects on Model Robustness", AISafety 2021

BETTER KNOW YOUR MODEL

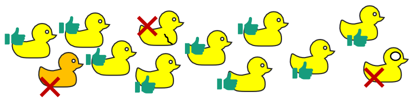


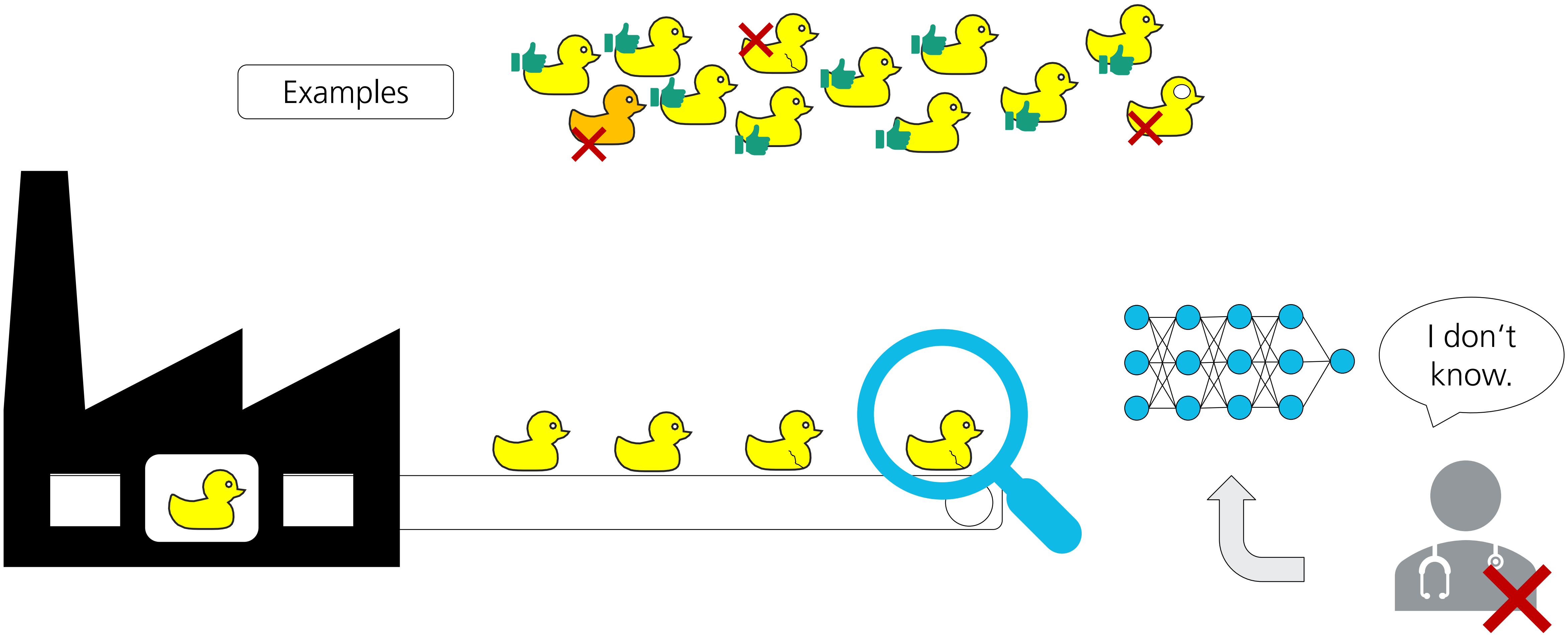
UNDERSTANDING THE IMPACT OF AI ON RELIABILITY

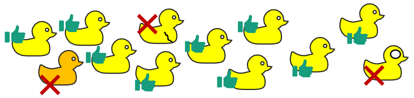


- What properties of AI need to be argued for them to be considered reliable?
- Which evidence, in which combination is required?

ACTIVE LEARNING: ASK THE EXPERT

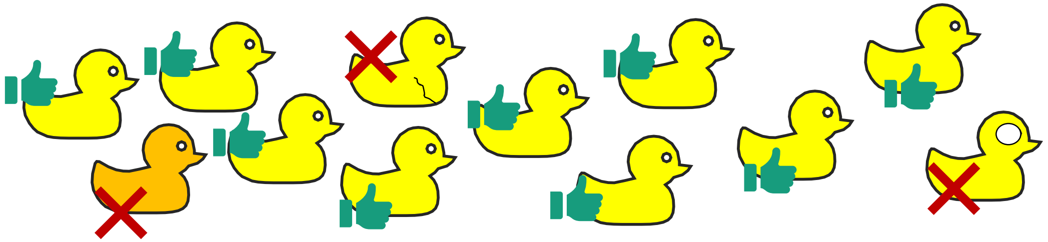


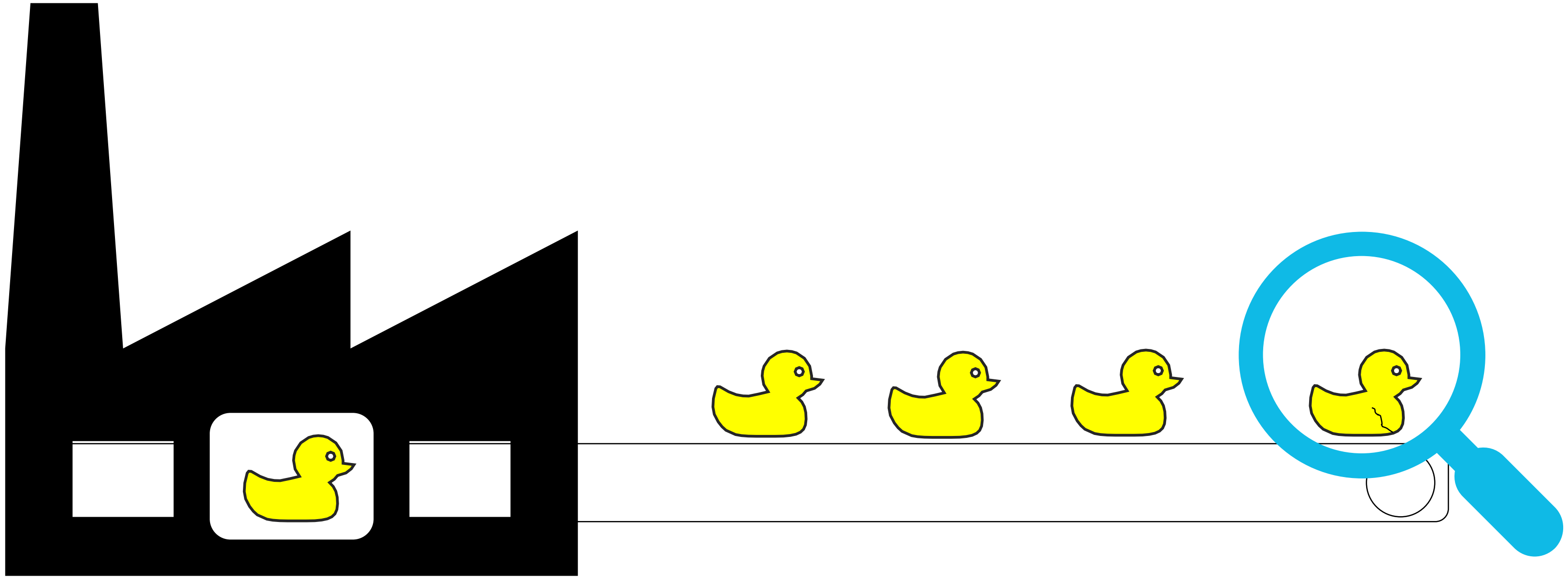


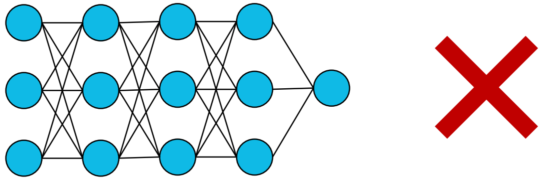


ACTIVE LEARNING: ASK THE EXPERT

Examples

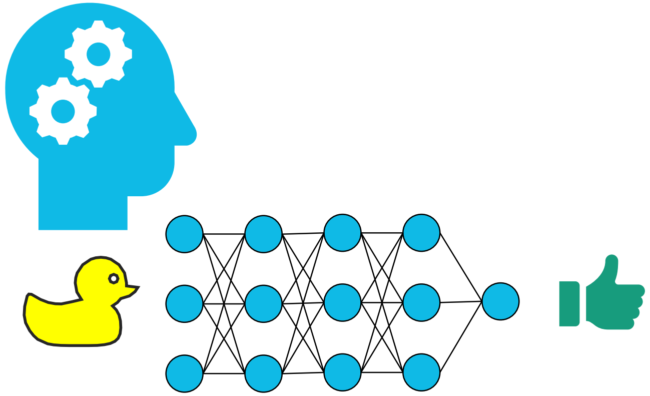


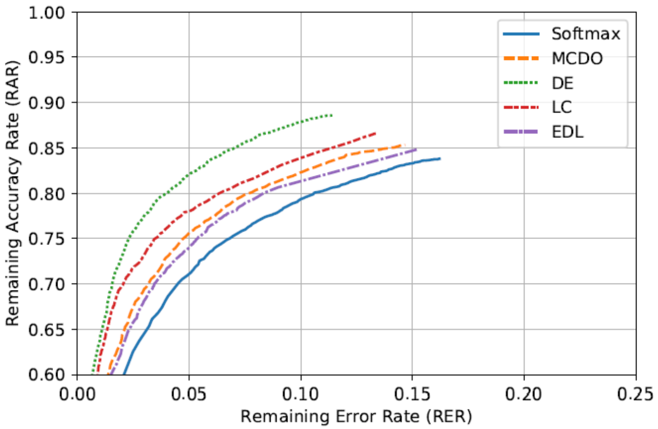


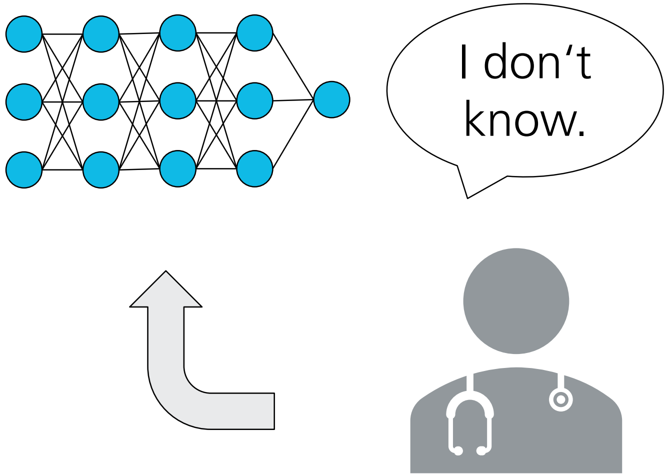


If **uncertainty is reliably estimated**, we can start with little data and a low performing model gradually **improving it in use**.

SUMMARY







Machine Learning can automate time consuming and expensive inspection and analysis task.

Uncertainty Estimation is a key component to reduce unwanted errors and improve reliability of the function.

Active Learning is a promising approach to reduce the cost and risk of data acquisition and preparation upfront.

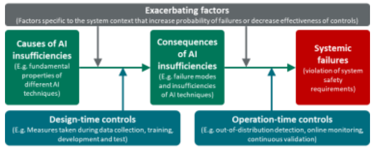
CONTACT

Karsten Roscher

Department Head Certifiable Artifical Intelligence

karsten.roscher@iks.fraunhofer.de

ONGOING RESEARCH – ROLE OF QUANTITATIVE RISK ASSESSMENT IN DETERMINING RELIABILITY OF AI



2. Evaluation of impact of ML insufficiencies on performance

Understand performance, generalization, robustness, fairness, explainability properties.

Example Evidences

- *Adversarial frequency, consistency, occlusion sensitivity, uncertainty quantification*

Is there a correlation to the residual failure rate?

1. Direct measurement of failure rate of ML function

E.g. based on test data

Example Evidences

- *Accuracy, Precision, Recall, mIoU,…*
- Based on limited number of test/samples (S)

How accurate are our performance predictions?

Definition of **quantitative acceptance criteria,** decompose to ML functions

What level of performance is required?

3. Evaluation of effectiveness of **design-time methods**

How rigorously were design-time measures for reducing insufficiencies applied.

Example Evidences

- *Training data selection criteria, Test scenario coverage, adversarial confidence loss*

What level of rigor in the development process is required?

4. Evaluation of effectiveness of **operation-time methods**

To what extent do architecture measures reduce residual failure rate.

Example Evidences

- *Out of Distribution Detection, Uncertainty quantification realism,*

What level of diagnostic coverage can be achieved?