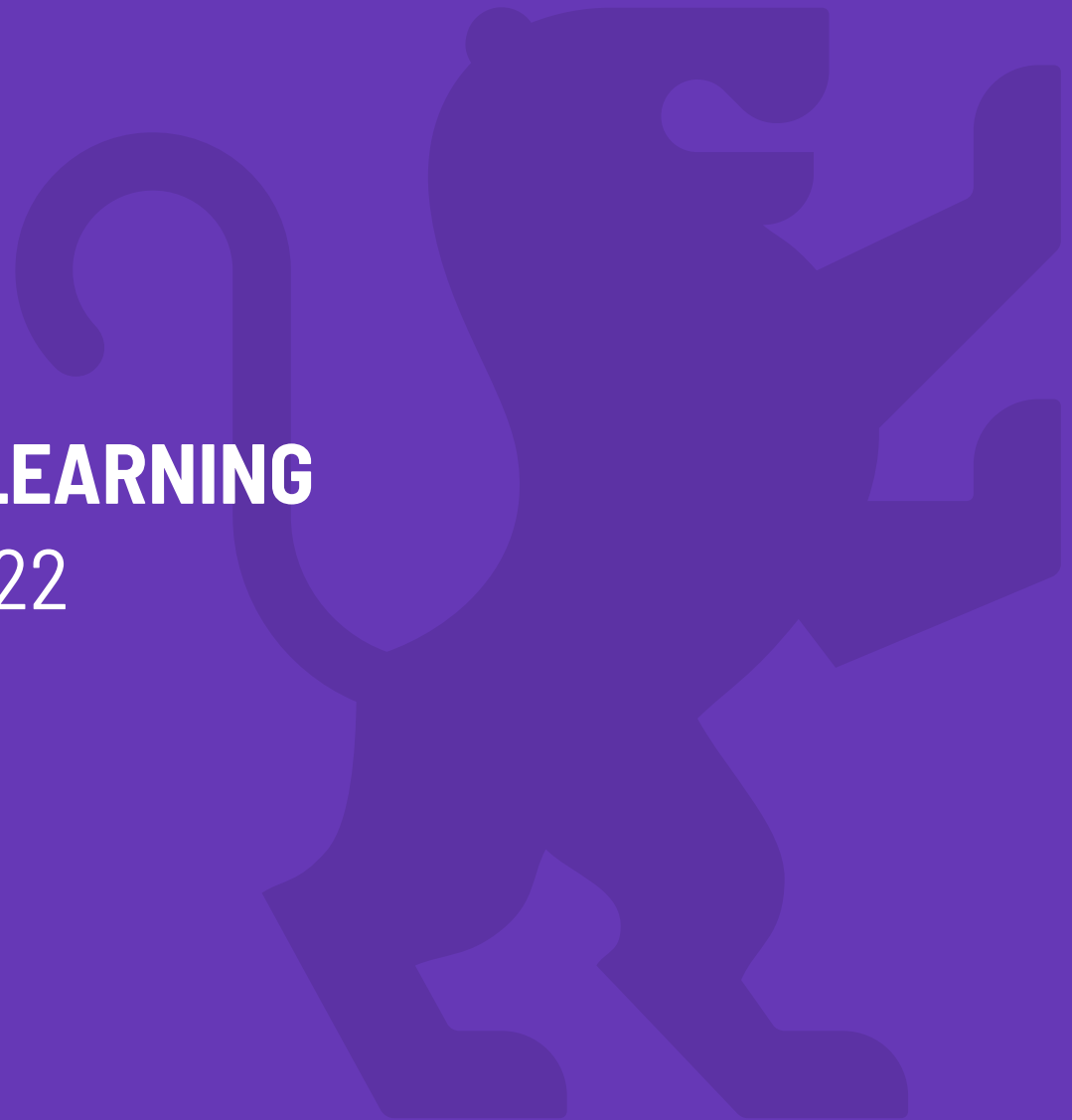




HOCHSCHULE
RAVENSBURG-WEINGARTEN
UNIVERSITY
OF APPLIED SCIENCES

APPLIED DEEP LEARNING MEETS SAFETY '22

Mark Schutera @RWU @DHBW





How you will be examined

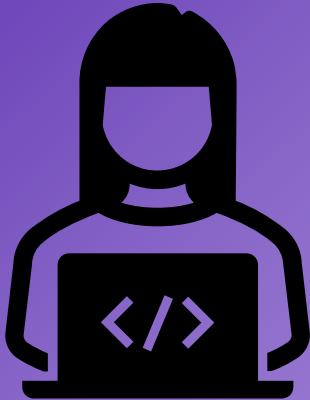
- Your contribution to the **weekly workshops** and **status stand ups**
- Points on potential **bonus tasks** during the time of the lecture
- **(40) Final report** on your „Safety concept for applied deep learning applications“ (max 5 pages) as pdf.
 - Motivate the underlying issue of safety and testing for deep learning perception applications
 - Overview of your novel concept proposal (your code contribution) to ensure safety and testing (This might include methods, processes and roles, with examples how this worked during the lectures)
 - Motivation and discussion of your findings and an outlook on your safety concept and how future students and teams should proceed with validation (outlook).
- **(60) Software Contribution** in the form of a python script (binder, github, jupyter notebook with environment or requirements).
 - Add value to a general **validation process** or a **safety feature** for predictive models, including a **validation_batch**
 - Examples are metrics libraries, loggers, notifiers, validation functionalities as wrappers, dda

Submission Deadline: 13.06.2022



What does Corona and Deep Learning have in common?

TESTING IS KEY.



Culture Wars

Right now we have a situation in which different disciplines with different approaches, and cultures are trying to work together.



Silicon Valley –

“Let’s disrupt, move
fast and break things”

Culture Wars

Right now we have a situation in which different disciplines with different approaches, and cultures are trying to work together.



Silicon Valley –

“Let’s disrupt, move fast and break things”



AI community –

“I detect balloons with 99% accuracy, this is amazing”

Culture Wars

Right now we have a situation in which different disciplines with different approaches, and cultures are trying to work together.



Silicon Valley –

“Let’s disrupt, move fast and break things”



AI community –

“I detect balloons with 99% accuracy, this is amazing”



Safety community –

“Impressive, but life critical system design requires something around 99.999999% to start with”

Culture Wars

Right now we have a situation in which different disciplines with different approaches, and cultures are trying to work together.



Silicon Valley –

“Let’s disrupt, move fast and break things”



AI community –

“I detect balloons with 99% accuracy, this is amazing”



Safety community –

“Impressive, but life critical system design requires something around 99.999999% to start with”



Automotive culture –

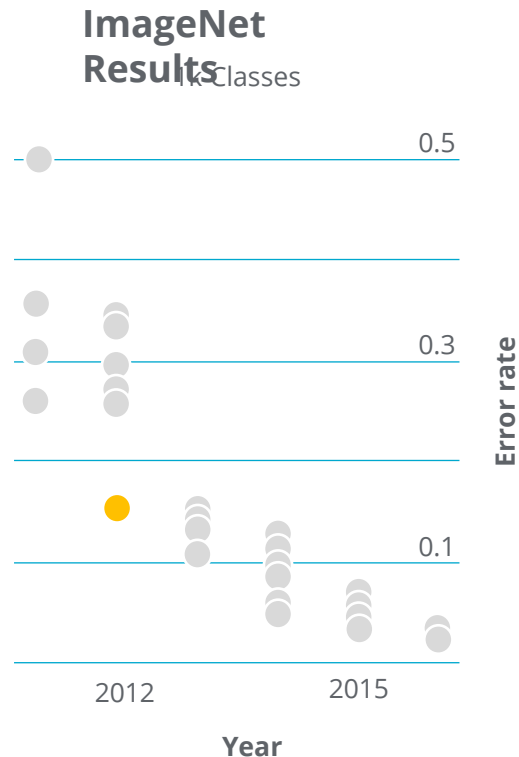
“Critical system failure? Probably it was driver error” – “wait what”

The ImageNet Revolution

Autonomous driving technology is motivated by **economic** and **safety** benefits for the customer.

So why is this possible now?

- 1 Deep Learning revolutionized perception
- 2 GPUs provide the compute

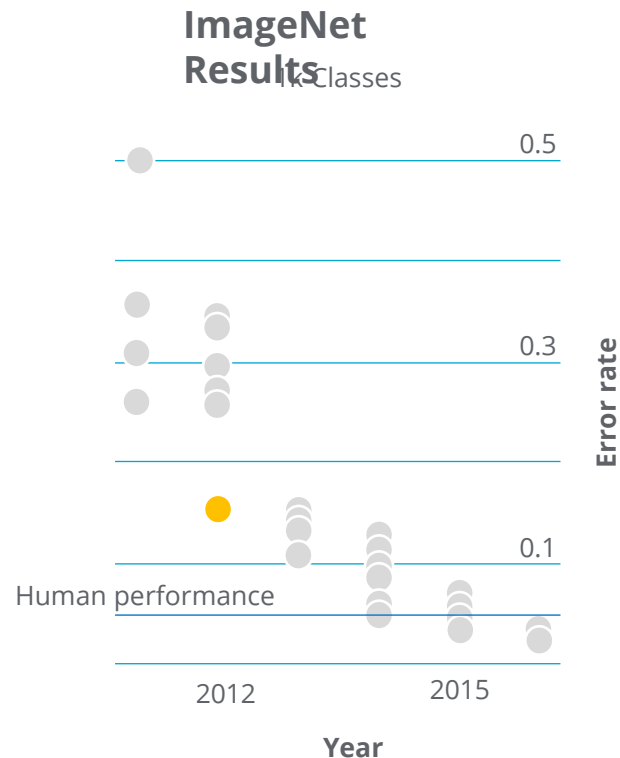


The ImageNet Revolution

Autonomous driving technology is motivated by **economic** and **safety** benefits for the customer.

So why is this possible now?

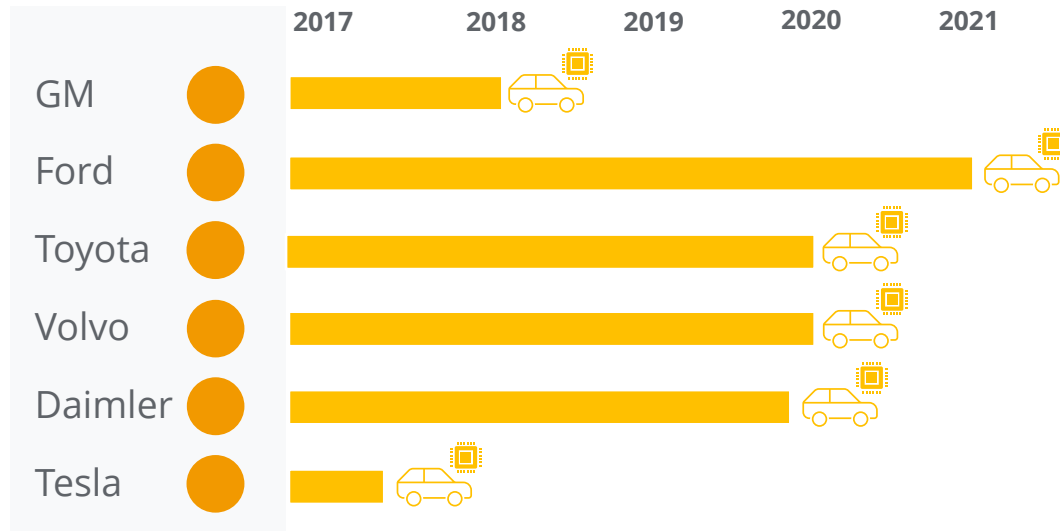
- 1 Deep Learning revolutionized perception
- 2 GPUs provide the compute



Because, Machine Learning

Machine learning is a **statistical approach**, and has a hard time to get over 99% accuracy – however not enough for safety critical systems.

It is **hard to prove** wheter a model performs as (good as) intended.

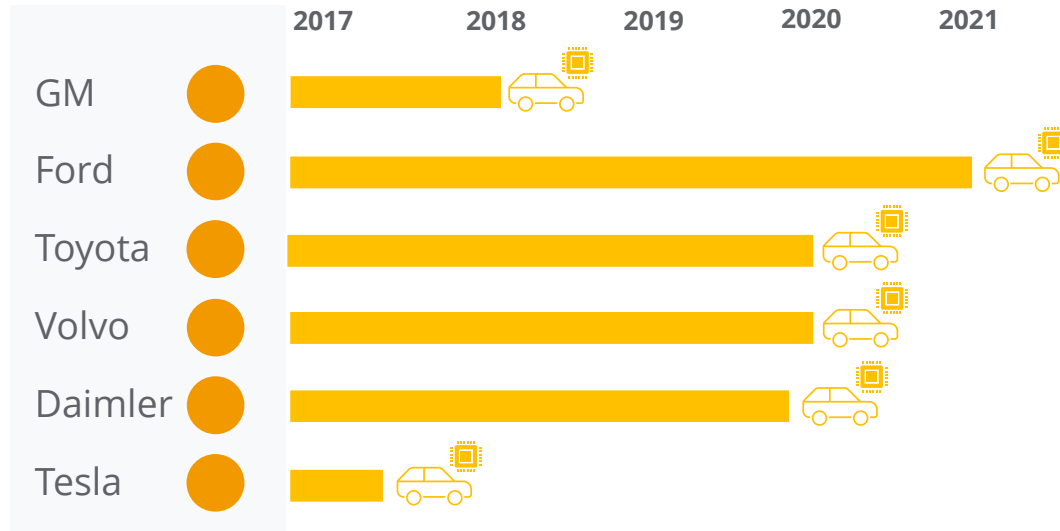


**The Hype
is real by
2021**

Because, Machine Learning

Machine learning is a **statistical approach**, and has a hard time to get over 99% accuracy – however not enough for safety critical systems.

It is **hard to prove** wheter a model performs as (good as) intended.










**No, calling it
autopilot does
not count**

Mum, Machine Learning breaks my safety process

How does Deep Learning work in general?

A Developer / Model Dialogue in 7 lines

Training

-  What is 3+5?
-  6
-  No, please try again.
-  7
-  What the heck dude.
-  8
-  Oh my god, it is intelligent!

Mum, Machine Learning breaks my safety process

So why is this so hard to validate?

Potential to overfit on data seen during training.

This makes gaps in the training and validation data critical.

Validation



8



Oh my god, it is intelligent!



What is 3+4?



8



Are you kidding me?

Mum, Machine Learning breaks my safety process

So why is this so hard to validate?

Potential to overfit on data seen during training.

This makes gaps in the training data critical. Further it is hard to validate behavior on potentially infinite unseen samples (unknown unknowns).

Deployment



What is 1-2?



97

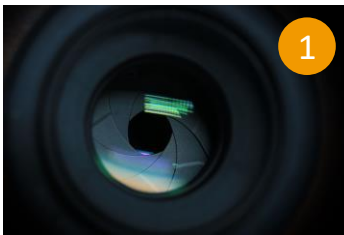


.. seriously?



AI Incidents

Why, it is about time to think about assessment, certification and safety.



1

2021-09-20 Amazon's AI Cameras punish drivers for mistakes they did not make



2

2021-09-20 Uber's Self-Driving Car Runs Red Light in San Francisco



3

2015-06-03 Google's photo tag identifies black people as Gorillas



4

2015-04-04 Google's Search has a CEO Gender Bias – 11% to 27%



5

2021-02-19 Youtube Algorithm blocks video on "black vs white" chess strategy



6

2016-06-30 Driver killed in crash with Tesla's autopilot active

The Origin Validation [Batch_0]

Set up your codebase, get the data, train the model. Think about how to validate your model, and set up a validation pipeline and report. Think about how to engineer a great test dataset, what are the characteristics that make it a great one? Think about metrics, how to select meaningful metrics? And think about how you can answer the question what is your model able to do? Which guarantees would you give with confidence? How can we measure this confidence? How confident are you about your models performance at this point? Think about a process, roles and concept that would enable you to answer that question as you improve your network starting today. Maybe you want to split your team in a performance (training) and safety (validation) group with the goal to outperform each other? Or are there different structures that would help with achieving your goals? The more you invest in processes and structured approaches now, the easier you will coast through the course.

The False Friend Validation [Batch_1]

The first thing to learn today, is that there is probably always an additional class you have not taken into consideration. Start thinking about a concept to tackle this real-world problem! Which approaches can you take?

If you already have added this class to your dataset and model, congratulations. Has this been pure chance? Do you already have a strategy for this in place? Would you have been less lucky with another traffic sign? Are you reporting your model performance over time (if not start doing so).

The second thing to learn is that this class is close to a class already in your dataset (Class 10). Chances are that your model thus gives you high confidence while completely predicting the wrong class. This is a behavior we also know from humans, but did we not expect neural networks to behave objective. Make this ungrounded confidence visible in your reports.



The Variance Validation [Batch_2]

This class has been within your model since day 1. So how did you perform on this class? Okay you also have been tricked, if you have not increased your training dataset, this class only has 42 training samples to learn from. Now being confronted with a few samples with increased variance (meaning same thing but slightly different), does your model break? Which role does class imbalance play in your validation dataset? How can you take this into account?

Are there approaches to tackle this? What can you do to tackle aleatoric uncertainty in your input data? Can this be done in a structured way? How can we detect potential uncertainty in input samples? How can we quantify this?

Further, with this validation you should have noticed that in deep learning we tend to give accuracies as a percentage (a relative measure), what are the absolute failure numbers for your model? Once deep learning is applied in a safety critical system, resorting to relative numbers only becomes ambiguous.



The Domain Shift Validation [Batch_3]

This class has been within your model since day 1, too. Looking at your validation, what did go wrong? Or what did you do, so this time the validation was easy to overcome? In case you have a proper augmentation pipeline set up by now, this domain adaptation might not have hurt you too much.

In that case how did you decide on this augmentation method? Why not others? Have you thought about rain or fog augmentations or something entirely different?

And what could go wrong when augmenting your data?

How can we augment samples? How can we anticipate domain changes? Are there continuous domain adaptation necessities?



The Coverage Validation [Batch_5]

You have come a long way, and if you realized how hard it is to validate a supposedly trivial perception system and understood perception task – you have learned your most important lesson of this course.

This validation will get to the bottom of your model, this is the closest we can get to an actual battle test in the context of this lecture.

How robust is your system, how extensive is your system. Can you come up with an estimation of how much you covered with your model?



Your Validation [Batch_6]

You have come a long way, and if you realized how hard it is to validate a supposedly trivial perception system. And how little one knows about a perception task that on the first look seems to be well understood – you have learned your most important lesson of this course.

To wrap up things this validation set will be yours to engineer, yours to design. This is your chance to tackle the potential weaknesses of your model that have remained uncovered. And at the same time you can show of how powerful your safety and validation strategies, approaches and processes have become by now.

Also remember that there is a documentation / report due at the end of the lectures. There is a reason this is called a deadline.





HOCHSCHULE
RAVENSBURG-WEINGARTEN
UNIVERSITY
OF APPLIED SCIENCES

STAY SAFE



github.com/schutera/DeepSafety



mark.schutera@zf.com



<https://www.linkedin.com/in/schuteramark/>