



Safety Critical AI in Autonomous Driving

Szilárd Németh

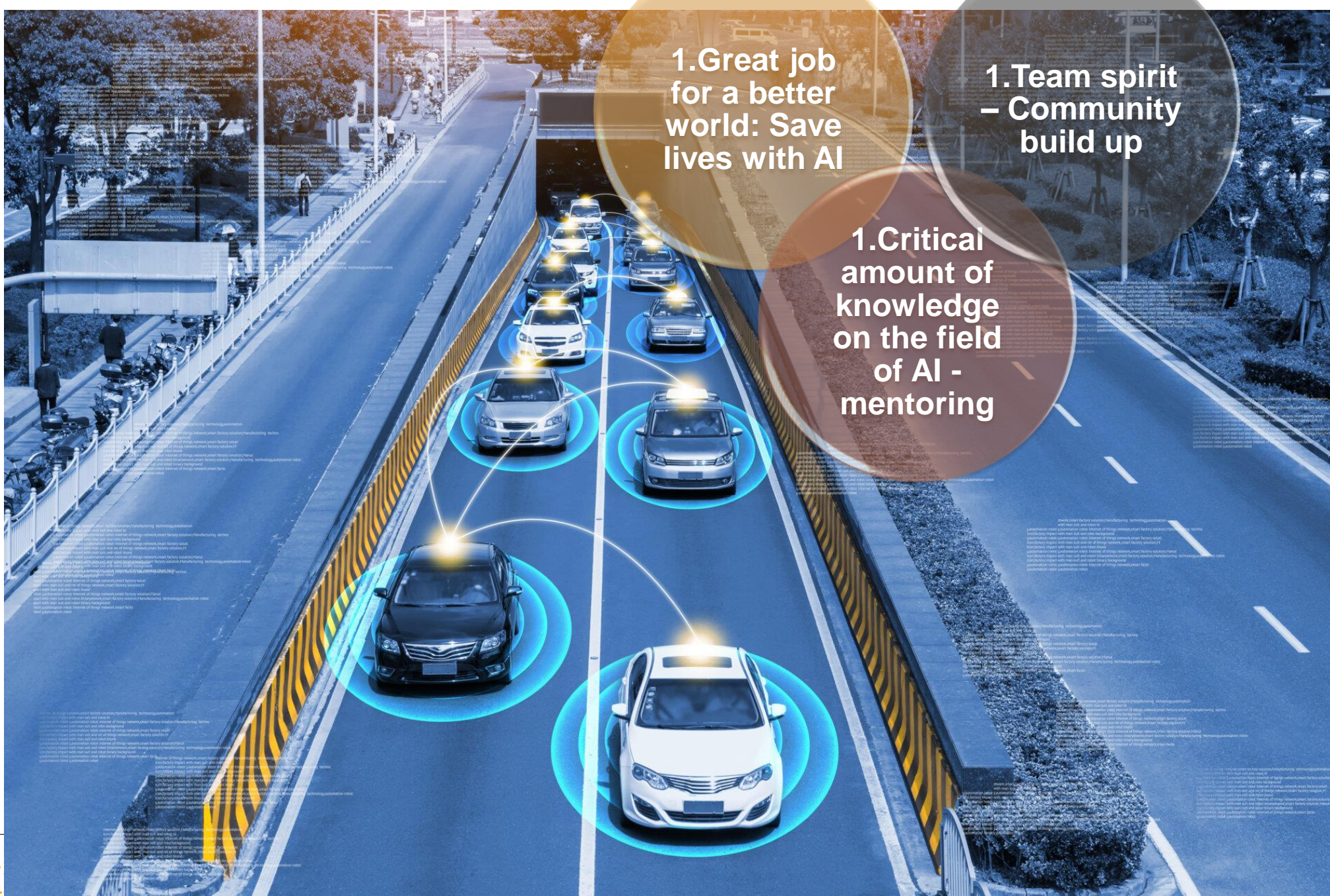
Machine Learning Expert
Artificial Intelligence Development Centre, Budapest

LET
YOUR
IDEAS
SHAPE
THE
FUTURE

Introduction – AI Development Center Budapest

- Established in 2018
- Critical amount of knowledge: 200+
- Sensing of the environment
- AI model & software development





1. Great job for a better world: Save lives with AI

1. Team spirit – Community build up

1. Critical amount of knowledge on the field of AI - mentoring

Safety critical AI in autonomous driving

*What is **autonomous driving**?*

*What is **AI**?*

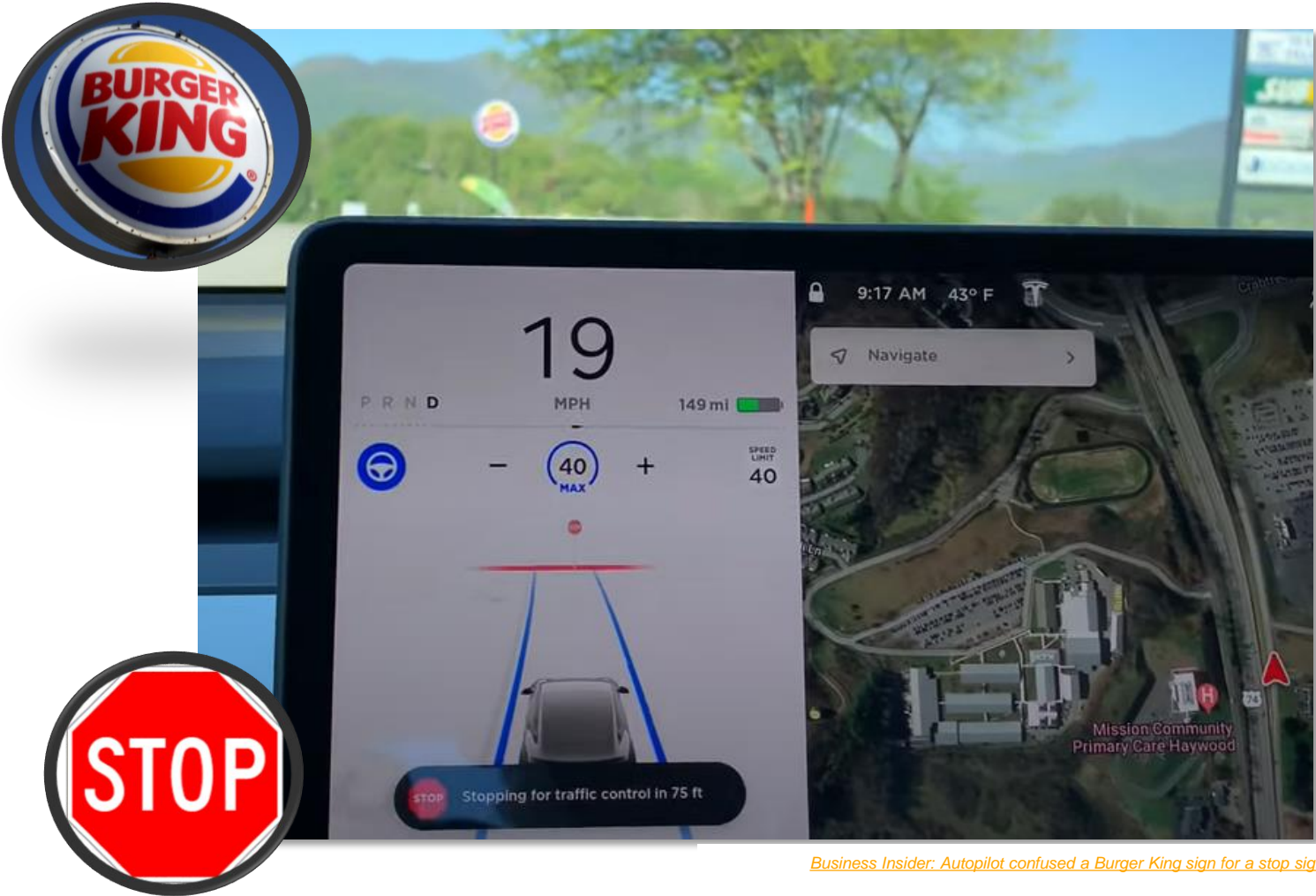
*What is **Safety**?*

How AI sees the world

REALITY



AI INTERPRETATION



Business Insider: Autopilot confused a Burger King sign for a stop sign

How humans see the world

Which square is darker?
A or B?



How humans see the world

Which square is darker?
A or B?



*Errors by human and errors by AI are **differently distributed***

What is Safety?

What is safety?

“the condition of being safe from undergoing or causing hurt, injury, or loss”
“a device designed to prevent inadvertent or hazardous operation”

BUT HOW?



<https://xkcd.com/1897/>

SO MUCH OF "AI" IS JUST FIGURING OUT WAYS TO OFFLOAD WORK ONTO RANDOM STRANGERS.

Functional safety (ISO 26262)

WHAT: *“Absence of unreasonable risk due to hazards caused by malfunctioning behavior of E/E systems.”*

HOW: tools, processes, guidelines

Regulatory landscape – a complex terrain to navigate

Current

Functional safety standards do not cover AI development

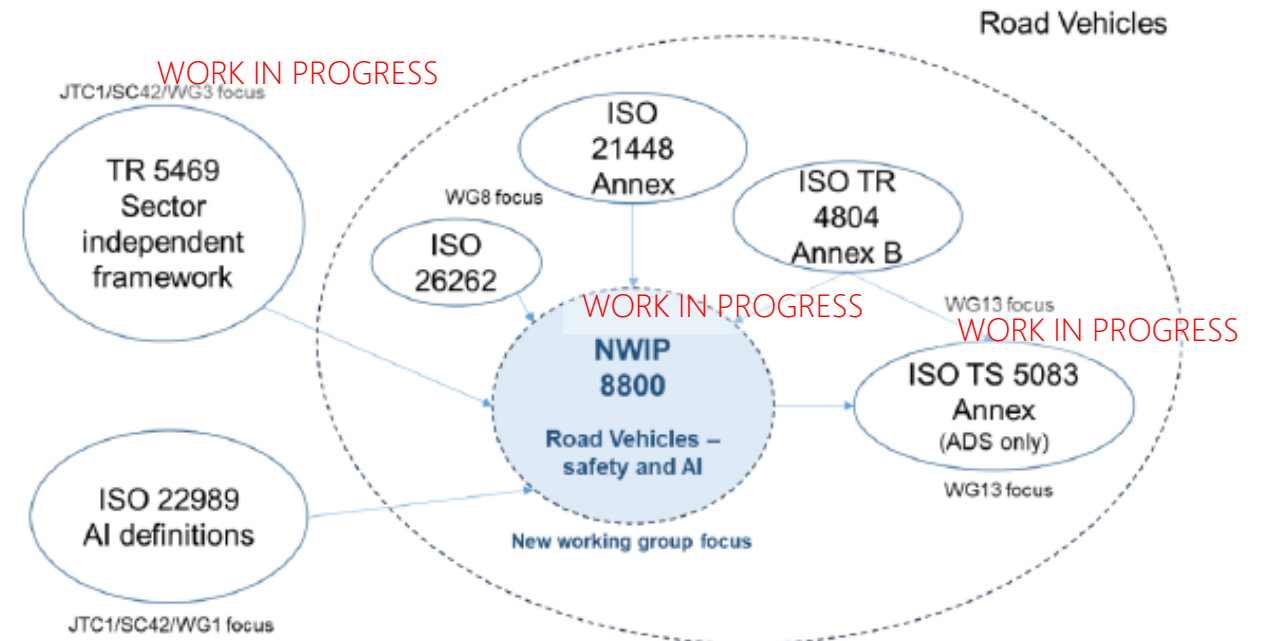
Work in progress

ISO 5469 - Artificial intelligence – Functional safety and AI systems

ISO PAS 8800:Road Vehicles – Safety and Artificial Intelligence

Implication

No active safety standard to support AI development



https://unece.org/sites/default/files/2021-09/GRVA-11-13e_0.pdf

What is Autonomous Driving?

Levels of autonomy



SAE J3016™ LEVELS OF DRIVING AUTOMATION

Liability starts here

	SAE LEVEL 0	SAE LEVEL 1	SAE LEVEL 2	SAE LEVEL 3	SAE LEVEL 4	SAE LEVEL 5
What does the human in the driver's seat have to do?	You <u>are</u> driving whenever these driver support features are engaged – even if your feet are off the pedals and you are not steering			You <u>are not</u> driving when these automated driving features are engaged – even if you are seated in “the driver’s seat”		
	You must constantly supervise these support features; you must steer, brake or accelerate as needed to maintain safety			When the feature requests, you must drive	These automated driving features will not require you to take over driving	

These are driver support features

What do these features do?	These features are limited to providing warnings and momentary assistance	These features provide steering OR brake/acceleration support to the driver	These features provide steering AND brake/acceleration support to the driver
Example Features	<ul style="list-style-type: none"> • automatic emergency braking • blind spot warning • lane departure warning 	<ul style="list-style-type: none"> • lane centering OR • adaptive cruise control 	<ul style="list-style-type: none"> • lane centering AND • adaptive cruise control at the same time

These are automated driving features

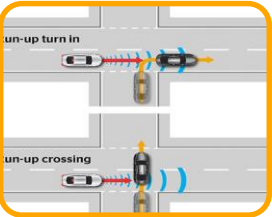
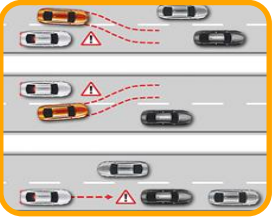
What do these features do?	These features can drive the vehicle under limited conditions and will not operate unless all required conditions are met	This feature can drive the vehicle under all conditions
Example Features	<ul style="list-style-type: none"> • traffic jam chauffeur • local driverless taxi • pedals/steering wheel may or may not be installed 	<ul style="list-style-type: none"> • same as level 4, but feature can drive everywhere in all conditions

For a more complete description, please download a free copy of SAE J3016: https://www.sae.org/standards/content/J3016_201806/

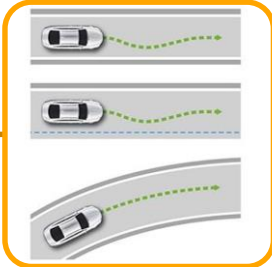


Functions in autonomous driving – L2/L3

Emergency Brake Assist



Lane Keeping Assist



Traffic Sign Assist



Automated Parking



Driver Monitoring



Implications

- **Complex, infinite** operational domain with unpredictable conditions
- **Liability** for L3-L5
- Wide range of **functions** for self-driving
- **Low probability** of **high impact** scenarios
 - Low probability: lack of data
 - High impact: safety implications
- **Full specification** of the domain is practically infeasible. *Motto: what is a pedestrian?*

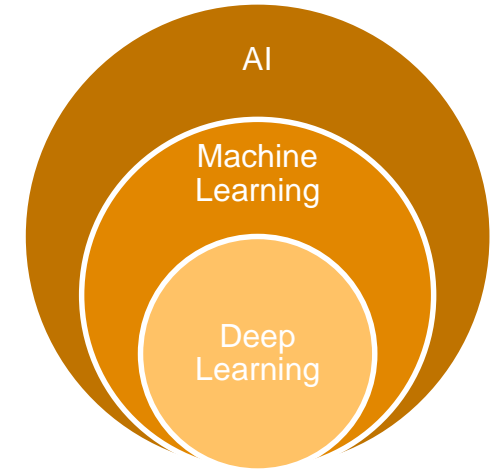
Colorful world



What is AI?

What is Artificial Intelligence (AI)?

- AI can be achieved in multiple ways, Machine Learning is one of those
- Deep Learning is concerned with scalable models that require less human intervention
- AI is flexible and powerful in realizing multiple tasks



object detection



https://en.wikipedia.org/wiki/Object_detection

lane detection



<https://arxiv.org/abs/2204.07335>

pose estimation



<https://arxiv.org/pdf/2211.03375.pdf>

semantic segmentation



<https://arxiv.org/pdf/2105.05633.pdf>

AI as a software

Software 1.0

Conventional (non-AI) development: hand crafted rules by **analyzing** and **decomposing** the problem

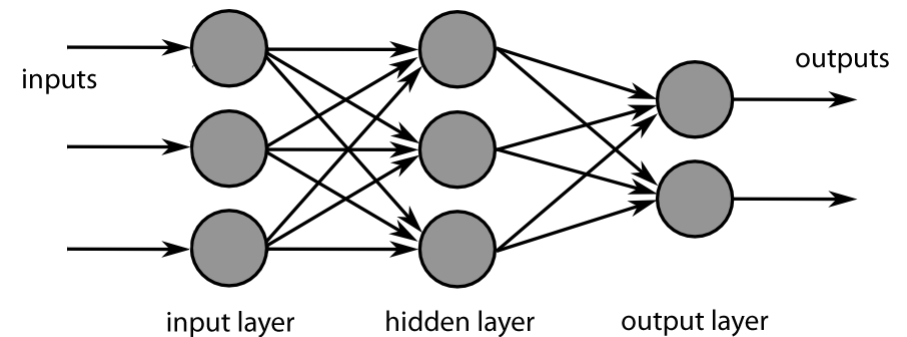
Software 2.0

AI: a data-driven method aim to **learn rules from data**

Software 1.0



Software 2.0



Challenges of AI – a safety perspective

Ian J. Goodfellow et al. "Explaining and Harnessing Adversarial Examples"

Explainability

Why did model fail under certain condition and works in others (even seemingly similar conditions)?
What is a fault model for AI?

Robustness

Sensitivity to even unobservable (to a human) variation?

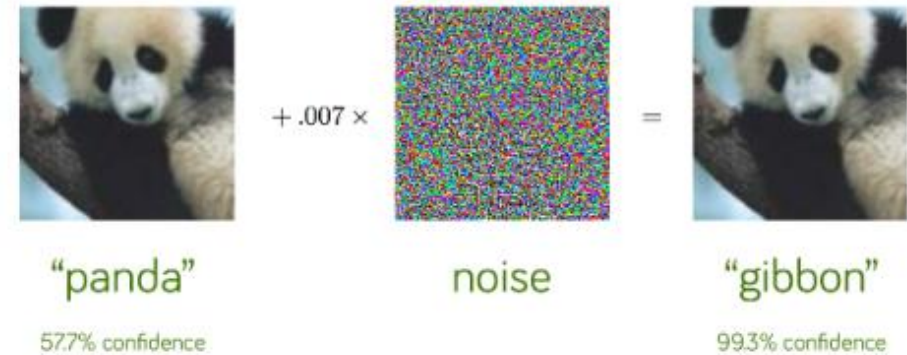
Specifiability

How to define data requirement based on AI output requirements?

Verifiability

How to analyze and measure performance?
How to use required tools or simulation data?

A funny mistake



And a serious one



<https://www.ft.com/content/3b0eaba6-38b5-11e8-8b98-2f31af407cc8>

Handling complexity

AI Safety for Autonomous Driving

ENVIRONMENT — complex operational domain

TECHNOLOGY — machine learning / artificial intelligence is a black box

ORGANIZATIONAL — ML looks like software engineering

REGULATORY — no live standard for AI safety

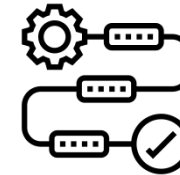
LEGAL — autonomous (versus assisted) driving results in liability

SOCIETY — what is the tolerable risk from a machine?

AI Safety – addressing challenges



Regulatory landscape



AI Methodology



Automotive Processes

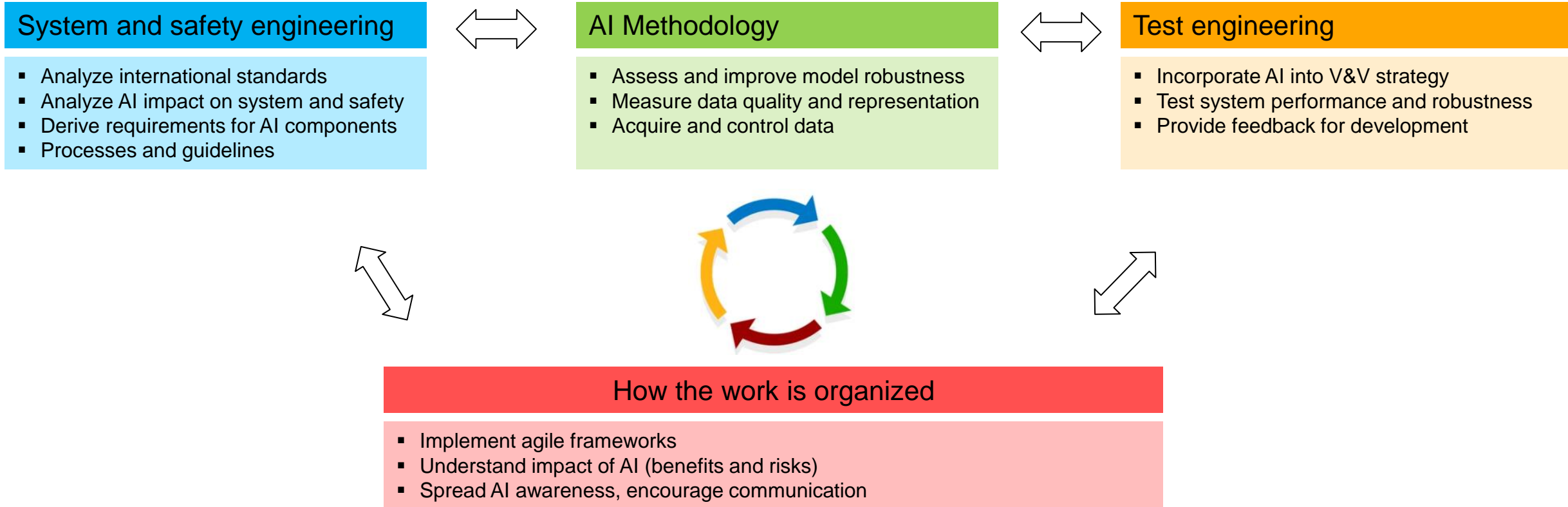


Data governance



Organization

AI Safety – structuring efforts



Best practices

- **Close cooperation** between functions (AI Experts, Safety Managers, System Engineers, V&V team, Software developers)
- Spread **AI awareness**, guilds are great
- Identify (simple) **data-driven components** and use them for a case study
- Do not try to just “won” challenges (e.g. as competitions on Kaggle), develop insight and understanding instead
- Spend time to **understand your data**, identifying flaws early on will pay off
- Find AI performance insufficiencies via exploratory analysis
- Involve (independent) quality organization in controlling data

Handling AI safety is a cross-functional job

Safe and Dynamic Driving towards Vision Zero



SensePlanAct



Szilárd Németh

Machine Learning Expert

AI Development Center Budapest

szilard.nemeth@continental.com