Open Source meets AI meets Functional Safety
How to reach compliance to ISO 26262?

- Florian Bogenberger -

Eclipse Summit 2026
30th of June 2026, Starnberg

exida.com GmbH
Prof.-Messerschmitt-Straße 1
D-85579 Neubiberg

About *exida.com* GmbH


Process Industry


End User

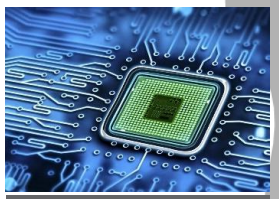

Automotive


Automation


Supplier


Robotics


Medical


Semiconductors

Training & Coaching

CRA – Cyber Resilience Act
Introduction to Functional Safety, Cyber Security, SOTIF
Gap analysis & Gap Closure
Domain or customer specific tailoring
On-site, online, webinars

Consulting Services

Safety, SOTIF, Cyber Security
AI Supported Development
Processes, Methodologies, Templates & Tools
Concepts, Analyses, Safety Case, …

Standardization

ISO 26262, IEC 61508, ISO 21448 …
ISO/SAE 21434, IEC 62433
ISO 8800, IEC 22440, …
etc.

CRA
Audits, Assessments & Certifications

Products, Processes
Organizations
People

About the Presenter

Dipl.-Ing. **Florian Bogenberger**

Managing Consultant for Functional Safety and Cybersecurity for innovative Projects, CFSE

1993 : Motorola Semiconductors, later Freescale Semiconductors
2008 : Elektrobit Automotive GmbH
2015 : exida.com GmbH

As Managing Safety Consultant, Florian is leading a team of technology specialists at *exida* and consults leading companies to implement new safety and security related systems with cutting-edge technologies including AI for Functional Safety, SOTIF and Cybersecurity.

Projects / Special Interest

- Requirements Engineering, Systems Engineering, ..
- SW-Development, HW-Development, Semiconductors
- Formal Specification & Verification
- Assessment, Audit, Certification

- High Performance Computing-Platforms, Autonomous Driving
- New Methods & Approaches to master Complexity

Mailto: florian.bogenberger@exida.com

Functional Safety is about …

… protecting
human life and health

HW faults & *systematic SW faults* can lead to severe failures and accidents …

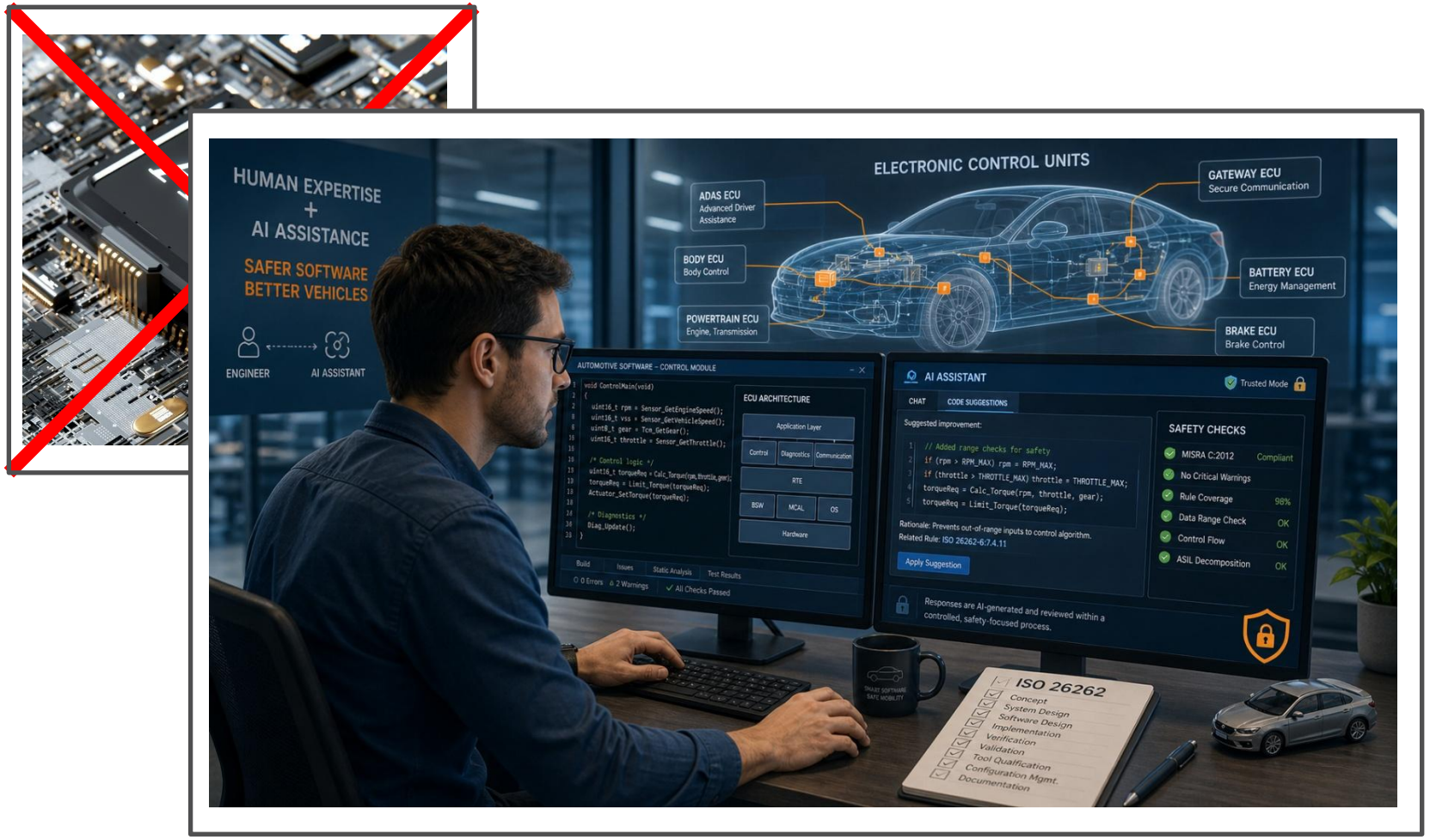
… people can be injured or killed

⇒ “best effort” engineering is not enough

⇒ law imposes significant responsibilities for functional safety on the manufacturer and its developers

Which Type of AI we mean …

In this Presentation we don´t talk about „Embedded AI“ … (ISO 8800, …)

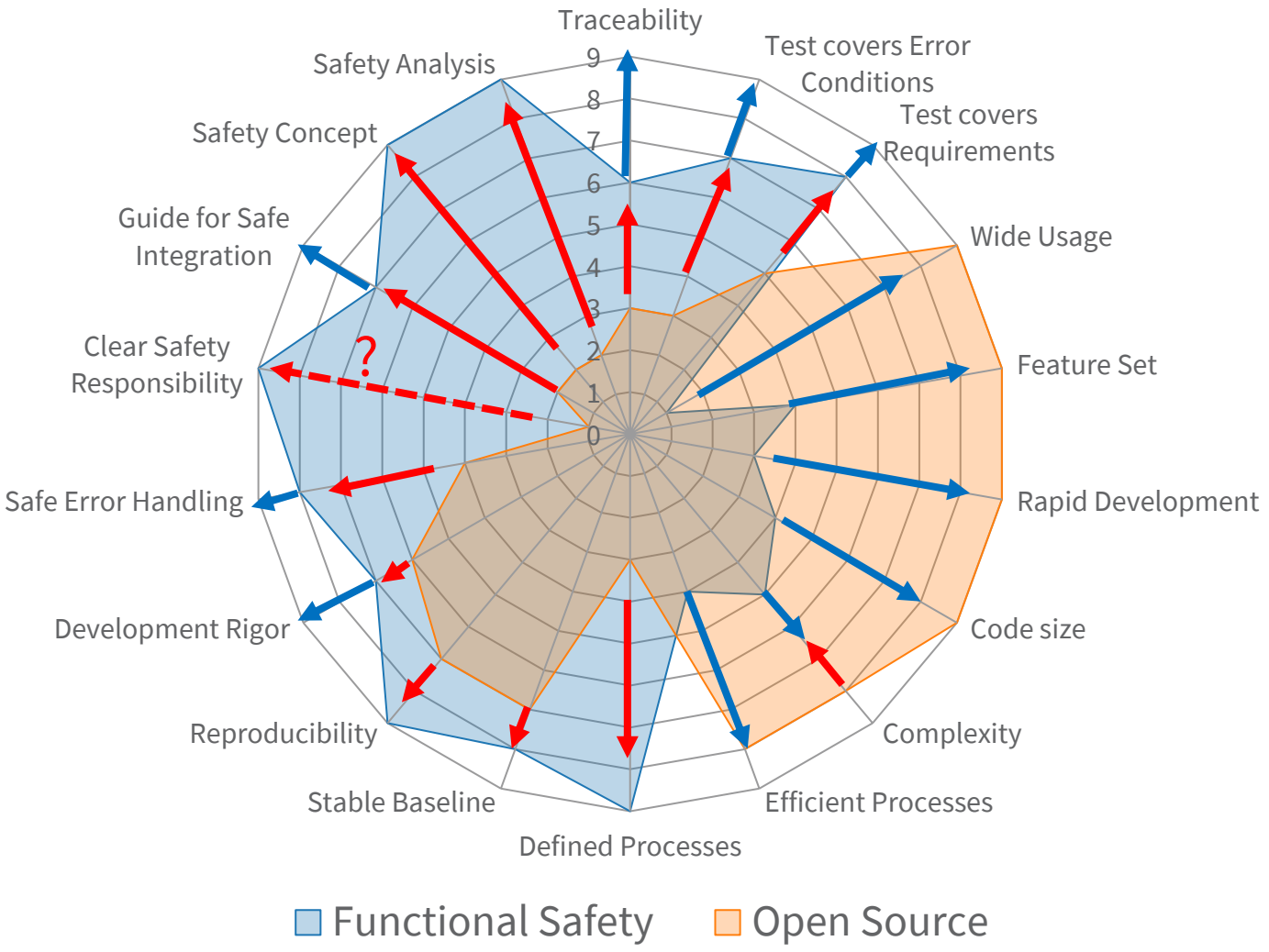


AI-Supported Development of Safety-related SW

Use-Cases throughout the Safety-Lifecycle

- directly or indirectly affecting embedded SW
- requirements, design, tools, tests, plans, …
- review, refine, verify, analyze, …
- track, control, build, monitor, …
- …

Open Source meets AI meets Functional Safety



Which Gaps …

- … can be narrowed with AI-Tooling?
- … can not be addressed by AI?
- … may be worsened by AI?

Human versus AI: ARC-AGI-3

ARC Abstraction and Reasoning Corpus (and Organization doing “Intelligence Tests” for AI)
AGI Artificial General Intelligence.

Conclusion: Today humans by far outperform the latest AI if it comes to „(real) learning and (fast) generalization“.

Model	Score	Actions	Published
Humans	100%	776	—
claude-opus-4-8	2%	644	31.05.2026, 22:49
claude-opus-4-7	0%	110	01.05.2026, 07:07
gpt-5.5-2026-04-23	0%	110	01.05.2026, 01:25
grok-4-20-beta-0309-reasoning	0%	110	14.04.2026, 19:43
gpt-5.4-2026-03-05	0%	110	14.04.2026, 19:43
gemini-3.1-pro-preview	0%	110	14.04.2026, 19:43
claude-opus-4-6	0%	110	14.04.2026, 19:43

Try it … 😉

References:
ARC Prize Foundation, “ARC Prize,” *ARC Prize*. [Online]. Available: https://arcprize.org. [Accessed: Jun. 26, 2026]
ARC Prize Foundation, “ls20,” *ARC Prize*. [Online]. Available: https://arcprize.org/tasks/ls20. [Accessed: Jun. 26, 2026]

Quick Shot – AI for Safety Tasks …

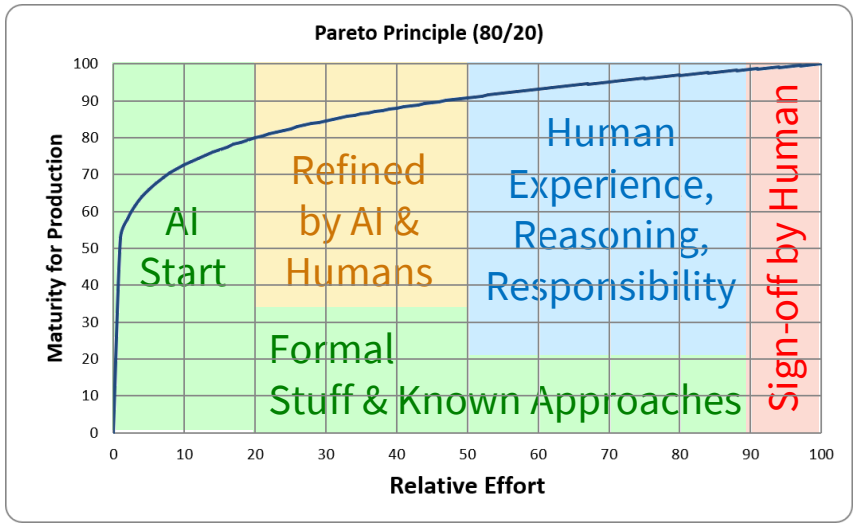
Work product of ISO 26262	Type of Work Product: Plan	Requirements	Design & Architecture	Implementation/Test	Verification & Analysis	Confirmation	Type of Safety Task: create ideas	create from scratch	create from template	create from spec	formal review	content review	thorough verification	deep analysis	generate traceability	check traceability
Safety plan																
Item definition																
Hazard analysis and risk assessment																
Functional safety concept																
Technical safety concept																
System architectural design																
Hardware architectural design																
Software architectural design			X													
Hardware safety requirements specification		X														
Software safety requirements specification		X														
System design specification																
Software unit design specif																
Software unit implementat																
Hardware design specificati																
Integration test specification					X											
System test specification					X											
Software unit test specification					X											
Test Implementation				X												
Verification report																
Validation report																
Safety analysis report (e.g. FTA/FM																
Confirmation review report																
Functional safety assessment repo																
Confirmation measures plan																
Safety case																
Production, operation, service and																
Configuration management plan																
Verification plan	X															

In General:
Results depends heavily on Input Spec (usually generated by humans)

AI flaws are not visible like human flaws

AI detects gaps in details overlooked by humans

However AI may overlook fundamental logical flaws

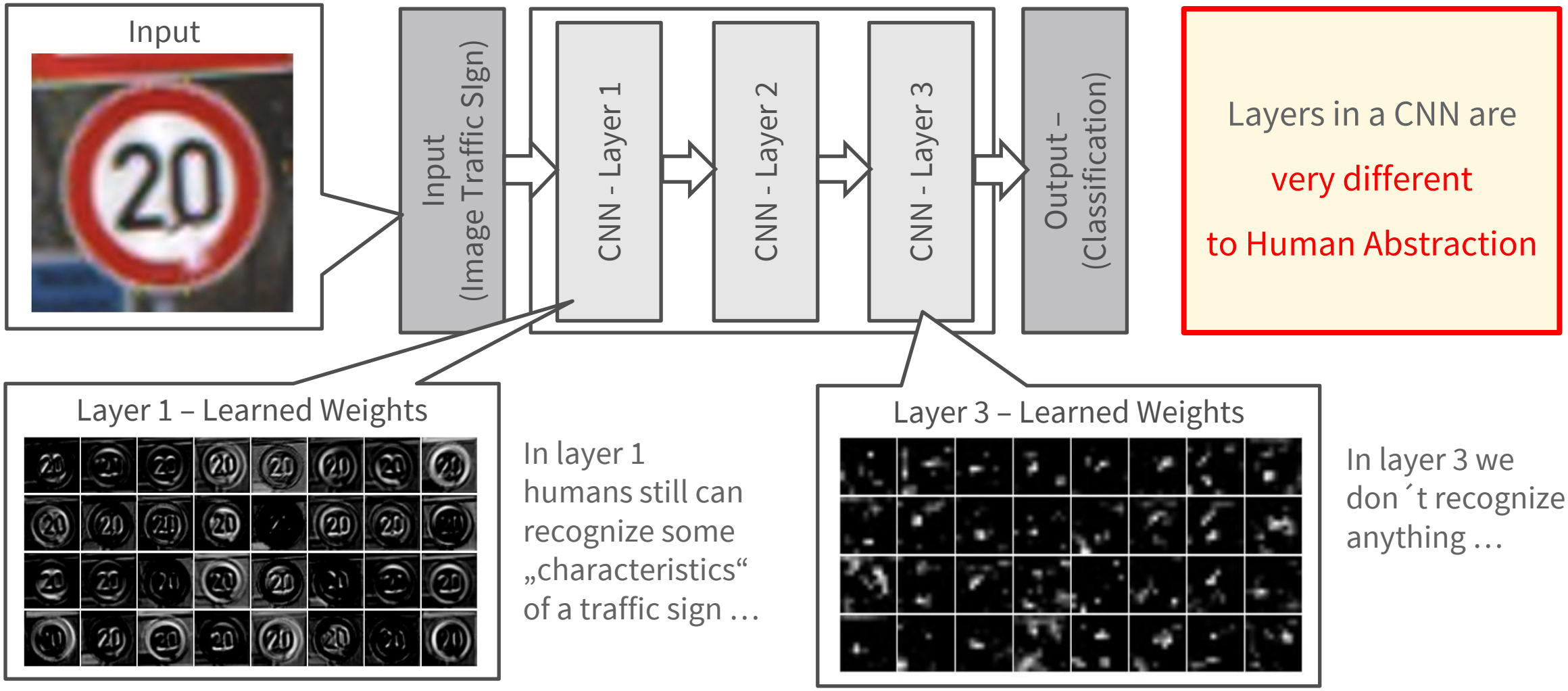


- good to start fast
- good to „sort the mess“ 😉
- good for „known“ approaches

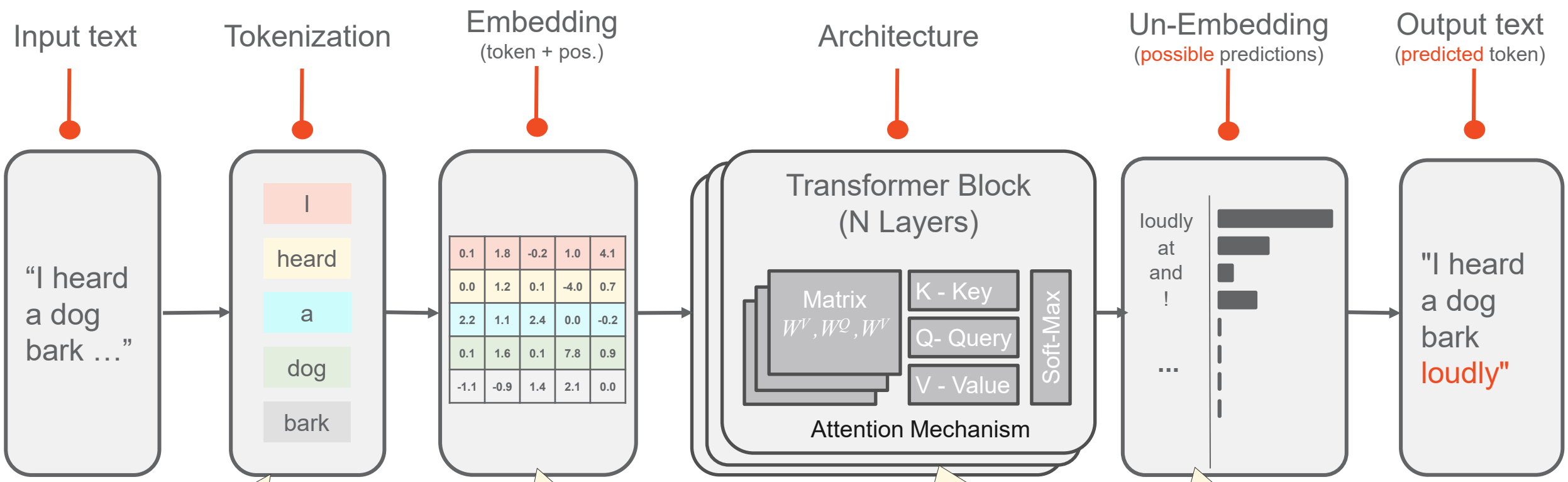
- good/bad specs generate good/bad results
- bad input may propagate undetected

- use with care for innovative approaches 😳
- don´t overload humans with „AI slop“ 😡
- more care needed towards SOP 🧐
- sign-off remains human responsibility

„Abstraction“ inside a CNN



Abstraction Errors in LLMs



Boundary & segmentation error, loss of multi-word, subword structure, ambiguity of punctation ⇒ systematic bias

Conflated similarity (numerically close but semantically misleading), domain mismatch, bias, loss of structure, stability & version issues

Draws “Conclusions” from Statistical Correlation, no Causal Reasoning

Rare but dangerous edge cases are likely to be overlooked

Top-probabilities can mislead interpretation, hide or exaggerate signals, hide rare but dangerous anomalies

„Physics“ of AI Failure

- Correlation over Causality
  Safety: Find most severe errors by identifying Cause-Event-Chains
- Trained to prioritize the most probable …
  Safety: Precisely the rare edge cases can be most dangerous.
- Trained to predict the expected
  Safety: Important is to predict the unexpected
  Safety: From tiny deviations conclude to underlying systematic faults
- Interpolation on sparse data leads to hallucination
  Safety: Unkown/uncertain areas must be identified and investigated

Quick Shot – AI for Safety Tasks …

Work product of ISO 26262	Type of Work Product: Plan	Requirements	Design & Architecture	Implementation/Test	Verification & Analysis	Confirmation	Type of Safety Task: create ideas	create from scratch	create from template	create from spec	formal review	content review	thorough verification	deep analysis	generate traceability	check traceability
Safety plan	X															
Item definition		X														
Hazard analysis and risk assessment report					X											
Functional safety concept		X	X													
Technical safety concept		X	X													
System architectural design			X													
Hardware architectural design			X													
Software architectural design			X													
Hardware safety requirements specification		X														
Software safety requirements specification		X														
System design specification			X													
Software unit design specification			X													
Software unit implementation				X												
Hardware design specification			X													
Integration test specification					X											
System test specification					X											
Software unit test specification					X											
Test Implementation				X												
Verification report					X											
Validation report					X											
Safety analysis report (e.g. FTA/FMEA)					X											
Confirmation review report						X										
Functional safety assessment report						X										
Confirmation measures plan	X					X										
Safety case						X										
Production, operation, service and decommissioning pla	X															
Configuration management plan	X															
Verification plan	X															

"Red" regions - Most critical area for *"Physics" of AI Failure*

This area requires
- thorough causal thinking
- understanding of physical and logical relations
- precise language not misguided by "most likely"
- thorough understanding of rare cases and edge cases
- capability to recognize systematic error patterns from small deviations
- real world ground truth

"Yellow" regions – Take care to avoid *"Physics" of AI Failure*

This area requires
- thorough verification of AI content
- must scale with amount of AI generated content
- potentially "diverse" tools & approaches for verification
- AI generalization must be checked thoroughly
- "looks good" but is likely to be incomplete, inconsistent or even wrong

What about …

? Tool Qualification ?

What about Safety-Competence?

How to detect AI-errors?

How to verify the results?

How to find dangerous faults not detected by AI?

CAN I RELY ON THE AI-TOOL?

What if the AI-tool hallucinates?

… and who takes the Responsibility?

ISO 26262 Tool Classification & “Qualification“

Risk \ Measure		Tool Error Detection (confidence to detect malfunction)		
		TD1 (high)	TD2 (medium)	TD3 (low/ unknown)
Dangerous Tool Impact	not possible TI1	TCL1	TCL1	TCL1
	possible TI2	TCL1	TCL2	TCL3

not dangerous, no problem - beneficial also for QM

dangerous, but easily to detect, no problem

dangerous, not easy to detect …

Potentially possible … but …

- AI-tooling by far too complex for classic tool qualification approach
- today´s ISO 26262 approach cannot be applicable 1:1
- *“Physics” of AI Failure* may cause weird
- …k box
- …cal reasoning
- …an input
- …

„Tool Qualification“ requires Approaches going far beyond today´s (and tomorrow´s) ISO 26262

Some Hints – How to use AI for Functional Safety

- Mark AI-assisted content distinguishable from human created content
- Ensure Reproducibility
- Document AI-Usage Guidelines
- Use AI capabilities for concise language

- Ensure that all critical decisions & content are under human responsibility that humans have efficient tools and measures to thoroughly verify AI content

- Only good Input Specs result in good Output
- Care about „what makes sense“, not „what looks good“

- Safety Management
  - ☺ generate and maintain plans
  - ☺ manage tickets and check for concise, clear content
  - ☺ check for consistency; check compliance to safety rules
  - ☹ AI closing tickets or confirming correct completions without humans

- ISO 26262 Work Products
  - ☺ use AI to generate initial WPs
  - ☹ do not generate Paper Tigers

- Bi-directional Traceability
  - ☺ use AI assistance to find and check traces
  - 😐 generate blindly thousands of links
  - ☹ trust completeness & consistency

- Implementation
  - ☺ use AI to generate SW following known patterns
  - ☺ verify thoroughly
  - ☹ never trust anything blindly

- Verification & Analysis
  - ☺ complement known test cases
  - 😐 complement req/tests with equivalence classes
  - 😐 complement missing tests from requirements and existing patterns
  - ☹ do not assume coverage of edge cases
  - ☹ do not assume coverage of edge cases
  - ☺ analyze carefully Cause-Event-Chains carefully

- Before SOP
  - ☺ ensure that all critical AI content has been verified thoroughly
  - ☺ prepare safety case with AI but …
  - 😐 … check safety argumentation thoroughly, create critical parts manually
  - 😐 thoroughly check content completion
  - ☹ AI must not sign-off
  - ☹ do not assume coverage of edge cases

The Red Line

Never let AI sign-off

?

⇒

?

Do not confuse correlation with causal reasoning

Do not overload humans with „AI slop“*)

Do not generate paper tigers

Never (t)rust results just because they „look good“

*) Linus Torvald, Jan26, Linux Kernel Mailing List

Conclusion

- AI can bring Open Source & Functional closer together
- AI tooling can be a key enabler & powerful accelerator but …
  - … it can introduce a new flavor of tricky, hidden and potentially dangerous faults
  - ... “good looking” results may hide wrong abstractions & conclusions
  - … statistical estimation of what is right is not enough
- Understand the *“Physics” of AI Failure*
- Finding AI faults is much more difficult than finding human errors
  ⇒ requires thorough safety expertise
- Huge amount of “AI slop” cannot be verified by human review

What AI cannot solve … **Responsibility remains with Humans**

Many Thanks for Your Attention!

florian.bogenberger@exida.com

Image Credits

Some illustrations in this presentation were generated using generative AI tools,
including **ChatGPT (OpenAI), Perplexity** and **Midjourney**, and were further selected and refined by the author.