Structuring the Safety Argumentation for DNN Based Perception in Automotive Applications

Gesina Schwalbe (*speaker*), Bernhard Knie (*discussion*)

Sep 15, 2020, WAISE 2020

Gefördert durch:

Bundesministerium
für Wirtschaft
und Energie

aufgrund eines Beschlusses
des Deutschen Bundestages

Ein Projekt entwickelt von der
VDA Leitinitiative
autonomes und vernetztes Fahren

KI-Absicherung – Safe AI for Automated Driving

Joint **research project** (07/2019 – 06/2022)

- *Vision:*

  We want to achieve that
  highly automated vehicles can be equipped with
  safety-approved AI-based systems.

- *Funding authority:* **German Federal Ministry** for Economic Affairs and Energy (BMWi)
- *Project partners:* **25 total** (automotive manufacturers, suppliers, technology providers, academic research partners)
- *Goals:* Training & test methods, **safety argumentation strategy**, industry consensus

Problem

- New challenges for safety case:
  - **Open world** problem
  - Specific **DNN insufficiencies**:
    DNNs are not hand-crafted and might learn **counterintuitive features**/relations/rules!

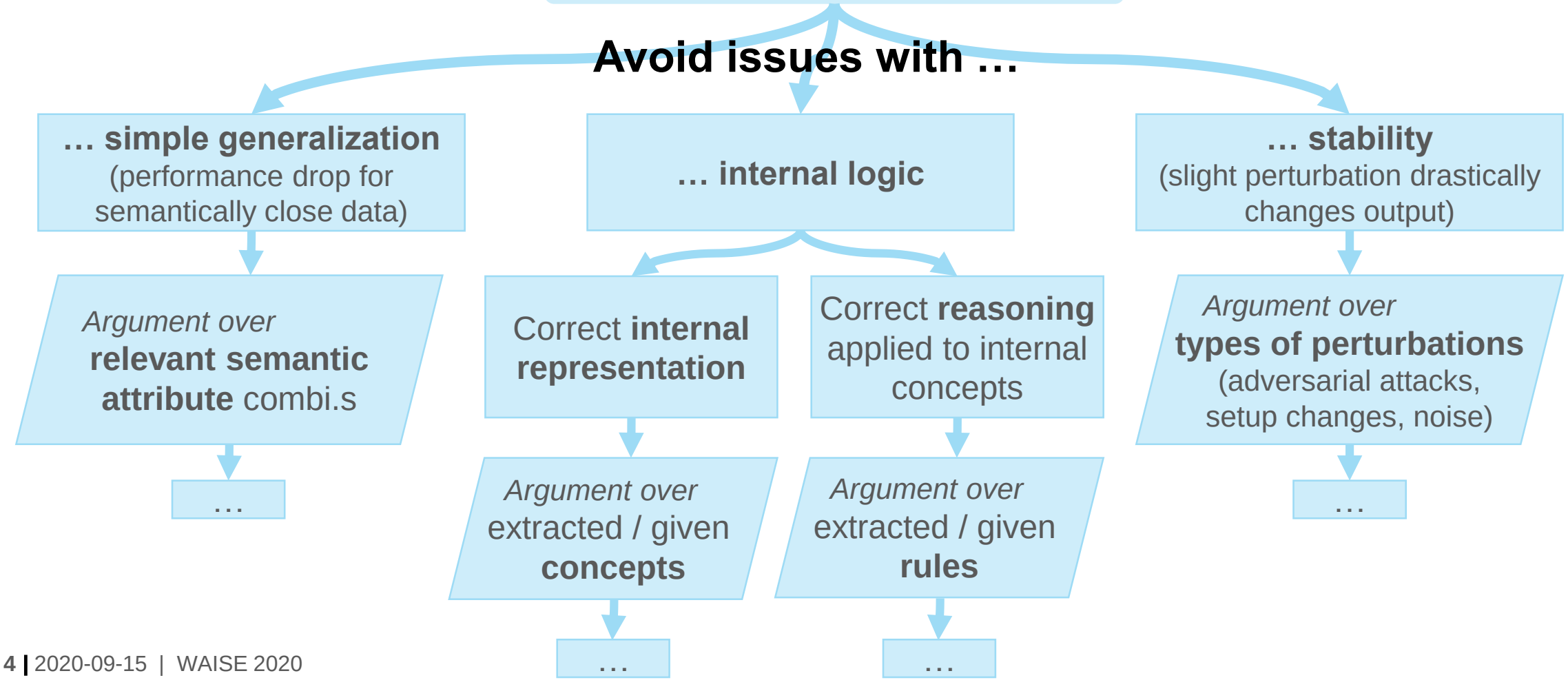
Properties of trained DNN models inherent to their technology, and with negative impacts for the use in safety-critical systems

- Existing literature / standards either of:
  - **Bottom-up:** Best practices, DNN-specific analysis methods
  - **Top-down:** domain/DNN specific hazards/failures

- **Needed: *complete* argumentation line!**

Top-down: DNN-specific Safety Requirements

DNNs not tied to human intuition → **identify & cover DNN insufficiencies**



Bottom-up: Required Evidences

Mechanisms during creation

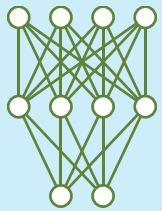
Specification & guidelines for:

- DNN design
- **Dataset** collection
- **Training**

Little experience available
→ Reliability not proven

Mechanisms on system level

Detect & prevent at:



Input

Internal state

Output

Complex internal state:
→ Errors hard to catch

Verification

- Testing
- (Semi-)formal
- Inspect **via proxy**

Validation

via traditional validation testing
(e.g. endurance run)

Ensure test data representativity

cover:

Experience

Semantic features

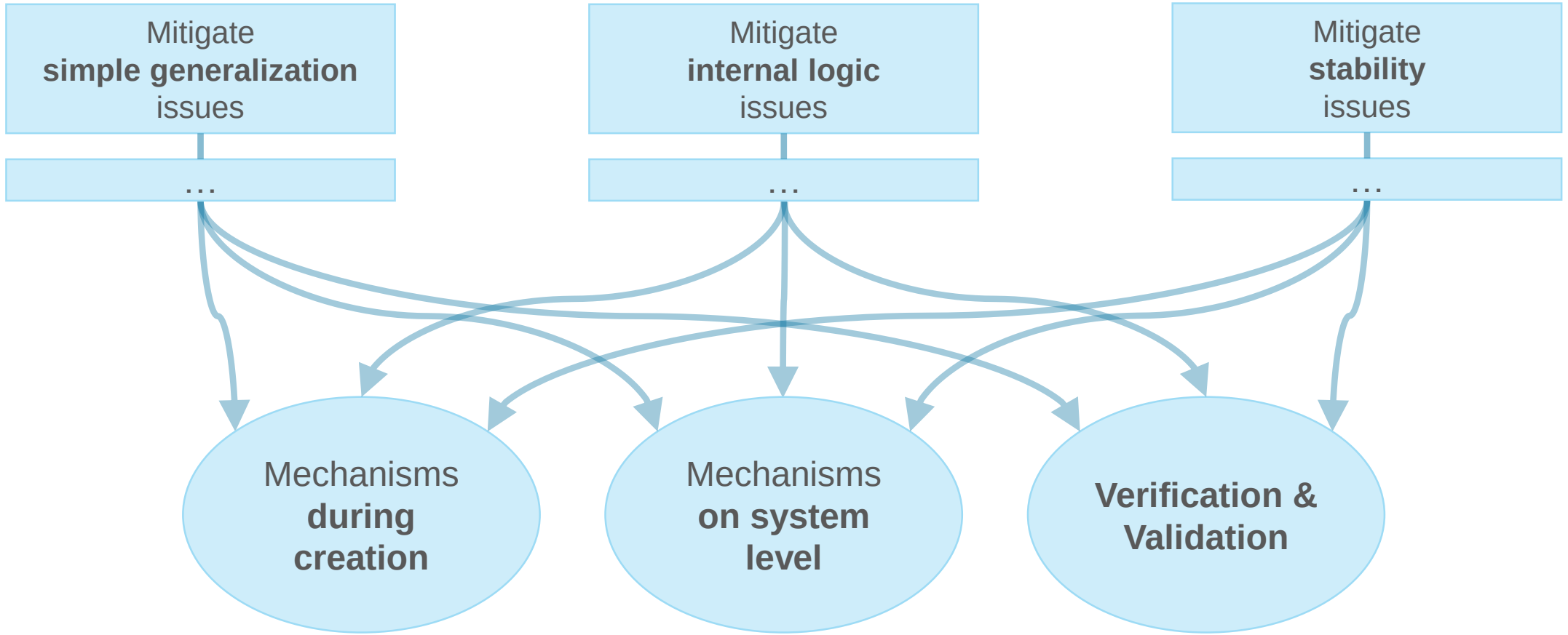
Learned features

Open world & complex internal state
→ Hard to achieve coverage

→ **All needed!**

Putting it all together …

Matrix: All requirements need all types of evidence



Thanks for your attention!

Contact:

- *Speaker:* Gesina Schwalbe (gesina.schwalbe@continental-corporation.com)
- *Discussion:* Bernhard Knie (bernhard.knie1@volkswagen.de)