fortiss

cogitron

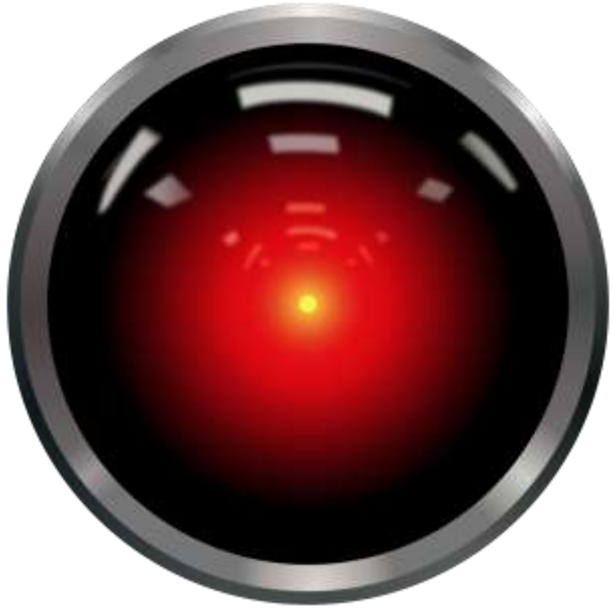
Dr. Henrik J. Putzer
(fortiss & cogitron)

Ein strukturierter Ansatz für verlässliche KI

- Standardisierung
- Autonome/Kognitive Systeme
- Safety & Security & <X>
- KI-basierte Komponenten

Singularität der KI

Maschinen übernehmen die Weltherrschaft



HAL 9000

Quelle: www.wikipedia.de

Künstliche Intelligenz ist unter uns: Schwache KI

Quelle: www.npr.org

Quelle: www.youtube.de

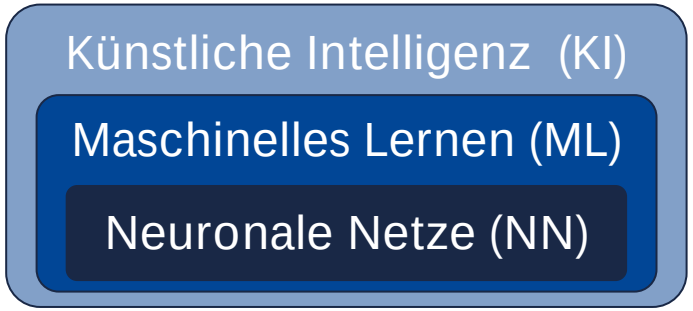
- KI besiegt weltbesten Go-Spieler (Lee Sedol vs. DeepMind)
- KI besiegt Profi-Spieler in Starcraft II (DeepMind)
- Ausassistent erledigt Aufgaben auf Zuruf
- Assistenten erkennen Verkehrsschilder und kritische Situationen (Fußgänger)
- **„schwache KI“ (Produkte) vs. „starke KI“ (Forschung)**

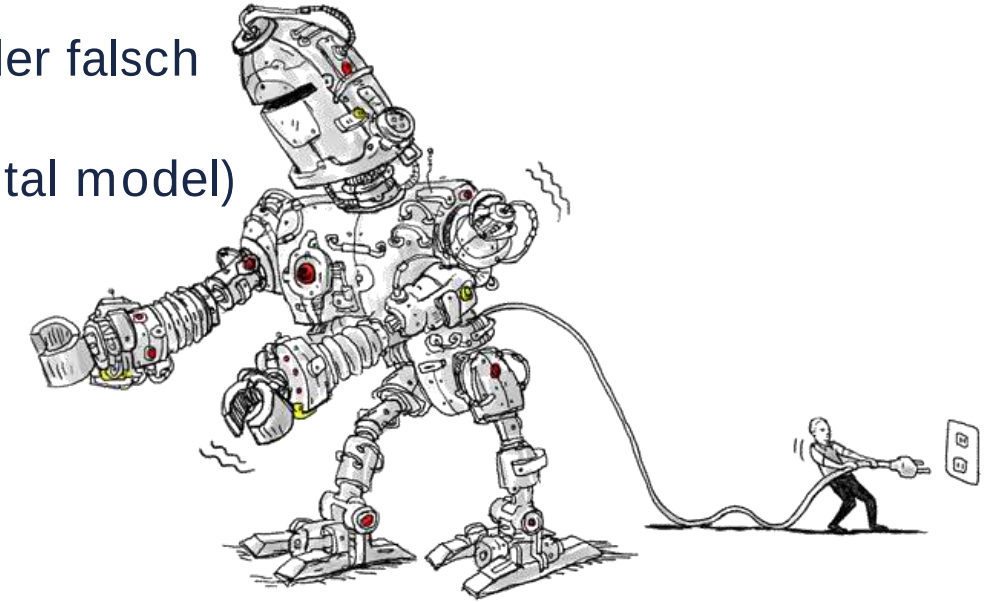
Quelle: www.amazon.de

Quelle: www. mobileye.de

Neuer Engineering-Ansatz: Neuronale Netze

- Singularität ist nicht unser Problem (aktuell)
  - NN ≠ Mystik
  - NN = Engineering-Methode (implizite Anforderungen aus Daten)
  - NN-Technologie (& Eigenschaften) noch nicht ganz verstanden
- Probleme mit Automation übertragen auf NN [Billings, 1997]
  - **complexity** –
    NN übernimmt Regelung; zu komplex für Menschen
  - **brittleness** –
    NN-Verhalten jenseits der Betriebsgrenzen undefiniert oder falsch
  - **opacity** –
    NN wird vom Nutzer nur unvollständige verstanden (mental model)
  - **literalism** –
    NN erfüllt Aufgabe ohne Kontext und Ziele
- Probleme der Praxis (akut)
  - **Zulassung** selbstfahrender Autos
  - **Qualifizierung** medizinischer Nano-Roboter



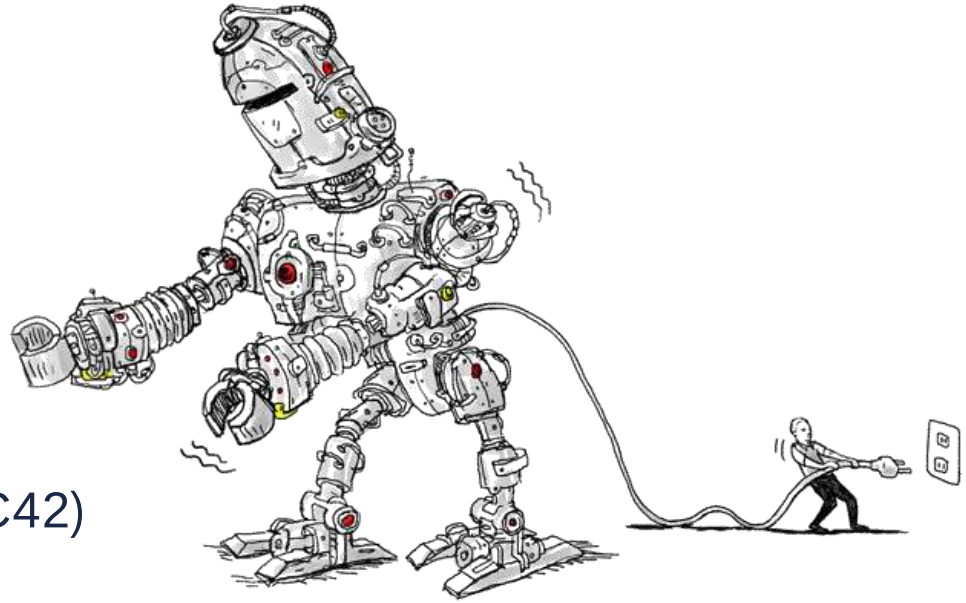


Quelle: www. gizmodo.com.au

Bestandsaufnahme DKE/VDE AK 801.0.8

Schwache KI:
Ethische Verantwortung
nicht in der KI
sondern beim Anwender

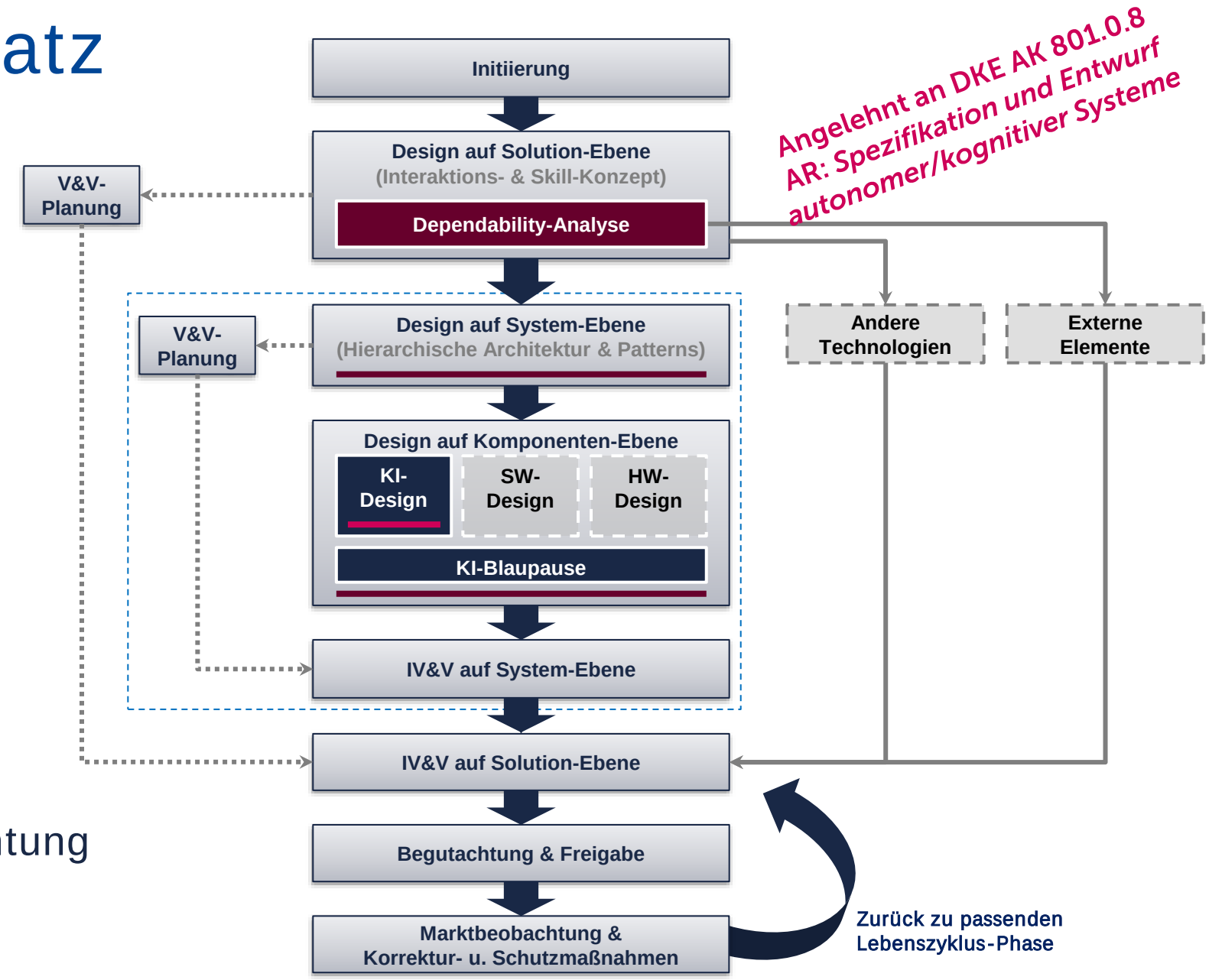
- Wo wir herkommen
  - KI | NN ist eine neue Engineering-Methode für Automation
  - KI | NN ist noch nicht mit allen Eigenschaften verstanden
  - KI | NN übernimmt sicherheitsrelevante Aufgaben aber ist nicht qualifizierbar
- Wo wir stehen
  - Neues Paradigma: autonome/kognitiven Systemen (mit KI | NN -Elementen)
  - Prozesse, Methoden, Maßnahmen, Werkzeuge
  - Kooperation von
    - **Forschung**: Fragen klären & KI verstehen
    - **Industrie**: Pilotprojekte & Anwendbarkeit
    - **Normung**: destillieren & dokumentieren
  - Laufender Entstehungsprozess (schnelllebiges Thema)
- Wo wir hinwollen
  - Anwendungsrichtlinie erstellen (ggf. mehrstufig)
  - Zur Standardisierung beitragen (z.B. DIN EN 61508, ISO SC42)



Quelle: www. gizmodo.com.au

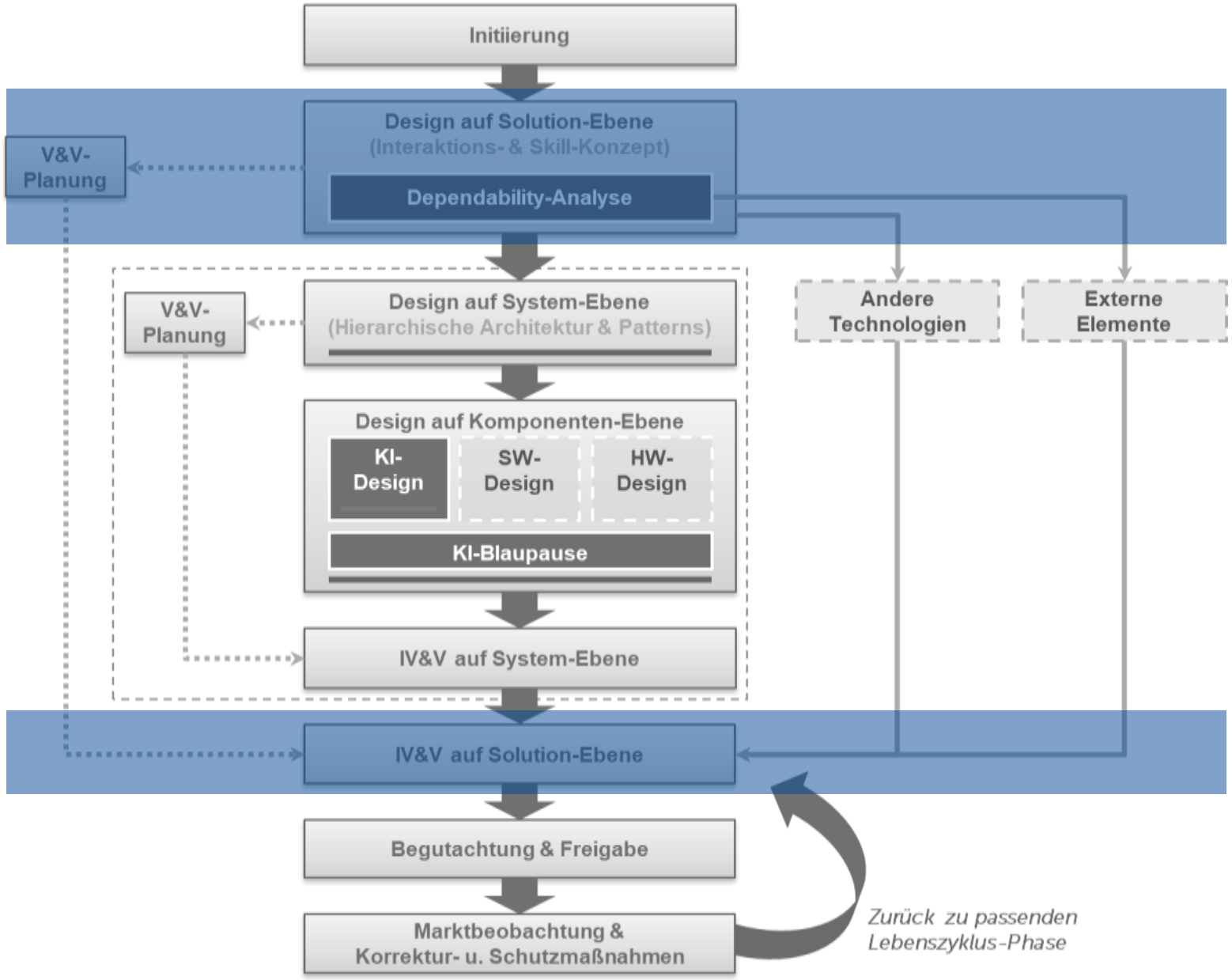
Strukturierter Ansatz
Lebenszyklus

- Entwicklungs-Lebenszyklus
  - Generischer Rahmen (Safety, Security, Usability, ...)
  - Leicht integrierbar (z.B. IEC 61508)
- Grundlegende Erweiterungen
  - Solution-Ebene
  - System-Ebene
  - KI-Design
  - Markteintritt & Marktbeobachtung



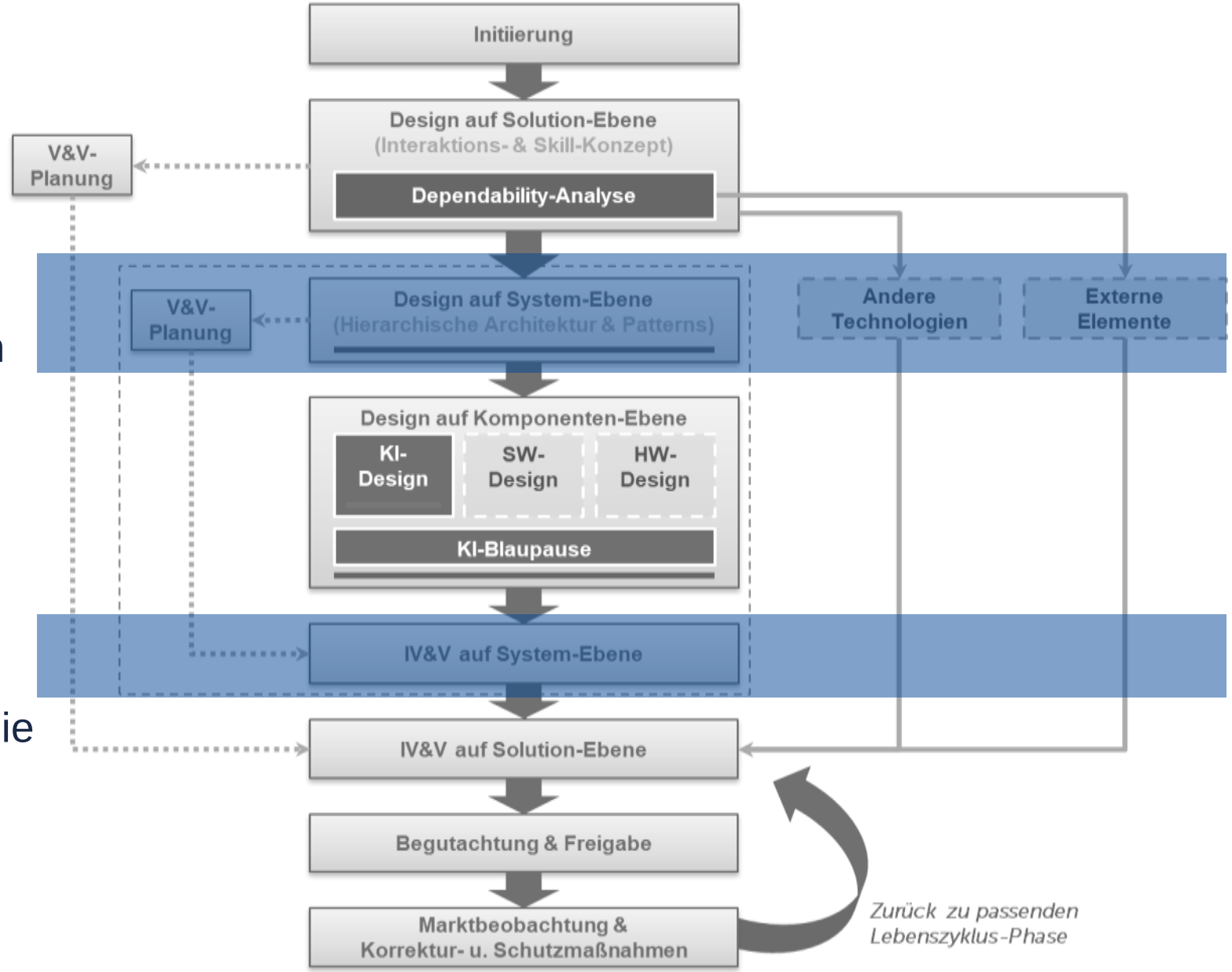
Solution-Ebene

- Solution-Konzept
  (KI ist nur ein Element)
  - Interaktions- & Kooperations-Konzept
  - Skill-Konzept
- Dependability-Analyse
  - Aspekte festlegen
    (Safety, Security, Usability, ...)
  - Bestehende Standards nutzen
- Dependability-Konzept
  - Gegenmaßnahmen festlegen
  - Maßnahmen integrieren in Solution-Konzept



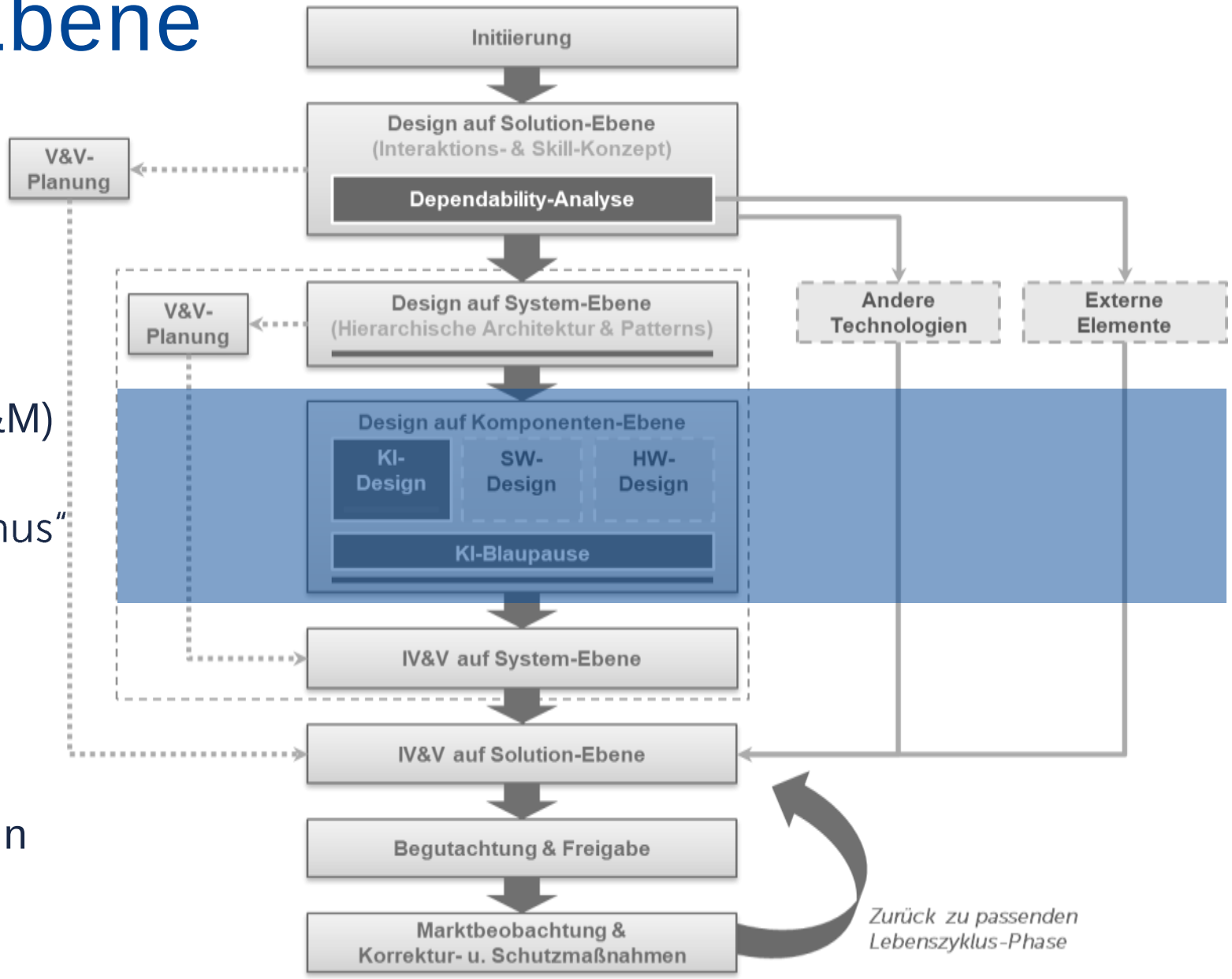
System-Ebene

- Hierarchisches Design (state-of-the-art)
  - Elemente detaillieren
  - Dependability-Anforderungen zuordnen
- KI-Design-Pattern
  - Redundanz
    - Klassische Technologien
    - Gleiche Technologie (NN-Ensembles)
    - Unterschiedliche Technologie (Konnektionismus & Symbolismus)
  - Monitoring
  - Online-Verifikation

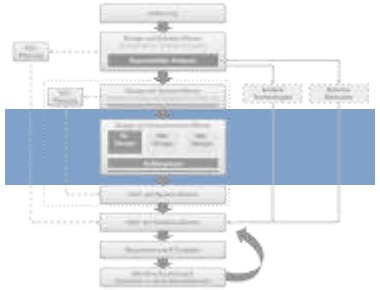


Komponenten-Ebene

- Komponente = HW + SW + KI
  - KI als weitere Technologie
  - SIL_{AI} = Auswahl von Maßnahmen & Methoden (M&M)
  - λ_{AI} = operationalisiert den „gefühlten Nicht-Determinismus“
- KI-Blaupause
  - Legt (Teil-) Prozess fest
  - Spezifisch je KI-Technologie
  - Erfüllt Anforderungen an Schnittstellen und Qualifikation („plug & play“ für Prozesse)



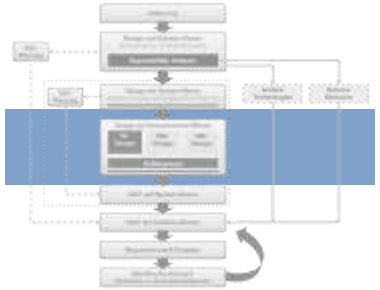
Komponenten-Ebene



- Komponente = HW + SW + KI
  - KI als weitere Technologie
  - SIL_{AI} = Auswahl von Maßnahmen & Methoden (M&M)
  - λ_{AI} = operationalisiert den „gefühlten Nicht-Determinismus“
- KI-Blaupause
  - Legt (Teil-) Prozess fest
  - Spezifisch je KI-Technologie
  - Erfüllt Anforderungen an Schnittstellen und Qualifikation („plug & play“ für Prozesse)

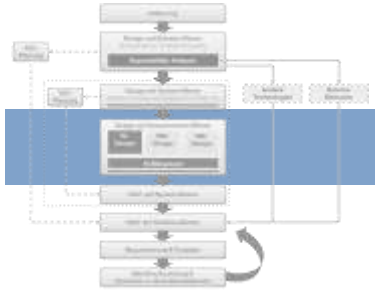
Fehlertyp	Maßnahme	HW	SW	AI
systematisch	Prozess, M&M, Ausbildung	SIL	SIL	SIL_{AI}
zufällig	*„Ausfallrate“*	λ	- -	λ_{AI}

Komponenten-Ebene & KI-Design

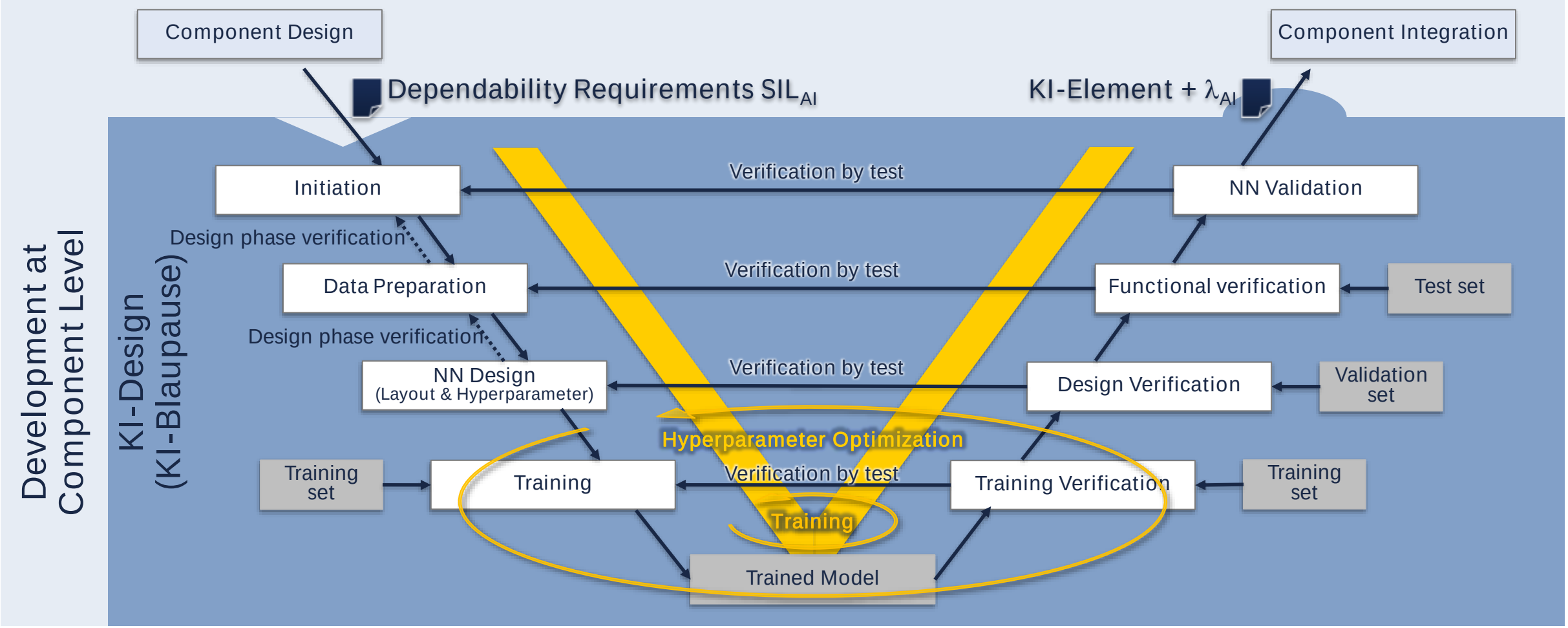


- Beispiel für eine KI-Blaupause: Supervised Neural Network Blueprint

KI-Design



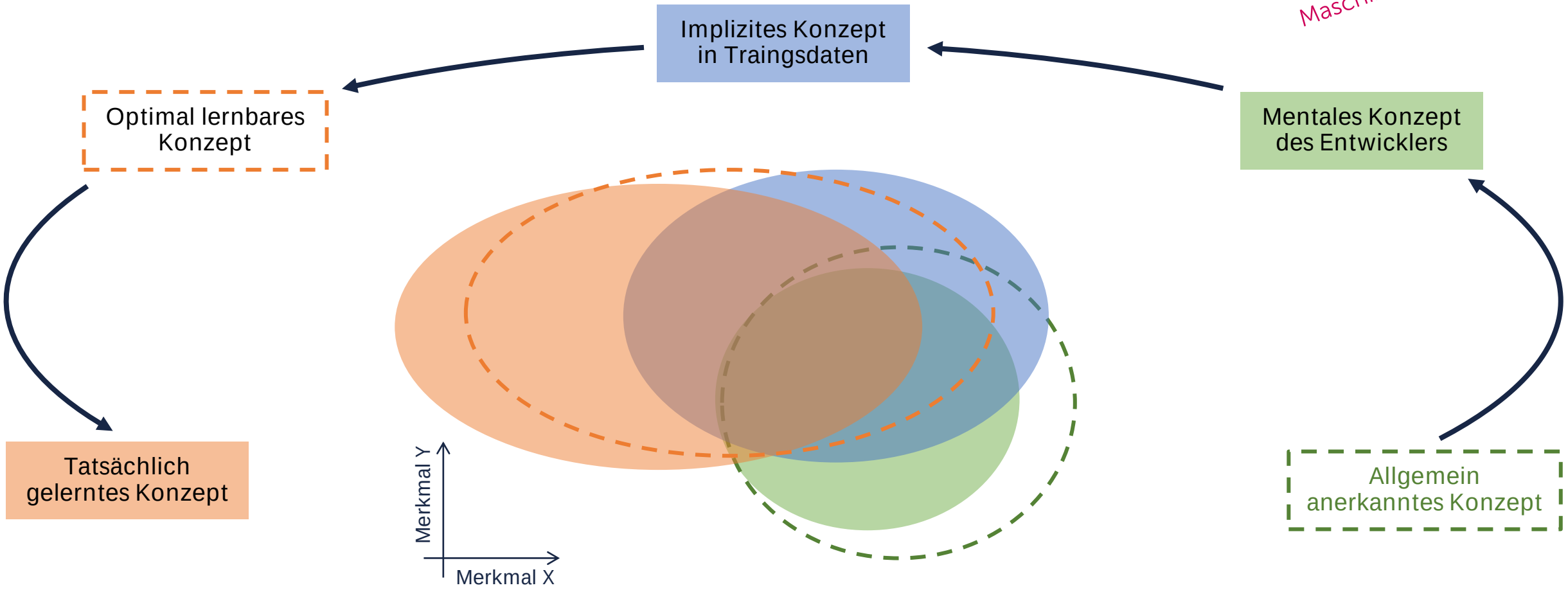
- Beispiel für eine KI-Blaupause: Supervised Neural Network Blueprint



Gefühlter Nicht-Determinismus und λ_{AI}

Analyse

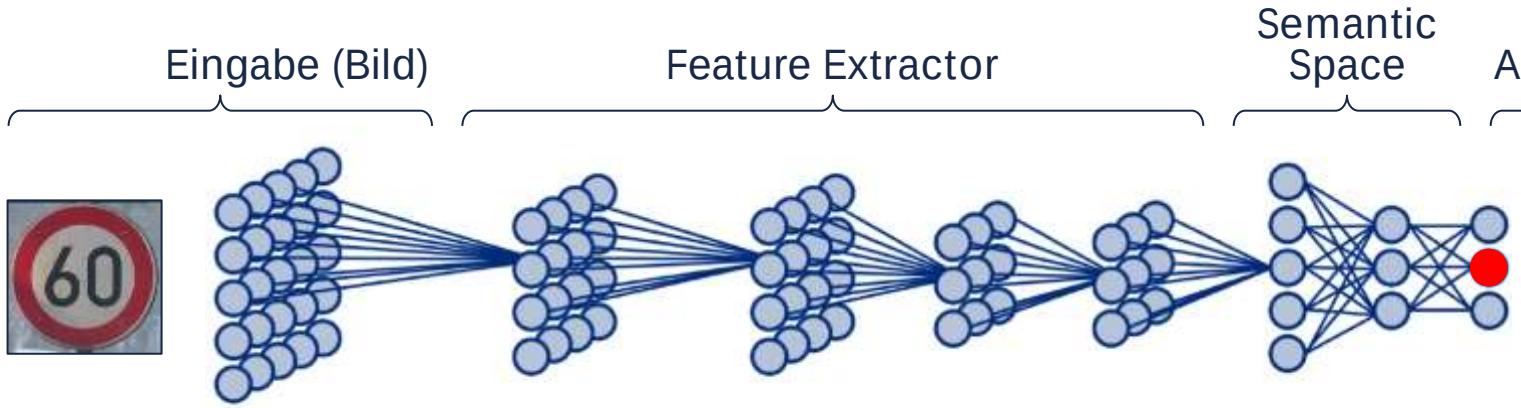
Beispiel:
überwachtes,
konnektionistisches,
Maschinelles Lernen



cogitron fortiss

Gefühlter Nicht-Determinismus und λ_{AI}

Beispiel: Verkehrszeichen-Erkennung



(siehe auch: https://distill.pub/2018/building-blocks/)

NNs sind immer ein **Kompromiss** zwischen **Genauigkeit und Robustheit**

Für „unknown unknowns“ ist Situationswissen und semantisches Wissen nötig.

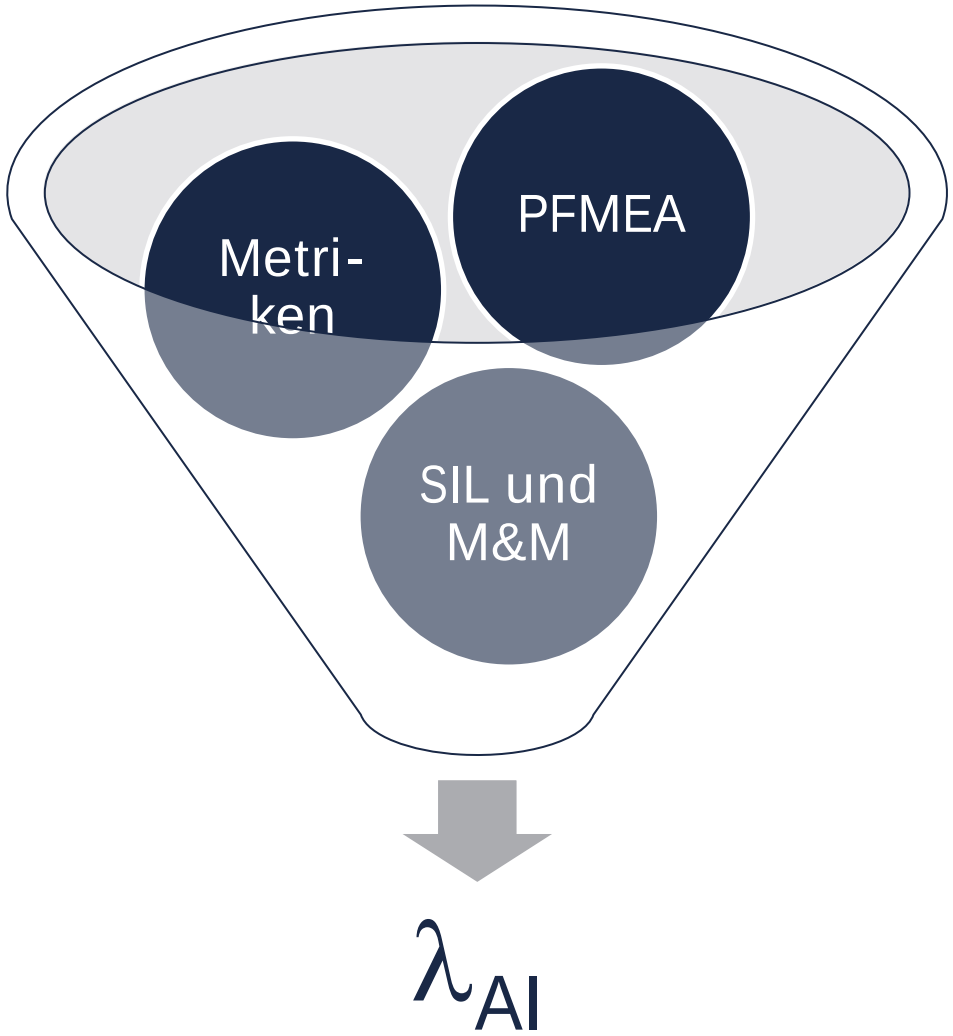
known unknowns | unknown unknowns

Quelle: www. google.de

Gefühlter Nicht-Determinismus und λ_{AI}

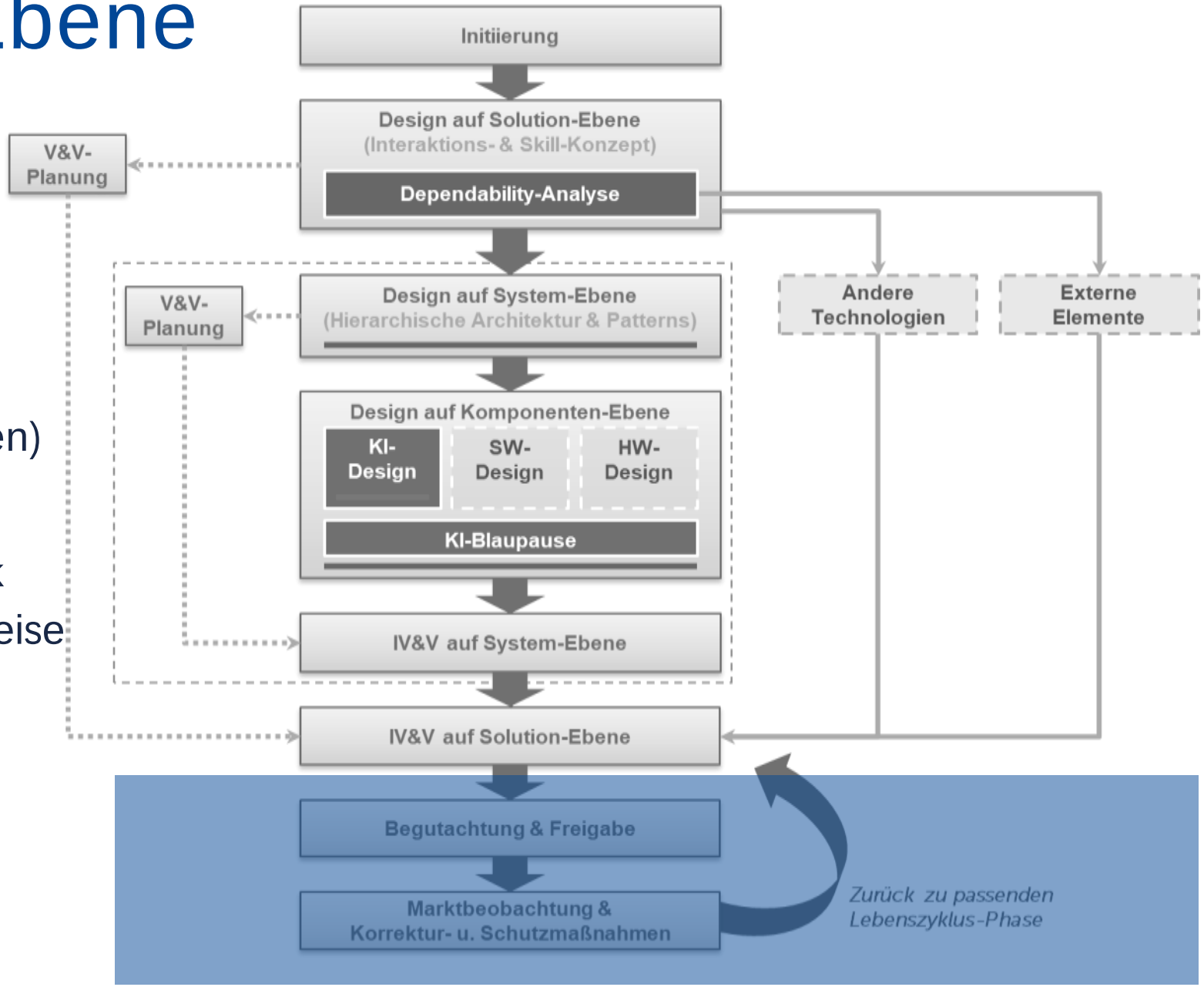
Ansatz

- Berechnungsmethode λ_{AI}
  - Individuell je KI-Blaupause
  - Einheitliche Vorgaben (Vergleichbarkeit)
  - Quellen: **Fehlermöglichkeiten** aus
    - Gewählten Maßnahmen & Methoden (M&M)
    - Erstellungsprozess (Prozess-FMEA)
    - Genutzte Metriken und deren Werte
    - und Fehlermöglichkeiten
- Offene Fragestellungen
  - Forschung: Metriken und M&M
  - Industrie: Erfahrung & Vergleichbarkeit
  - Normung: „good practices“ & Struktur



Komponenten-Ebene

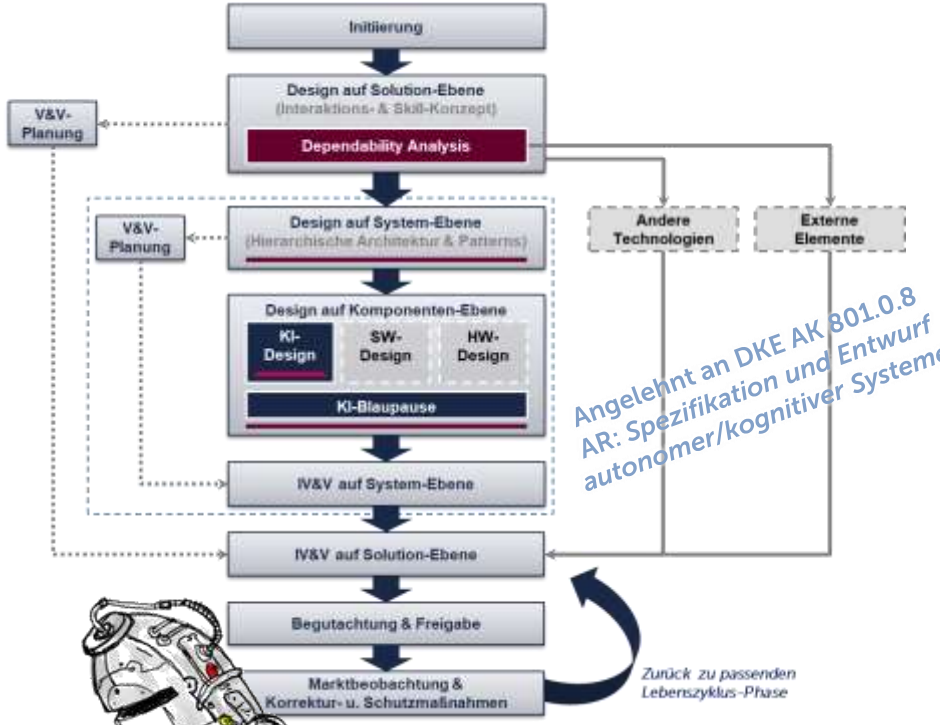
- Markteintritt
  - Begutachtung
  - Risk-Benefit-Analyse
  - Freigabe (ggf. Einschränkungen)
- Marktbeobachtung
  - Ähnlich CAPA, Medizintechnik
  - Beobachtung der Nutzungsweise (Anteile von SOTIF)
  - Kontinuierliches Lernen
  - Security: Erkennung neuer Angriffe & Gegenmaßnahmen



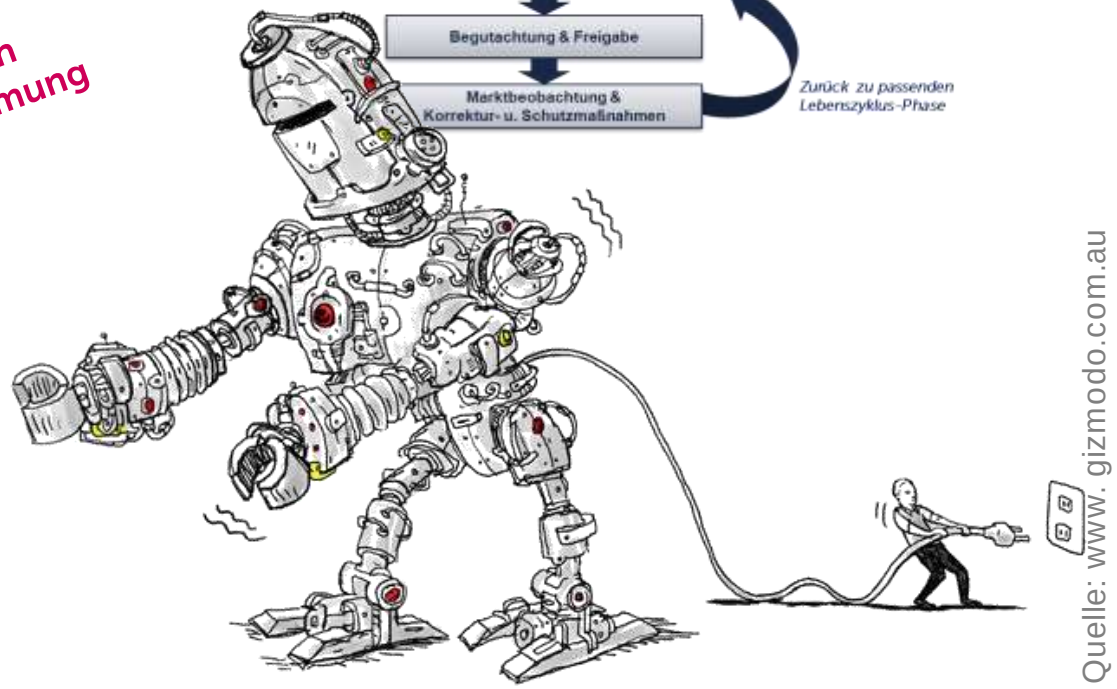
Zusammenfassung

- Die Zukunft wird von KI geprägt sein.
  Diese KI muss verlässlich sein.
  (Engineering, Produkte)
- Strukturierter Ansatz für KI-Entwicklung nötig.
  (Lebenszyklus, Entw.prozess, Maßnahmen&Methoden)
- VDE/DKE AK 801.0.8 arbeitet daran.
  - Solution-Ebene
  - System Level
  - KI-Design
  - Markteintritt & Marktbeobachtung
- Gestalten Sie die Zukunft der KI mit!
  - Beiträge: Forschung, Pilotprojekte, Finanzierung
  - fortiss & cogitron sind dabei – und Sie?

Zusammenarbeit nötig von
Forschung + Industrie + Normung



Angelehnt an DKE AK 801.0.8
AR: Spezifikation und Entwurf
autonomer/kognitiver Systeme



Quelle: www.gizmodo.com.au

Danke für Ihr Interesse.

Fragen?

Contact

fortiss GmbH
Guerickestraße 25
80805 Munich
GERMANY
www.fortiss.org
info@fortiss.org

Dr. Henrik J. Putzer
putzer@fortiss.org
henrik.putzer@cogitron.de

Presentation by

Dr. Henrik J. Putzer is the head of autonomous systems and sensor systems at fortiss. On the basis of a computer science diploma (correlated to embedded systems) and a degree in aeronautical and space technology (correlated to cognitive systems, AI and human factors) he gained several decades of practical industrial experience as a consultant and certified SPICE assessor for embedded systems and processes cross industries. He was part of the ISO 26262 core team and contributed to other standards. Latest contributions include the DKE AR 801.0.8 "Spezifikation und Entwurf autonomer/kognitiver Systeme" as well as the autonomous system research, AI topics and their application to the autonomous car and drone platforms at fortiss.