



Roadmap

Safety, Security, and Certifiability of Future Man-Machine Systems

Ergebnisse des SafeTRANS-Arbeitskreises Resilient, Learning, and Evolutionary Systems

Ko-Leitung: Prof. Dr. Werner Damm (SafeTRANS/OFFIS Institut für Informatik), Peter Heidl (Robert Bosch GmbH)

Impressum

SafeTRANS e.V.
Escherweg 2
26121 Oldenburg

Tel.: +49-(0)441 / 9722 503
Fax: +49-(0)441 / 9722 502
E-Mail: info@safetrans-de.org
Website: www.safetrans-de.org

Design: Franziska Griebel, SafeTRANS
Bildnachweise: Titelseite: agsandrew/fotolia, Seite 7: tawanlubfah/fotolia, Seite 24: Grafik-Montage SafeTRANS, Bilder Adobe Stock
Datum: Dezember 2019

Roadmap



Safety, Security, and Certifiability of Future Man-Machine Systems

Inhalt

Executive Summary	11	5	Verantwortlichkeit und Reflexion	75
Teilnehmende Organisationen	17	5.1	Zusammenfassung	75
1 Motivation und Einleitung	19	5.2	Detaillierte Beschreibung	76
1.1 Motivation	19	5.2.1	Analyse der Veränderungen von Verantwortung	76
1.2 Einleitung	21	5.2.2	Capabilities	76
2 Systemstärke	25	5.2.3	Herausforderungen und Forschungsfragen	77
2.1 Zusammenfassung	25	5.2.4	Benötigte Technologien und Zeithorizont	79
2.2 Detaillierte Beschreibung	26	6 Integrität und Zertifizierung	83	
2.2.1 Analyse	26	6.1 Zusammenfassung	83	
2.2.2 Capabilities	28	6.2 Detaillierte Beschreibung: Integrität	86	
2.2.3 Herausforderungen und Forschungsfragen	31	6.2.1 Analyse	86	
2.2.4 Benötigte Technologien und Zeithorizont	33	6.2.2 Capabilities	91	
3 Kontext	41	6.2.3 Technische Herausforderungen und Forschungsfragen	91	
3.1 Zusammenfassung	41	6.3 Detaillierte Beschreibung: Absicherung	93	
3.2 Detaillierte Beschreibung	42	6.3.1 Analyse	93	
3.2.1 Analyse	42	6.3.2 Capabilities	94	
3.2.2 Capabilities	46	6.3.3 Herausforderungen und Forschungsfragen	95	
3.2.3 Herausforderungen und Forschungsfragen	50	6.4 Benötigte Technologien und Zeithorizont	99	
3.2.4 Benötigte Technologien und Zeithorizont	52	7 Roadmap	109	
4 Kooperation	59	8 Handlungsempfehlungen	115	
4.1 Zusammenfassung	59	9 Referenzen	117	
4.2 Detaillierte Beschreibung	60			
4.2.1 Analyse	60			
4.2.2 Capabilities	66			
4.2.3 Herausforderungen und Forschungsfragen	67			
4.2.4 Benötigte Technologien und Zeithorizont	69			

Vorwort

Seit mehr als einem Jahr bereitet SafeTRANS in seinem Arbeitskreis über Resiliente, Lernende, und Evolutionäre Systeme eine Fortschreibung der Nationalen Roadmap on Embedded Systems (NRMES)¹ vor, in der für zukünftige Mensch-Maschine-Systeme (Cyber-Physische Systeme, die für und mit Menschen zusammen arbeiten) in vier Aggregationsstufen² solcher Systeme die zentralen Entwicklungsperspektiven entlang von fünf Komplexitätsdimensionen³ charakterisiert werden. Aufbauend auf den Kernkompetenzen von SafeTRANS steht dabei die Frage der **Nachweisbarkeit der Beherrschung der Komplexität** solcher Systeme im Vordergrund. Das vorliegende Dokument beschreibt detailliert diese Aggregationsstufen sowie die identifizierten Komplexitätsdimensionen, identifiziert Forschungs- und Entwicklungsherausforderungen sowie die zur Lösung dieser Herausforderungen benötigten zentralen Technologien und ordnet diese zeitlich gemäß den Ausbaustufen in den einzelnen Komplexitätsdimensionen an. Begleitet wird dieses Dokument durch ein Positionspapier⁴, das die wesentlichen Erkenntnisse dieser Roadmap zusammenfasst.

Mit der Nationalen Roadmap Embedded Systems und folgenden Forschungsroadmap-Dokumenten⁵ hat SafeTRANS bereits wegweisende Beiträge zu den Forschungs Herausforderungen zu eingebetteten und cyber-physikalischen Systemen geleistet, die in ihrer Einbeziehung der Mensch-Maschine Interaktion und -Kooperation sowie der Vernetzung mit der Cloud wegweisend waren und sind, und die sowohl nationale wie europäische Förderstrategien nachhaltig beeinflusst haben.

In einer nun – 10 Jahre nach Erscheinen der NRMES – erfolgenden Rückschau können wir feststellen, dass die dort getroffenen Vorhersagen in nahezu allen Punkten eingetroffen sind. Im Rückblick kann dabei bestätigt werden, dass – basierend auf einer hinreichend langen Erfahrung in der Entwicklung und den Sicherheitsnachweisen von eingebetteten Systemen – Voraussagen über die weitere Entwicklung dieser Themenfelder getroffen werden konnten, die sehr genau und präzise waren und sind.

Wir stellen jedoch auch zwei wesentliche Abweichungen der realen Entwicklung von den Voraussagen der NRMES fest: Zum einen fokussiert die NRMES – dem damaligen Entwicklungsstand für Eingebettete und Cyber-Physische Systeme entsprechend – sehr stark auf Einzelsysteme. Vernetzung und Kooperation von Systemen untereinander und mit Menschen wurden zwar als wichtige Themen

¹ SafeTRANS/ZVEI (Hrsg.). Nationale Roadmap Embedded Systems. 2009

² Individuelle (Mensch-Maschine-)Systeme, Gruppen von Systemen, homogene Kollektive von Systemen und heterogene Kollektive von Systeme

³ Systemstärke, Kontext, Kooperatoin, Verantwortung und Reflexion, Integrität und Zertifizierung

⁴ SafeTRANS (Hrsg.). Safety, Security, and Certifiability of Future Man-Machine Systems - Positionspapier. 2019

⁵ SafeTRANS (Hrsg.). Hochautomatisierte Systeme: Testen, Safety und Entwicklungsprozesse. 2017

SafeTRANS (Hrsg.). Eingebettete Systeme in der Automobilindustrie. 2015

Alle Dokumente zu finden unter: <http://www.safetrans-de.org/de/Aktivitaeten/Roadmapping.php>

erkannt; in ihrer heute sichtbaren Bedeutung sowohl in Bezug auf dadurch ermöglichte Systemfähigkeiten als auch auf die aus ihnen resultierenden Forschungs Herausforderungen wurden beide Themen jedoch unterschätzt. Zum anderen werden in der NRMES für die Themenbereiche „Nachweisbar sichere Umwelterkennung (Perzeption)“ sowie „Nachweisbar sichere Vorhersage der Weiterentwicklung von Situationen (Prediktion)“ (beide unter anderem im Forschungsfeld „Verteilte Echtzeit-Situationserfassung und Lösungsfindung“ der NRMES) deutlich zu optimistische Vorhersagen bezüglich des Zeitraums, in dem diese Themen beherrschbar sind, durchgeführt.

Zusammen tragen diese Punkte wesentlich zur Steigerung der Komplexität zukünftiger Mensch-Maschine Systeme bei; die Frage der **Nachweisbarkeit der Beherrschung dieser Komplexität** wird damit noch einmal um Größenordnungen schwieriger, als bereits in der NRMES identifiziert.

Wir erleben hier und heute, wie Teilaspekte dieser Fragestellungen zunehmend in der Fachöffentlichkeit, in öffentlichen Foren und Expertenrunden diskutiert werden:

- Wie sicher ist autonomes Fahren?
- Darf KI sicherheitsrelevante Entscheidungen treffen?
- Wie gut sind kritische Infrastrukturen vor Cyber-Attacken geschützt?



So relevant und berechtigt diese Fragen sind, so berühren sie doch nur Fragmente eines Komplexitätsraums, der in einem wahrhaft atemberaubenden Tempo durch Forschung und Entwicklung neuer Lösungsoptionen aufgerollt wird. Zukünftige Mensch-Maschine Systemen werden in der Lage sein, ihre Zielsetzungen in einer sich ändernden Umgebung umzusetzen, sei es durch den Grad an Wahrnehmung, Intelligenz- und Kooperationsfähigkeit, welche solche Systeme besitzen werden, sei es durch die Fähigkeit zur Entwicklung völlig neuer Fertigkeiten oder der Gestaltung der Umwelt, sei es durch die Fähigkeit zunehmend höhere Grade an Selbst- und Fremdreiflexion in ihr Handeln einbeziehen zu können. **Die zentrale Frage wird nicht sein, ob oder bis wann wir solche Mensch-Maschine-Systeme entwickeln können, sondern ob wir ihre Komplexität noch beherrschen.** Mit jeder Zunahme an Komplexität entlang der in der Roadmap aufgezeigten Komplexitätsdimensionen und den für sie identifizierten Ausbaustufen wird die Frage nach der Beherrschung der Systeme und ihrer (Sicherheits-)Eigenschaften brisanter, in ihrer vollen Kombination erreichen wir die Komplexität von Modellen des Klimawandels. Dem gegenüber steht ein extremer sowohl gesellschaftlicher wie auch wirtschaftlicher Nutzen, den die Entwicklung solcher Systeme unbestritten mit sich bringen wird, der jedoch nur dann realisiert werden kann, wenn die notwendigen Qualitätseigenschaften wie Sicherheit (im doppelten Sinne von Safety und Security), Verfügbarkeit und weiterer nachweisbar gegeben ist.

Von jeher hat SafeTRANS einen Schwerpunkt auf die Beherrschbarkeit der Komplexität der Systeme gelegt, die in den ursprünglich betrachteten Kerndomänen in verkehrstechnischen Applikationen sich konkretisiert nach Verfahren, Methoden, und Prozessen zum Nachweis der Sicherheit und Verfügbarkeit IT-basierter Lösungen zur Sicherung der Mobilität. Mit der SafeTRANS Roadmap für hochautonome Systeme wurde dabei der Schwerpunkt auf eine Komplexitätsdimension gelegt, und diese bereits unter Einbeziehung von Experten aus dem Energiesektor erweitert. Die durch SafeTRANS auf europäischer Ebene initiierten Groß-Projekte wie CESAR⁶, CRYSTAL⁷ oder ENABLE-S3⁸ bezogen weitere Sektoren wie die Produktionsautomatisierung und eHealth mit ein. Aufbauend darauf spiegelt diese Roadmap auch die aus diesen Projekten und Roadmaps gewonnenen Erfahrungen. Die identifizierten Komplexitätsdimensionen und -skalen sind für alle Domänen relevant, die im folgenden Sinn sicherheitskritisch, wirtschaftlich kritisch oder kritisch für die Gesellschaft sind:

- sicherheitskritisch: Fehlerhafte Verhalten oder Cyber-Attacken können Menschenleben gefährden.
- wirtschaftlich kritisch: Fehlerhafte Verhalten oder Cyber-Attacken können drastische wirtschaftliche Schäden verursachen.
- kritisch für die Gesellschaft: Fehlerhafte Verhalten oder Cyber-Attacken können KRISIS (kritische Infrastrukturen) lahmlegen.

Unsere Gesellschaft kann es sich nicht leisten, von Mensch-Maschine Systemen abhängig zu sein, die nicht einem den Grad an Kritikalität reflektierenden Grad an Nachweisbarkeit der Beherrschbarkeit der Komplexität solcher Systeme unterliegen. Dies beinhaltet als Spezialfälle schon die in verschiedenen Kontexten, z. B. den Leitprojekten des VDA oder in der Plattform Lernende Systeme, diskutierte Fragen nach Testen und Zulassung solcher Systeme. Die vorgelegte Roadmap geht jedoch weit über diese Klassen von Systemen hinaus und charakterisiert allgemein die Herausforderungen zum Nachweis der Sicherheit und Verfügbarkeit solcher Systeme.

Aus industrieller Sicht zielt die Roadmap auf eine mit dem Zusammenwachsen von Branchen notwendige ganzheitliche, einheitliche Sicht, welche Prinzipien im Nachweis von funktionaler Sicherheit, IT-Sicherheit und Verfügbarkeit bei der Einführung solcher Systeme zu berücksichtigen sind. Dies betrifft zum einen die Fortschreibung und Weiterentwicklung entsprechender Industriestandards, aber auch die Klärung von Fragen, welche Regularien und Richtlinien hier auf staatlicher Seite erforderlich sind, etwa in der Frage der Nachvollziehbarkeit von Entscheidungsfindungen KI-basierter Systeme in hohen Kritikalitätsklassen. Eine solche Diskussion kann durch diese Roadmap durch die Strukturierung in domänenübergreifende Komplexitätsdimensionen vorbereitet werden.

⁶ Projekt CESAR: Laufzeit: 2009-2012. Partner: 53. EU-Förderprogramm: ARTEMIS

⁷ Projekt CRYSTAL: Laufzeit: 2013-2016. Partner: 68. EU-Förderprogramm: ARTEMIS

⁸ Projekt ENABLE-S3: Laufzeit: 2016-2019. Partner: 71. EU-Förderprogramm: ECSEL

Executive Summary

Wie können wir sicherstellen, dass zukünftige Mensch-Maschine-Systeme, in denen Menschen mit digitalen Systemen interagieren, von der Gesellschaft akzeptiert werden? Wir fokussieren dabei auf Anwendungsklassen, die entweder sicherheitskritisch (d. h. Ausfälle oder Cyberangriffe gefährden das Leben von Menschen), wirtschaftlich kritisch (d. h., Ausfälle oder Cyberangriffe führen zu drastischen finanziellen Verlusten) oder kritisch für die Gesellschaft (d. h., Ausfälle oder Cyberangriffe führen zum Ausfall von KRITIS – kritischen Infrastrukturen) sind. Eine notwendige Bedingung für die gesellschaftliche Akzeptanz solcher Systeme ist die Gewährleistung von Qualitätseigenschaften, insbesondere funktionaler und IT-Sicherheit sowie Verfügbarkeit. Gesellschaftliche Akzeptanz verlangt jedoch mehr: Es stellen sich Fragen des Vertrauens, der Verständlichkeit und der Einsicht in die einer automatisierten Entscheidungsfindung zugrunde liegenden Wertesysteme.

Wir analysieren die Herausforderungen zur Sicherstellung dieser Eigenschaften von zukünftigen Mensch-Maschine-Systemen bezüglich zwei Kategorien, dem Vernetzungsgrad und der Komplexität. Der Vernetzungsgrad bildet die Basis der Analyse. Wir unterscheiden vier Aggregationsklassen (siehe Abbildung 3, Seite 17):

- **Einzelne HCPS** (Human-Cyber-Physical Systems/Mensch-Maschine-Systeme), wie hochautomatisierte Fahrzeuge, Züge, Flugzeuge, Schiffe.
- **Gruppen von HCPS**, die entweder ad-hoc formiert werden oder als dauerhaftes Organisationsprinzip etabliert sind (wie Platoons, ein Team aus Ärzten, Pflegepersonal und medizinischen Geräten einer Intensivstation oder eine Fertigungslinie mit Maschinen und dem betreuenden Personal).
- **Homogene Kollektive von HCPS**, welche durch die Führung des Kollektivs eine optimale Nutzung von (Anwendungsklassen spezifischen) Ressourcen ermöglichen und den Teilnehmern des Kollektivs Dienste anbieten (wie etwa Verkehrsleitplanung, Energieverbundnetze). Dabei gehören alle Teilnehmer des Systems einer Anwendungsdomäne an.
- **Heterogene Kollektive von HCPS** integrieren Teilnehmer verschiedener Anwendungsdomänen in ein Gesamtsystem, um durch dessen Führung eine holistische domänenübergreifende Ressourcennutzung und die Einhaltung übergeordneter Vorgaben (wie etwa zur Erreichung von Klimazielen) zu ermöglichen (wie etwa Smart Cities oder Krisenmanagement-Verbünde).

Wir strukturieren den Entwurfsraum für solche Systeme nach den folgenden **fünf Komplexitätsdimensionen** (teilweise mit Untergruppen/-dimensionen). Die Komplexität einer Applikation wird dabei aus Sicht des Systems gemessen, das die Applikation realisiert. In Anlehnung an eine etablierte Terminologie in der Automobilindustrie bezeichnen wir dieses nachfolgend als „Egosystem“. Diese relative Bewertung trägt dem Rechnung, dass ein und dieselbe Aufgabe für ein System leicht, für ein anderes schwer zu lösen ist, je nach Stärke des Systems. Der Begriff System bezeichnet dabei wie im obigen Abschnitt dargestellt je nach Aggregationsstufe ein einzelnes technisches System oder einen einzelnen Menschen, eine Gruppe, ein homogenes oder heterogenes Kollektiv von Menschen und technischen Systemen.

Systemstärke. In der Dimension **Stärke** messen wir Fähigkeiten des Egosystems, welche ihm erlauben, die für eine Anwendung gegebenen Zielsetzungen alleine erfolgreich zu verfolgen. Dies wird umso eher gelingen, umso stärker die **Intelligenz** des Systems ist, umso besser es sich an sich verändernde Umfeldbedingungen anpassen kann, also welche **Evolutionsfähigkeiten** es besitzt, und letztlich mit welchem Grad an **Autonomie** es ausgerüstet ist.

Kontext. In dieser Dimension messen wir die durch das System zu beherrschende Komplexität der Umgebung des Egosystems. Zur Umgebung des Egosystems zählen alle Systeme (technische Systeme, Menschen) sowie deren Eigenschaften, deren Kenntnis für die Durchsetzung der Ziele des Egosystems relevant sind (ohne deren Kenntnis eine Zielerreichung nicht möglich ist). Desgleichen zählen hierzu alle relevanten physikalischen Phänomene. Wir analysieren die Komplexität des Kontext getrennt für **physikalische Umgebungsaspekte** und **Menschen oder technische Systeme** in der Umgebung des Systems.

Kooperation. Immer dann, wenn die eigene Stärke nicht ausreicht, um in einem gegebenen Kontext seine Ziele durchzusetzen, kann das Egosystem andere Systeme oder Menschen der Systemumgebung anfragen, für einen abgestimmten Zeitraum das Egosystem in der Umsetzung seiner Ziele zu unterstützen. Wir trennen wiederum in **Kooperation mit anderen Systemen und Menschen** und in **Kooperation mit der Cloud und der Infrastruktur**; gemeinsam ist diesen Aspekten, dass die potenziellen Kooperationspartner das Egosystem tendenziell unterstützen.

Verantwortung und Reflexion. Das zunehmende Maß an Autonomie von Systemen erfordert die Festlegung eines Konzepts der Systemverantwortung: Wenn nicht mehr der Mensch die endgültige Kontrolle hat, wer übernimmt dann die Verantwortung für die Aktionen eines Systems? Im Allgemeinen beschreibt Verantwortung eine Beziehung zwischen einem Subjekt (in unserem Kontext ein autonomes System) und einem Objekt, die von den Handlungen des Subjekts beeinflusst wird, und bezeichnet die Fähigkeit des Subjekts, seine Handlungen so zu reflektieren, bewerten und auszuwählen, dass sie konform sind zu bestehenden juristischen, ethischen und moralischen Normen.

Integrität und Absicherung. Um das übergeordnete Ziel der Sicherung der Qualität trotz wachsender Komplexität zu erreichen, sind Maßnahmen zur Sicherung der Integrität und der nachprüfbarer Absicherung der Umsetzung der Qualitätsvorgaben zu ergreifen. Die Komplexitätsdimension der **Systemintegrität** widmet sich der Herausforderung, durch Konstruktion sicherzustellen, dass alle Quellen für die Entscheidungsfindung konsistent und vertrauenswürdig sind. In der Dimension **Absicherung** unterscheiden wir Ausbaustufen, die je nach Risikoklasse des Egosystems realisiert werden müssen.

Für diese Dimensionen (und Untergruppen) identifizieren wir jeweils dimensionsspezifische Skalen, um die verschiedenen Komplexitätsstufen der Dimension bzw. im Umkehrschluss die Ausbaustufen der benötigten Fähigkeiten und Fertigkeiten der Systeme zu charakterisieren. Abbildung 1 auf der folgenden Seite zeigt das so entstehende Diagramm.



Abbildung 1: Komplexitätsdimensionen

Die Roadmap ist dem entsprechend wie folgt aufgebaut: In den folgenden Kapiteln 2 bis 6 wird für jede der identifizierten Dimensionen zunächst eine ausführliche Beschreibung der Dimension, einschließlich Untergruppen und dimensionsspezifischer Skalierung gegeben. Für Leser, die ausschließlich an den Ergebnissen dieses Roadmap-Prozesses interessiert sind, ist es ausreichend, jeweils dieses erste Unterkapitel zu lesen. Die ausführliche Herleitung dieser Ergebnisse pro Dimension erfolgt dann in den weiteren Unterabschnitten dieser fünf Kapitel. Hierzu erfolgt zunächst eine Analyse der für die jeweilige Dimension charakteristischen komplexitätsrelevanten Eigenschaften. Anschließend werden jeweils die konkreten forschungs- und entwicklungsbezogenen Herausforderungen aufgelistet entsprechende Forschungsfragen identifiziert und die für die Lösung der Herausforderungen relevanten Technologien aufgelistet und bzgl. ihrer Verfügbarkeit zeitlich eingeordnet. Kapitel 7 schließlich fasst die Ergebnisse dieser fünf Abschnitte zusammen und gibt eine zeitliche Anordnung der identifizierten Anforderungen an. Es entsteht eine vollständige Roadmap, die hier auch als Bild 2 abgebildet ist.

Teilnehmende Organisationen

Organisation	Name
Carl von Ossietzky Universität Oldenburg	Andreas Hestermeyer
Deutsches Forschungszentrum für Künstliche Intelligenz GmbH (DFKI)/Universität Bremen	Prof. Dr. Dieter Hutter
Deutsches Zentrum für Luft- und Raumfahrt (DLR)	Prof. Dr. Barbara Lenz
Fraunhofer-Institut für Experimentelles Software Engineering IESE	Patrick Feth
Fraunhofer-Institut für Experimentelles Software Engineering IESE	Dr. Daniel Schneider
FZI Forschungszentrum Informatik	Carl-Philip Hohl
FZI Forschungszentrum Informatik	Dr. Thomas Meyer
INGenX Technologies GmbH	Jörg Krüger
ITK Engineering GmbH	Pascal Vollmer
OFFIS Institut für Informatik	Dr. Lars Fischer
OFFIS Institut für Informatik	Prof. Dr. Axel Hahn
OFFIS Institut für Informatik	Dr. Christoph Mayer
OFFIS Institut für Informatik	Dr. Michael Siegel
Robert Bosch GmbH (Co-Leitung des Arbeitskreises)	Peter Heidl
SafeTRANS (Co-Leitung des Arbeitskreises)	Prof. Dr. Werner Damm
SafeTRANS	Jürgen Niehaus
Siemens AG	Dr. Cornel Klein
Siemens AG	Martin Rothfelder
Siemens Mobility GmbH	Bernhard Evers
Universität Hamm-Lippstadt	Prof. Dr. Achim Rettberg
Universität Hamm-Lippstadt	Charles Steinmetz

1 Motivation und Einleitung

1.1 Motivation

Disruptive Veränderungen in Gesellschaft, Bildung, Arbeitswelt und Wirtschaft, Sicherheit, und individueller Unabhängigkeit kündigen sich vielschichtig an. Diese werden durch eine rasante Entwicklung der technischen Systeme – sogenannter Cyber-Physical Systems, kurz: CPS – möglich. Mit diesem Beitrag wollen wir antizipierte Veränderungen in den kommenden Jahren und Jahrzehnten aufzeigen und die damit verbundenen Herausforderungen beleuchten. Daraus wollen wir relevante Forschungsfragen für eine Forschungsroadmap ableiten. Wir verstehen die Arbeit als einen Beitrag zur Gestaltung der Rahmenbedingungen in Gesellschaft, Wirtschaft und Forschung zur Sicherung von Wohlstand, Sicherheit und Lebensqualität der Menschen in unserem Gemeinwesen.

Ausgangspunkt dieser Roadmap sind die folgenden Annahmen:

- CPS werden sich weiterentwickeln von individuellen, einzelnen Systemen zu Gruppen von Systemen, Kollektiven von Systemen und domänenübergreifenden heterogenen Kollektiven von Systemen.
- Dies wird massiven Einfluss auf die Gesellschaft und auf das alltägliche Leben haben, die es zu gestalten gilt. Die Entwicklung geht in Richtung einer „Human-Machine Society“, in der Maschinen Rollen übernehmen die bisher nur Menschen vorbehalten sind.
- Die Entwicklung der Systeme 2030ff lässt sich durch folgende, pointierte Aussagen charakterisieren:
 - Maschinen bauen Maschinen
 - Maschinen lernen von Maschinen und von Menschen
 - Maschinen treffen eigenverantwortlich Entscheidungen und handeln autonom
 - (Gruppen von) Maschinen kooperieren mit (Gruppen von) Menschen.
 - Maschinen bestimmen über Menschen (treffen Entscheidungen für und über Menschen)
 - Maschinen sind für ihre Handlungen verantwortlich
 - Maschinen sind omnipräsenter Teil einer Welt, in der Menschen leben

Daraus ergeben sich einerseits wirtschaftliche Chancen, die den Erfolg einer Industrienation wie Deutschland bestimmen und den Wohlstand der Gesellschaft auch in der nächsten Generation sicherstellen. Andererseits ergeben sich daraus jedoch auch sehr große gesellschaftliche und technische Herausforderungen.

Diese Roadmap betrachtet daher bewusst neben den technischen Dimensionen auch nicht-technische Herausforderungen in der Gestaltung solcher Systeme, so dass sie sowohl aus individueller Sicht, wie auch aus gesellschaftlicher Sicht den Menschen dienen und auf eine hohe Akzeptanz stoßen. Wir können und wollen uns dem technischen Fortschritt nicht verschließen, deshalb müssen wir ihn gestalten. Dabei stellen sich folgende Fragen:

- Wie beherrschen wir, was wir bauen?
- Wollen wir, was wir bauen?
- Wie garantieren wir die Eigenschaften der Systeme, die wir bauen?
- Welche Rolle spielen Menschen in dieser Welt der Technik?
- Können wir schützen, was wir bauen?

Der technische Fortschritt folgt gesellschaftlichen Strömungen, aber er verändert die Gesellschaft auch. Um Technik so zu gestalten, dass sie die Entwicklung einer humanen Gesellschaft fördert, stellen sich folgende Fragen:

- Welcher Ethik folgen Maschinen?
- Wie fördern Maschinen die menschliche Entwicklung?
- Gibt es – wie für den Menschen – auch einen Würde-Begriff für (intelligente) Maschinen?
- Privatheit: Welche Informationen dürfen Maschinen nutzen?
- Wieviel Transparenz, Vertrauen, Kontrolle ist notwendig?

Hinweis zur verwendeten Sprache:

Die hier betrachteten HCPS werden zunehmend „autonom“, sie „handeln“ selbstständig, werden „intelligenter“, wählen aus Handlungsalternativen aus („entscheiden“), „schätzen ab“, welche Auswirkungen ihre Handlungen für sie selbst und die sie umgebene Umwelt haben und „übernehmen Verantwortung“ für ihr Handeln. Wir nutzen diese – bisher für die Beschreibung von Menschen und ihrem Verhalten vorbehaltenen – Begriffe hier, da uns die Sprache keine einfache andere Möglichkeit der Beschreibung dieser Eigenschaften bietet. Wir wollen damit jedoch keine Aussage über die Vergleichbarkeit von maschineller und menschlicher „Intelligenz“, die Anwendbarkeit des „Autonomie-Begriffs“ auf Maschinen oder gar ethische Implikationen bzgl. einer eventuellen „Maschinen-Würde“ geben, sondern lassen diese und verwandte Fragen bewusst offen.

1.2 Einleitung

Die antizipierte Entwicklung von technischen Systemen muss aus den Blickwinkeln von Menschen und Gesellschaft, von Technik und System und von Forschung und Entwicklung betrachtet werden. Zugehörige Leitfragen, denen diese Roadmap folgt, sind:

- Welche systemischen und technischen Veränderungen werden erwartet?
- Welche Auswirkung hat diese technische Entwicklung auf Menschen und Gesellschaft?
- Welche Herausforderungen für Forschung und Entwicklung ergeben sich daraus?

Die hier betrachteten sogenannten Cyber-Physical Systems (CPS) entwickeln sich zu einem dahingehend, dass ihr Vernetzungsgrad zunehmend steigt. Dieser Vernetzungsgrad reicht von einfachen, individuellen CPS über Gruppen und Kollektive von CPS zu heterogenen, domänenübergreifenden Kollektiven von CPS. Zum anderen steigt auch die Komplexität der Systeme gemäß den durch sie zu beherrschenden Aufgaben und den dazu nötigen Fähigkeiten.

Vernetzungsgrad

- **Individuelle Cyber-Physical Systems:** Dies sind CPS, deren Komponenten gemäß einer vorgegebenen hierarchischen Struktur zusammenwirken und deren Verhalten sich durch a priori (d.h. zur Designzeit) entwickelte Konzepte beschreiben lässt. Diese CPS haben eine feste Systemgrenze, d. h., es ist zu jeder Zeit festgelegt, welche Komponenten bzw. Sub-Systeme Teil des betrachteten Gesamtsystems sind und welche nicht. Individuelle CPS sind nur in vorgegebenen Kontexten (sinnvoll) einsetzbar, kooperieren wenig oder gar nicht und haben typisch eine sehr begrenzte Selbst-Wahrnehmung. Beispiele für individuelle CPS sind hochautomatisierte Fahrzeuge, einzelne Komponenten in Energienetzen (z.B. „Smarte Zähler“) oder einzelne Geräte zur medizinischen Versorgung oder Behandlung von Krankheiten.
- **Gruppen und Kollektive von CPS:** Als Gruppe bezeichnen wir dabei eine Menge von individuellen CPS, die zur Erreichung gemeinsamer Ziele zeitlich befristet zusammenarbeiten. Beispiele hierfür sind etwa Platoons von Personen- oder Lastkraftwagen, die sich zur Erreichung des gemeinsamen Ziels „Treibstoff sparen“ zeitlich befristet zusammensetzen und sich ggfs. zu einem späteren Zeitpunkt wieder auflösen. Solange die Gruppe besteht, reduzieren die individuellen CPS die Priorität ihrer eigenen Ziele zu Gunsten des gemeinsamen Ziels, und sie agieren und kooperieren entsprechend. Sobald ein Mitglied der Gruppe eigene Ziele höher priorisiert als das Gemeinziel oder es nicht (mehr) zur Erreichung des Gemeinziels beitragen kann oder „will“, muss es die Gruppe verlassen (alternativ kann in bestimmten Gruppen auch das Gruppenziel angepasst werden). Auch Gruppen von CPS haben eine Systemgrenze, die sich jedoch dynamisch ändert (dadurch, dass individuelle CPS der Gruppe beitreten bzw. diese verlassen). Bei **Kollektiven von CPS** handelt es sich ebenfalls um zeitlich befristete Zusammenschlüsse. Mitglieder eines Kollektivs können individuelle CPS oder Gruppen von CPS sein. Anders als in Gruppen ordnen sich die Teilnehmer hier jedoch nicht bedingungslos den gemeinsamen Zielen unter. Stattdessen handeln sie fortwährend untereinander Wege und Möglichkeiten aus, wie und in welchem Ausmaß sie eigene und globale Ziele gegeneinander priorisieren und in welchem Ausmaß sie zu diesen Zielen beitragen werden. Die Systemgrenze von Kollektiven ändert sich ebenfalls dynamisch durch Beitritt und Verlassen von Mitgliedern. Sowohl **Gruppen** als auch **Kollektive von CPS** haben einen gemeinsamen Kontext (Umgebung, Domäne), der sich über die Zeit ändern kann (dynamischer Kontext).
- (Heterogene) **domänenübergreifende Kollektive von CPS.** Dies sind zielgerichtet gebildete Zusammenschlüsse (Netzwerke) von individuellen CPS und Gruppen von CPS, die in der gleichen

Welt arbeiten, aber unterschiedliche Kontexte (Domänen) haben, wie beispielsweise Mobilität oder Energie. Solche Zusammenschlüsse werden gebildet, um übergeordnete Ziele z.B. aus den Bereichen Security, Gesundheit, Umwelt(schutz) und weitere zu erreichen. Die Zusammenarbeit dieser Kollektive wird dabei z.B. durch Zuordnung (oder Vorenthalt) von Ressourcen oder durch Ansporn- und Belohnungssysteme gesteuert.

Hierarchie der CPS-Systeme:



Abbildung 3: Die der Roadmap zu Grunde liegenden Stufen des Vernetzungsgrades von Systemen

Komplexitätsdimensionen

- **Systemstärke:** Die Stärke eines Systems bemisst sich aus den „Kenntnissen“, den Fähigkeiten und Fertigkeiten eines Systems. Hierzu gehören neben der Wahrnehmungsfähigkeit insbesondere der Grad der Autonomie, der Grad der Intelligenz, sowie die Evolutionsfähigkeit. Die Stärke eines Systems ist umso größer, je mehr und/oder bessere Fähigkeiten es in diesen Unterdimensionen hat.
- **Kontext:** Die Komplexität des Kontexts bzw. der Umgebung, in der das System funktionsfähig/aktiv ist. Genauer: die Komplexität desjenigen Teils der Umwelt, den das System für eine korrekte Funktionsweise kennen und handhaben können muss. Schon für individuelle CPS umfasst diese relevante Umwelt mehrere komplexitätsrelevante Komponenten: (a) Die Umgebung selbst, also alle Gegenstände und Objekte im Umfeld des Systems sowie deren statische und dynamische Eigenschaften, (b) Menschen im Umfeld des Systems, inklusive ihren Eigenschaften, Wünschen und Handlungen, (c) andere technische Systeme und ihr Verhalten sowie (d) Vernetzung mit der Cloud oder anderen Informationsnetzwerken. Für CPS mit höherem Vernetzungsgrad werden weitere Aspekte des Kontexts relevant und steigern somit dessen Komplexität.

Gemeinsam ist diesen Facetten des Kontextes, dass diese entweder unkontrollierbare physikalische Phänomene beinhalten, die möglicherweise kritisch für die Zielerreichung des Egosystems sind, oder andere Systeme (Menschen oder technische Systeme), die jeweils ihre eigene Zielsetzungen verfolgen, welche möglicherweise konfligierend zu den Zielsetzungen des Egosystems sind. Objekte und Systeme im Kontext können somit dem Egosystem „freundlich gesonnen“ sein, d. h., auch potenziell mit ihm kooperieren; sie können antagonistisch wirken,

d. h., gegen die Ziele des Egosystems wirken; oder sie sind neutral gegenüber dem Egosystem und dem von ihm verfolgten Zielen. Grundsätzlich unterscheiden wir, welches Wissen über diese Aspekte des Kontextes dem Egosystem bekannt sind: Im einfachsten Fall - determiniert - sind dem System die relevanten Umgebungselemente bekannt und für die Prädiktion der Weiterentwicklung dieser Umgebungsartefakte liegen genügend genaue (stochastische) Modelle vor - dies wird in der Regel nur für sehr restriktive Umgebungs-kontexte zutreffen. In den meisten Fällen sind Objekte des Kontextes nur partiell bekannt. Die Roadmap differenziert hier weiter, ob die Objekte bekannt sind oder zusätzlich noch Dynamikmodelle vorliegen. Schließlich kann es sein, dass das System sich in einem vollständig unbekanntem Kontext befindet (und sich diesen dann erst selbst erschließen muss).

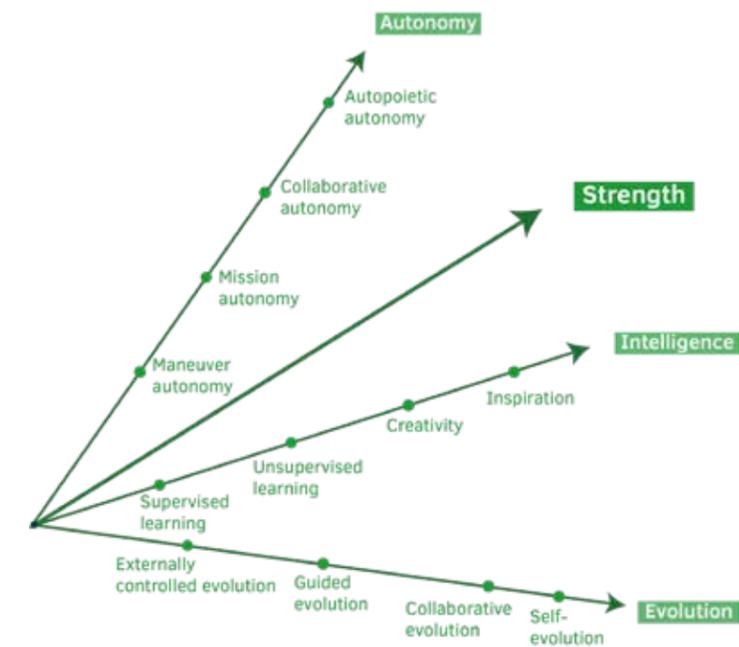
- **Kooperation:** Die Fähigkeit zur Zusammenarbeit mit anderen technischen Systemen und mit Menschen (als Nutzer) sowie die Fähigkeit zur Zusammenarbeit mit durch die Cloud oder die Infrastruktur bereitgestellten Diensten und Services ist ein weiteres Maß für die Komplexität eines Systems, die zur Systemstärke beiträgt, die aufgrund ihrer Wichtigkeit hier jedoch gesondert beschrieben wird. Ausgerichtet auf die oben dargestellten Vernetzungsgrade von Systemen bemisst sich diese zunächst am Umfang der möglichen Kooperationen (z.B. Perceptions-Kooperation zur Erstellung gemeinsamer Lagebilder, Aktions-Kooperation und weitere) sowie an verschiedenen hierzu benötigten Fähigkeiten des Systems, wie der Kommunikationsfähigkeit, der Transaktionsfähigkeit, der Verhandlungsfähigkeit, der Autonomiefähigkeit, der Konfliktfähigkeit und weiterer.
- **Verantwortung und Reflexion:** In dem Maße, in dem die Stärke von Systemen – also ihre Fähigkeiten und Fertigkeiten, insbesondere ihre Autonomiefähigkeit – zunimmt, muss auch die Reflexionsfähigkeit dieser Systeme – also die Selbst- und Fremdbeobachtung über den eigenen Zustand sowie die Fähigkeit, daraus geeignete Schlüsse bzgl. der eigenen Handlung und Handlungsfähigkeit zu ziehen – zunehmen. Gleichzeitig stellt sich die Frage nach der Fähigkeit, Verantwortung für das eigene Handeln zu übernehmen; dies ist keine ausschließlich technische Frage, wobei diese Roadmap jedoch auf die zugehörigen technischen Aspekte fokussiert. Beide Aspekte zusammen bilden eine weitere Dimension, entlang derer die Komplexität der Systeme zunimmt.
- **Integration und Absicherung** sind die letzte hier behandelte Komplexitätsdimension, die den Aufwand und die benötigten Fähigkeiten und Fertigkeiten misst, die zur Sicherstellung der Systemintegrität in jedem Betriebsszenario und damit auch zum Nachweis der Sicherheit solcher Systeme benötigt werden. Neben den klassischen Fähigkeiten und Aufwänden, die hierzu während der Systementwicklung benötigt werden, spielen hier zunehmend auch Fähigkeiten der Systeme selbst – z.B. zum Monitoring, zur on-line Validierung und ähnliche – eine Rolle.

Dieses Roadmap-Dokument folgt der oben beschriebenen Gliederung in Komplexitätsdimensionen, die in den Kapiteln 2 bis 6 ausführlich beschrieben werden. Für jedes dieser Kapitel erfolgt zunächst eine zusammenfassende Beschreibung der jeweiligen Komplexitätsstufen. Die folgenden Abschnitte erläutern dann die Details dieser Zusammenfassung: Die oben beschriebenen Vernetzungsgrade bilden dabei jeweils die Grundlage für eine Analyse der systemischen Veränderungen sowie die daraus abgeleitete Identifikation der benötigten System-Fähigkeiten („Capabilities“). Aus diesen Fähigkeiten wiederum werden technische Herausforderungen extrahiert, Forschungsfragen identifiziert, sowie benötigte Technologien aufgelistet und bezüglich ihrer Verfügbarkeit zeitlich eingeordnet.

In Kapitel 10 schließlich werden die identifizierten System-Fähigkeiten (Capabilities) in eine zeitliche Reihenfolge gebracht. Es entsteht eine Forschungsroadmap, die das zentrale Ergebnis dieses Dokuments ist.

2 Systemstärke

2.1 Zusammenfassung



Als Systemstärke bezeichnen wir die Fähigkeiten des Egosystems, die ihm erlauben, die für eine Anwendung gegebenen Zielsetzungen alleine erfolgreich zu verfolgen. Dies wird umso eher gelingen, umso stärker die Intelligenz des Systems ist, umso besser es sich an sich verändernde Umfeldbedingungen anpassen kann, also welche Evolutionsfähigkeiten es besitzt, und letztlich mit welchem Grad an Autonomie es ausgerüstet ist. Die folgende Abbildung zeigt diese Dimension sowie die jeweilige Skala der Ausbaustufen:

Ausbaustufen Autonomie

1. Funktionsorientierte Autonomie: Eine einzelne Funktion, z. B. Einparken, wird sicher ausgeführt.
2. Missionsorientierte Autonomie: Komplexe Abfolgen von Aktivitäten entlang eines Plans werden sicher ausgeführt.
3. Kooperative Autonomie erweitert das Selbstverständnis des Systems vom Ich zum Wir. Dadurch werden Ziele und Aufgaben durch abgestimmtes Verhalten mit anderen Systemen erreicht.
4. Während bei den bisherigen Formen der Autonomie davon ausgegangen wird, dass der Kontext in dem Systeme agieren bekannt ist, können sich autopoietisch autonome Systeme unbekanntem Kontext selbst erschließen und darin allein oder kooperativ handeln.
- 5.

Ausbaustufen Intelligenz

1. Überwachtes Lernen: Das System lernt unter Anleitung Objekte bzw. Elemente des Kontextes zu erkennen, zu analysieren, daraus Handlungsbewertungen vorzunehmen und auf den Kontext einzuwirken, um so seine Ziele zu erreichen.
2. Nicht überwachtes Lernen: Das System lernt selbständig Objekte bzw. Elemente des Kon-

textes zu erkennen, zu analysieren, daraus Handlungsbewertungen vorzunehmen und auf den Kontext einzuwirken, um so seine Ziele zu erreichen.

3. **Kreativität:** Das System ist selbständig in der Lage vorher nie angewandte Kombination von Fertigkeiten in der Wahrnehmung, Analyse, Handlungsableitung und Einwirkung auf den Kontext so zu kombinieren, dass vollständig neue Lösungsansätze zur Erreichung der Ziele des Egosystems gefunden werden. Dies schließt die Fähigkeit ein, zu erkennen, dass die eigene Stärke nicht ausreicht, um die Ziele zu erfüllen, und solche Systeme des Kontextes zu identifizieren, die als Kooperationspartner gewonnen werden müssen, um seine Ziele zu erfüllen.
4. **Inspiration:** Das System kann selbständig sich völlig unbekannte Kontexte so erschließen, dass darin enthaltene Strukturmerkmale identifiziert werden und Zielsetzungen und Strategien so erweitern, dass es aufgrund seiner Kreativität in diesem vorher unbekanntem Kontext seine Ziele erreichen kann.

Ausbaustufen Evolution

1. **Kontrollierte Evolution:** Die Fähigkeit des Systems, sich unter Kenntnis seines eigenen Integritäts- und Gesundheitszustandes so zu reorganisieren, dass trotz Verletzungen von Integritätsbedingungen und Teilausfällen von Systemen ein maximaler Grad an noch möglicher Zielerfüllung erreicht wird.
2. **Geführte Evolution:** Die durch Stakeholder geführte Weiterentwicklung des Systems aufgrund von im Betrieb oder in der Cloud auf der Basis von Digital Twins erkannten Fehlern in der Wahrnehmung, Analyse, Strategiebildung oder Handlung des Systems zur Behebung dieser Schwachstellen.
3. **Selbstevolution:** Die durch das System selbst geführte Weiterentwicklung des Systems aufgrund von im Betrieb erkannten Fehlern in der Wahrnehmung, Analyse, Strategiebildung oder Handlung des Systems zur Behebung dieser Schwachstellen.
4. **Kooperative Evolution:** Die Fähigkeit einer Gruppe oder eines Kollektivs zur selbst geführten Weiterentwicklung der Systeme der Gruppe (Selbstevolution der konstituierenden Systeme).

2.2 Detaillierte Beschreibung

2.2.1 Analyse

Mit „Systemstärke“ erfassen wir Fähigkeiten und Eigenschaften von Systemen, die es erlauben, seine Zielsetzung in beliebigen Kontexten durchzusetzen. Der Begriff kann spieltheoretisch begründet werden; in dieser Sichtweise „spielt“ das System gegen seinen Kontext und gewinnt dann, wenn es seine Ziele – gegebenenfalls auch in einer beliebig feindlichen Umgebung – durchsetzen kann. Wir analysieren in diesem Abschnitt, welche Eigenschaften dazu beitragen, das System zu stärken und welche Capabilities benötigt werden, um Kontexte zunehmender Komplexität zu beherrschen.

Eigenschaften

Die dominanten Veränderungen der Systemstärke zukünftiger Systeme lassen sich mit den Eigenschaften Strukturdynamik, Evolution, Intelligenz und Autonomie charakterisieren.

- **Strukturdynamik** bedeutet, dass Systeme sich gesteuert oder spontan in ihrer Struktur verändern, indem sie sich zu neuen Systemverbänden zusammenschließen oder wieder auflösen.

Dies sind die in der Einleitung eingeführten Vernetzungsgrade (von individuellen CPS bis hin zu heterogenen Kollektiven von CPS).

- **Evolution** bedeutet, dass das CPS zur Betriebszeit neue Features erhält, die hauptsächlich durch neue Software bestimmt sind. Dabei kann man sich die folgenden Ausbaustufen und Mischformen vorstellen.
 - **(Extern) Kontrollierte Evolution** (kontrolliert durch und in der Firma/Organisation, die das System entwickelt und gebaut hat), getrieben aus Erkenntnissen im Feld, mit virtueller Freigabe in der Entwicklungsumgebung und Nachladen der neuen SW in das System im Feld. Das einzelne System spiegelt seine Erkenntnisse in die Cloud in einen „Digitalen Twin“. Dort kann eine Auswertung über eine Flotte erfolgen, die zu Evolutionsentwicklungen führt. Über ein geeignetes Verteilsystem wird die neue SW in die Systeme im Feld verteilt. Die Intelligenz steckt im Kopf des Entwicklers.
 - **Geführte Evolution in der Cloud und Freigabe in der Cloud:** Die Erkenntnisse der Digitalen Twins werden ausgewertet und gelernt, neue adaptierte Funktionen werden durch die Analyse abgeleitet und realisiert. Die Verifikation und Validierung findet in der virtuellen Entwicklungsumgebung statt und über ein geeignetes Verteilsystem wird die neue SW in die Systeme im Feld verteilt. Die Intelligenz steckt zentral in der Cloud.
 - **Selbstevolution:** Das System macht Erfahrungen über seinen Betrieb, analysiert diese und entwickelt Veränderungen. Diese werden on Board oder in der Cloud verifiziert und validiert sowie im System freigeschaltet. Dieser Ansatz braucht sehr viele Ressourcen im System selbst. Diese Systeme entwickeln sich individuell, so dass Erfahrungen einzelner Systeme immer schwieriger von der Flotte genutzt werden können. Die Intelligenz sitzt im System.
 - **Kooperative Evaluation:** Wie Selbstevaluation, allerdings nutzen die einzelnen Systeme auch Erfahrungen anderer Systeme desselben Typs, beziehen diese in ihre Erkenntnisse und die sich daraus ergebenden Änderungen ein und teilen die Änderungen den anderen Systemen mit. Es entsteht eine gemeinsame Weiterentwicklung/Evolution.
 - **Intelligenz** beinhaltet die Fähigkeit, bisher unbekannte Kontexte so zu erschließen, dass die Ziele des Systems auch in den bisher unbekanntem Teilen des Kontexts durchsetzbar sind. Dadurch wird ein System sich morgen intelligenter verhalten als es das heute tut; es wird sich an einen dynamischen, offenen Kontext anpassen können. Intelligenz stützt sich auf (überwachtes oder nicht überwachtes) Lernen, wird aber darüber hinaus zunehmend Fähigkeiten subsumieren, die wir bei Menschen mit Kreativität und Inspiration bezeichnen:
 - **Kreativität:** Jede neue, noch nicht da gewesene, von wenigen Menschen gedachte und effektive Methode, Ziele zu erreichen⁹
 - **Inspiration:** Die Fähigkeit, für das System relevante neue Ziele zu kreieren
 - **Autonomie** bedeutet, dass ein System, ohne externen Einfluss in einem Kontext agiert. Dabei ist die Frage, wie komplex der Kontext sein kann, in dem ein System autonom agiert.
- Wir unterscheiden – in Anlehnung an die SafeTRANS Roadmap *Hochautomatisierte Systeme* – die folgenden Grade an Autonomie:
1. **Funktionale automatisierte Systeme** können begrenzte, klar definierte Aufgaben (Manöver) autonom erfüllen, wie z.B. automatisches Einparken, automatisches Landen oder die automatische Abarbeitung einer durch einen Operator im Vorfeld geplanten Mission. Diese

⁹ In Anlehnung an Guilford, Joy Paul/Hoepfner, Ralph. The analysis of intelligence. Weinheim (u.a.). Beltz 1976

Systeme können während des Betriebs nicht lernen; die Kooperation mit anderen Systemen ist auf den Austausch von Kontextinformationen beschränkt.

2. **Missionsorientierte Systeme** haben die Aufgabe, situationsabhängig eine ungeplante Kette beherrschbarer und bekannter Situationen zu durchlaufen. Dabei können verschiedene Optimierungskriterien wie die Minimierung des Zeit- oder Ressourcenbedarfs eine Rolle spielen. Planungs- und Optimierungsberechnungen werden zur Laufzeit durchgeführt. Diese Systeme können während des Betriebs nicht lernen; die Kooperation mit anderen Systemen ist auf den Austausch von Informationen über den Kontext und über das System selbst beschränkt. Beispiele hierzu sind der Highway-Pilot oder die Durchführung von Gebietserkundungen.
3. **Kollaborative Systeme** sind Systeme wie Roboter, Fahrzeuge, Schwärme, die z.B. einfädeln lassen oder die zur Unfallvermeidung miteinander kooperieren. Solche Systeme sind zur Erfüllung ihrer Mission in der Lage, mit anderen Systemen und Menschen zu kooperieren und ihre Wahrnehmungen, Interpretationen, Ziele, Pläne und Aktionen miteinander abzustimmen. Die Systeme tauschen mit ihren Kooperationspartnern relevante Kontextinformationen aus, sind jedoch nicht lernfähig
4. **Autopoietische Systeme**¹⁰ sind Systeme, die ihre Perzeption, ihre Interpretationen, ihre Aktionen und ihre Kooperationsmöglichkeiten selbstständig erweitern und sich mit anderen Systemen darüber austauschen können (inklusive der Weitergabe von erlerntem Verhalten). Diese Systeme zeigen somit menschenähnliches Verhalten. Die Fähigkeit des nicht-überwachten Lernens ist das wesentliche Charakteristikum dieser Systemklasse.

2.2.2 Capabilities

Um die oben dargestellten Ausbaustufen zu erreichen, müssen HCPS mit den folgenden Fähigkeiten (Capabilities) ausgestattet sein: Sense, Predict, Analyze, Reflect, Decide, Certify, Act, Cooperate, Learn, Evolve, Create. Bevor wir diese Capabilities im Einzelnen beschreiben geben wir hier zunächst anhand eines Beispiels an, wie diese Capabilities angewandt werden können.

- (sense and reflect) Das System reflektiert den aktuellen Grad der Umweltwahrnehmung und befindet diesen für nicht ausreichend Handlungsentscheidungen zu fällen.
- (learn) Dies führt zum Einsatz der Lernfähigkeiten, um unbekannte Objekte in der Umgebung und deren Dynamik zu erlernen.
 - (act) ggfs. initiiert das System hierzu Experimente um den aktuellen Kontext zu untersuchen...
 - (learn) ... und nutzt seine Lernfähigkeiten, um neue, relevante Objekte zu identifizieren und/oder neue, verbesserte Vorhersagemodelle für deren Verhalten und/oder neue, bessere Handlungsstrategien für sein eigenes Verhalten zu ermitteln.
 - (evolve) Durch diesen Lernprozess wird ggfs. eine Systemveränderung initiiert, d.h. das System passt z.B. seine interne Struktur und Verhaltensweise an die neu gelernten Optimierungskriterien an...
 - (create) und erzeugt dabei ggfs. auch neue Verhaltensweisen oder Fähigkeiten (z.B.

¹⁰ Maturana, H.R./Varela, F.J. The cognitive process. Autopoiesis and cognition: The realization of the living. Springer Science & Business Media. 1980. Seite 3: "An autopoietic machine is a machine organized (defined as a unity) as a network of processes of production (transformation and destruction) of components which: (i) through their interactions and transformations continuously regenerate and realize the network of processes (relations) that produced them; and (ii) constitute it (the machine) as a concrete unity in space in which they (the components) exist by specifying the topological domain of its realization as such a network"

verbesserte Sensor-Fusion Fähigkeiten), die für die zukünftige Identifikation der neu erlernten Umgebungs-Artefakte notwendig sind (und die vor ihrem erstmaligen Einsatz entsprechend validiert bzw. zertifiziert werden müssen).

- (analyze) Wenn der Grad der Umweltwahrnehmung ausreichend hoch ist (bzw. das System eine ausreichend hohe Konfidenz in diesen Grad der Umweltwahrnehmung hat), wird das System seine Optionen zur Erreichung seiner Ziele analysieren und entsprechende Handlungsstrategien entwerfen und bzgl. Risiko und Erfolgswahrscheinlichkeit bewerten.
- (reflect) Vor der Auswahl einer Handlungsstrategie reflektiert das System die Auswirkungen der zur Umsetzung notwendigen Handlungen auf andere Systeme und auf die Umwelt, sowie deren Konformität zu normativen Wertesystemen und Regularien.
- (act) Schließlich setzt das System eine dieser Strategien durch Initiierung der entsprechenden Handlungen um.
- (cooperate) Ist das System nicht stark genug, die Strategie selbst umzusetzen, so initiiert es Kooperationen mit anderen Systemen
- (certify) Dies gibt die Fähigkeit des Systems an, zu jedem Zeitpunkt seine Konformität zu entsprechenden Regularien, Regeln und Qualitätsmaßen (bzgl. funktionaler und IT-Sicherheit, Verfügbarkeit, Gesetzen und Verordnungen, gesellschaftlichen Akzeptanzkriterien, usw.) darstellen zu können.

Im Einzelnen beschreiben diese Capabilities die folgenden Fähigkeiten:

Sense Behandelt die Frage, was eine Systemkategorie als Ganzes oder in Teilen von sich (Selbstwahrnehmung) und seinem relevanten Kontext (Kontextwahrnehmung) wahrnehmen muss um handlungsfähig zu sein.

Analyze Behandelt die Frage, was eine Systemkategorie als Ganzes oder in Teilen an Information über den Kontext oder sich selbst selektieren, erkennen und interpretieren kann.

Predict Behandelt die Fähigkeit, mögliche Zukünfte in der Weiterentwicklung von System und Kontext vorhersagen zu können

Reflect Beinhaltet die Fähigkeit, die Auswirkungen von eigenen Handlungsalternativen auf den Kontext einschließlich relevanter Normen und Regularien bewerten zu können. Diese Capability behandeln wir ausführlich in Abschnitt 5 dieses Dokuments.

Decide Behandelt die Frage, was eine Systemkategorie als Ganzes oder in Teilen entscheiden können muss. Dabei sind folgende Aspekte zu beachten:

Erklärbarkeit von Entscheidungen; kollektive versus individuelle Entscheidungen; Auflösen von Zielkonflikten; rationale, moralische, intuitive Entscheidungen; Ansammlung von Karma (Systemhistorie); Aushandeln von Entscheidungskriterien und Prioritäten im Systemverbund.

Act Behandelt die Frage, wie eine Systemkategorie als Ganzes oder in Teilen in ihrem Kontext handelt, wie Handlungen im Inneren des Systemverbundes vollzogen werden, wie der Systemverbund seine Handlungen koordiniert und dabei keinen Systemteil überfordert (Verfügbarkeit von Ressourcen und Fähigkeiten), wie ein Systemverbund Verantwortung für seine Handlungen übernimmt bzw. auf Teile delegiert (Überlappung mit Kooperation). Handlungen, die auf die Adaption des Systems gerichtet sind, um besser (effektiver und effizienter) im Kontext zu wirken.

Evolve

- Systeme bilden eine neue Systemgrenze und Systemkomponenten ändern ggf. ihre Relationen

zueinander.

- Mit der veränderten Systemstruktur ändert sich auch der Systemkontext und ggf. die Beziehungen zum Systemkontext. (Bsp. Ist ein Fahrzeug Teil eines Platoons, reagiert es nicht mehr individuell auf Ereignisse in der Umwelt.)
- Da einzelne Systeme sich im Verbund noch unabhängig weiterentwickeln können, muss die Kompatibilität z.B. bei Upgrades einzelner Systeme im Verbund sichergestellt werden.
- Systeme des Verbunds haben neue gemeinsame Ziele, die nicht immer kompatibel mit den Zielen eines einzelnen Systems sind.
- Verbünde von Systemen nutzen gemeinsame Informationen, die sich aus den Informationen der Teilsysteme aggregieren.
- Ein Systemverbund kann neue Wirkketten darstellen, die nur im Verbund möglich sind und sich aus den Fähigkeiten der Einzelsysteme ergeben.
- Systemverbünde erzeugen positive (gewollte) und negative (nicht gewollte) „emergente“ Funktionen. Beide gilt es zu beherrschen.
- In Systemverbänden müssen Entscheidungen gemeinsam getroffen werden, dabei kann es zu Konflikten kommen.
- Verbundspezifische Kontextinformationen werden ausgetauscht und aktuell gehalten. Jedes System trägt seinen Teil zum gemeinsamen Kontextbild bei.
- Systemdynamik muss verteilt synchronisiert werden (Schwarm).
- Kommunikation muss hochverfügbar, sicher und in real time ablaufen sowie broadcasting unterstützen um die Wirkketten des Systemverbundes ohne Einschränkungen zu ermöglichen.
- In Systemverbänden wirken typischerweise verschiedene Realisierungstechnologien zusammen, die auf verschiedenen Systemtheorien beruhen (z.B. signalbasierten, eventbasierten, informationsbasierten, KI Ansätzen, u.a.m.)

Cooperate Die Fähigkeit der Kooperation wird in Kapitel 4 ausführlich dargestellt.

Create Behandelt die Frage, wie eine Systemkategorie als Ganzes oder in Teilen seinen Verhaltensraum (Self-extension) und ggf. seinen Kontext erweitert. Je nach Systemklasse kann sich die Selbstadaptation auf unterschiedliche Teile und Fähigkeiten des Systems oder seiner Teile beziehen. Dazu zählen zum Beispiel:

- Restriktionen des Systemverhaltens,
- Einschränkung auf einen bestimmten operativen Kontext
- Anpassung und gesetzliche Maßnahmen zur Regelung von Verhaltensweisen in verschiedenen Kontexten (z. B. Ländergrenzen)
- Anpassung und verschiedene Incentive Systeme
- Fähigkeit zur Bewertung der Auswirkungen von Systemveränderungen auf den Kontext
- Erweiterung der Erkennungsfähigkeiten, der Vorhersagefähigkeiten, der Planung - Entscheidungsfähigkeiten, der Handlungsfähigkeiten und der Lernfähigkeiten.

Learn Der Lernzyklus stellt sich grob vereinfacht wie folgt dar:

- Festlegen des Gegenstands der Betrachtung (worüber will ich etwas lernen)
- Aufstellen einer Hypothese über den Gegenstand der Betrachtung (Modell und Wechselwirkung mit seiner Umwelt)
- Aufbau eines Experiments bei dem der Gegenstand gezielt stimuliert und beobachtet werden kann.
- Interpretation des Beobachteten und ggf. Anpassung der Hypothese (Modell und Wechselwirkung mit der Umwelt)

Lernen vollzieht sich in folgenden Feldern:

- **Wissensebene:** Was weiß das System über das zu Lernende? Wie repräsentiert das System das Wissen intern? Inklusiv der Entscheidung welches Wissen das System braucht
- **Integration von Wissen:** Wie integriere ich neues Wissen in bestehendes bzw. wie setze ich es in Beziehung zu bestehendem Wissen oder wie ändert sich bestehendes Wissen dadurch?
- **Handlungsebene:** Wie verwendet das System das Wissen um Entscheidungen zu treffen oder um mit dem Objekt in Interaktion zu treten und beurteilt die Konsequenzen im Ganzen?

Certify Siehe Kapitel 6 dieses Dokuments.

2.2.3 Herausforderungen und Forschungsfragen

Problemanalyse – Systemanalyse

stellt die Frage „Verstehen wir diese Systeme mit ihren Eigenschaften in ihrem Kontext vollständig und widerspruchsfrei, so dass wir sie modellieren und spezifizieren, bauen und verifizieren können?“

Aspekt Strukturdynamik

- Mit welchen Abstraktionen modellieren wir Systeme mit Strukturdynamik und ihre Eigenschaften?
- Welche Sprachen und Theorien haben wir, um Systeme mit Strukturdynamik zu spezifizieren, formalisieren und damit automatisieren zu können?
- Wie beschreiben wir ihre Wechselwirkungsdynamik mit dem Kontext vor dem Hintergrund der Strukturdynamik?

Aspekt System Evolution im Betrieb

- Mit welchen Abstraktionen modellieren wir Systeme mit ihren sich durch upgrade veränderbaren Konfigurationen von Eigenschaften, um die gewünschten und unerwünschten Effekte der Veränderungen beurteilen zu können?
- Welche Sprachen und Theorien haben wir, um Systeme mit veränderbaren Eigenschaften zu spezifizieren, formalisieren und damit automatisieren zu können?
- Wie beschreiben wir ihre Wechselwirkungsdynamik mit dem Kontext vor dem Hintergrund der Feature Evolution im Betrieb?

Aspekt Intelligenz

- Mit welchen Abstraktionen modellieren wir Lernzyklen auf verschiedenen Interaktionsebenen (Objektklasse, Verhalten oder Intention)?
- Modellierung von Hypothesen, Design of Experiments und Interpretationen
- Mit welche Sprachen und Theorien spezifizieren wir Lernzyklen bei denen sich die Wissensbasis und die Handlungsmuster verändern?
- Wie beschreiben wir ihre Wechselwirkungsdynamik mit dem Kontext vor dem Hintergrund des Gelernten?
- Wie beschreiben wir die Auswirkungen in Systemen mit konfliktbehafteten Individualzielen?

Aspekt Autonomie

- Mit welchen Abstraktionen modellieren wir Systeme mit autonomen Fähigkeiten auf den verschiedenen Autonomiestufen, um zu verstehen, welche Komplexität im Kontext sie beherrschen können? Dabei geht es um:
 - die Modellierung der Wahrnehmung und der sich daraus ergebenden Situationen

- die Modellierung der Interpretation und der Prädiktion
- die Entscheidung über Handlungsoptionen und ihre Auswirkungen auf die Einschränkungen bzw. das Risiko, welche sich daraus auf Objekte bzw. Subjekte im Kontext ergeben
- Welche Sprachen und Theorien spezifizieren wir, um autonome Systeme in verschiedenen Klassen von Kontexten zu formalisieren und damit automatisieren zu können?
- Wie beschreiben wir ihre Wechselwirkungsdynamik mit den verschiedenen Klassen von Kontexten vor dem Hintergrund der verschiedenen Ebenen von Autonomie?

Systementwurf und Realisierung

stellt die Frage: „Wie bauen wir diese Systeme so, dass sie ihren Zweck erfüllen, keine Gefahr von ihnen ausgeht, sie den Menschen dienen, mit möglichst wenig Betriebsmitteln auskommen und der Entwicklungsaufwand wirtschaftlich bleibt?“ Es ergeben sich Herausforderungen für die folgenden Aspekte:

- Kompositionelle Architekturen für Strukturdynamik von Systemen
- Mechanismen zur Auf/Abwärtskompatibilität für die Systemevolution
- “Self aware” und “self verifying” Mechanismen und Architekturen für die dynamische Anpassung der Systeme und Systemverbände durch Evolution, Lernen oder Autonomie.
- Mechanismen zur konsistenten Erweiterung von Wissen und Interpretation
- Architektur und Mechanismen, um automatisiert lernen zu können (d.h. spezifizieren des Scope über den etwas gelernt werden soll, Hypothese, DoE und Interpretation)
- RTE für die Ausführung von Lernen, Autonomie und Synchronisation von Strukturdynamischen Systemen.
- Mechanismen zur Self-Awareness und Self-Adaption von Systemen
- Design for Verification and Validation
- Unterbindung von mit maliziöser Absicht eingebrachten Verhaltens

System Verifikation und Validierung

stellt die Frage „Wie kann ein Nachweis für eine belastbare Freigabe von strukturdynamischen, evolutionären, lernenden, autonomen Systemen geschaffen werden?“ Da die Freigabe nicht mehr durch Personen erfolgen wird, sondern spontan dann, wenn sich Systemverbände bilden und auflösen, muss das alles virtualisiert werden.

- Wie formalisiert man virtuelle Verifikationsumgebungen?
- Wie formalisiert man die Bildung virtueller Verifikationsfälle?
- Was muss im System beobachtet werden können und wie aggregiert man das Wissen für die Freigabefrage?
- Wie formalisiert man die Bildung von spontanen Freigabekriterien?
- Wie formalisiert man die Systemanpassungen, die durch Erkenntnis-virtuellen Verifikationsprozess entstehen?
- Wie werden diese Fähigkeiten durch die Systemarchitektur, die Designmechanismen des Systems bzw. des Systemverbundes unterstützt?

2.2.4 Benötigte Technologien und Zeithorizont

Die folgende Tabelle gibt für jede Ausbaustufe dieser Komplexitätsdimension die größten Herausforderungen sowie die wichtigsten benötigten Technologien zu ihrer Realisierung an und ordnet diese zeitlich ein. Diese Analyse ist die Grundlage für die in Kapitel 7 dargestellte Roadmap. (Tabelle siehe folgende Seite.)

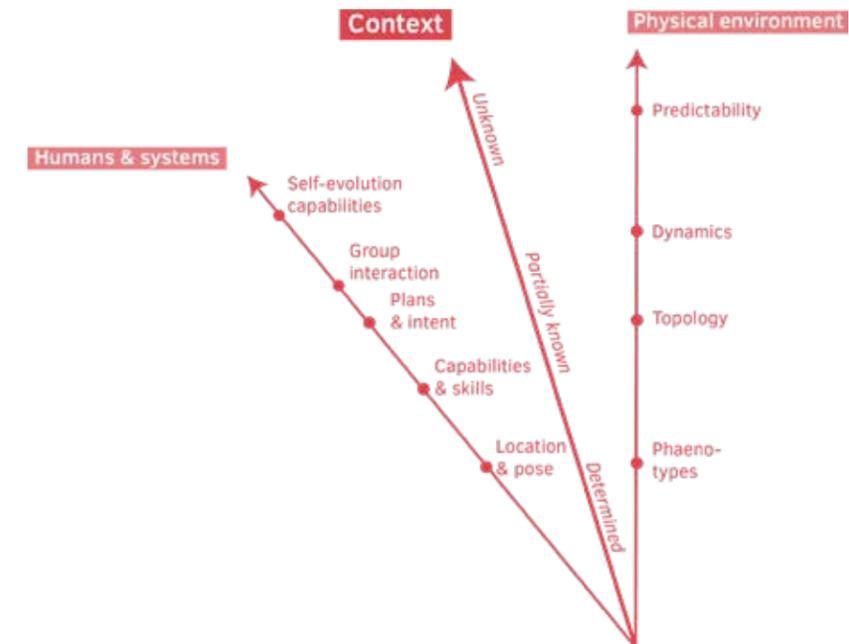
Dimension	Ausbaustufen	Major Challenge	Kerntechnologie	Kurz 2027	Mittel 2035	Lang 2045ff	
Autonomie Entscheidungsfähigkeit in komplexen Kontext	Manöver - Autonomie	Nachweisbar verlässliche Perception	Theorie perfekter Perception		Komplexitätsreduziert	Vollumfänglich	
		Nachweisbar vollständige Situationsbeherrschung	Theorie perfekter Kontextbeschreibung und seiner Evolution	Komplexitätsreduziert		Vollumfänglich	
			Sichere Architektur (incl. Laufzeitumgebung) autonomer Systeme		Komplexitätsreduziert	Vollumfänglich	
	Mission - Autonomie	Handlungsstrategie entwickeln und situativ anpassen	Hierarchische Spiele unter unvollständiger Information mit teil-rationalen Agenten		Komplexitätsreduziert	Vollumfänglich	
			Sichere Architektur autonomer Systeme			Komplexitätsreduziert	Vollumfänglich
	Cooperation - Autonomie	Ziele formulieren und makeln	Spieletheorie		Komplexitätsreduziert	Vollumfänglich	
		Konflikte und Verhandlungsfähigkeit	Spieletheorie		Komplexitätsreduziert	Vollumfänglich	
		Kompostionalität und Kooperationsfähigkeit	Holonische Systemarchitektur			Komplexitätsreduziert	Vollumfänglich
		Selfawareness	Integritätstheorie für Systeme			Komplexitätsreduziert	Vollumfänglich
		Vertrauenswürdigkeit	Systemtheorie für Vertrauenswürdigkeit			Komplexitätsreduziert	Vollumfänglich
	Autopoietische - Autonomie	Hypothesen Bildung/Anpassung	Theorie zur Erschließung von Unbekanntem				Vollumfänglich
		Systemintegrale Experimental- und Lernumgebung	Lernende, adaptive Systemarchitektur				Vollumfänglich
Intelligenz Die Fähigkeit sich unbekanntes zu erschließen	Supervised Learning	Adversary Sensibilitäten erkennen und beherrschen	Theorie statistischer Netzwerke		Komplexitätsreduziert	Vollumfänglich	
		Trainingsendekriterien	Theorie statistischer Netzwerke		Komplexitätsreduziert	Vollumfänglich	
		Sicherheitsnachweis mit statistischen Korrektheitsaussagen	Sichere Architektur autonomer Systeme			Komplexitätsreduziert	Vollumfänglich
	Unsupervised Learning	Hypothesen Bildung/Anpassung	Theorie zur Erschließung von Unbekanntem				Vollumfänglich
		Verlässliches Training mit Metawissen	Theorie dynamischer statistischer Netzwerke			Komplexitätsreduziert	Vollumfänglich
			Erweiterung der sicheren Architektur Lernender Systeme um Metawissenskonzepte				Vollumfänglich
	Kreativität	Analogien (für Probleme und/oder für Lösungen) erkennen und verallgemeinern	Theorie der Strukturähnlichkeit von inhomogenen Objektklas			Bei gegebener Problem-/Lösungsontologie	Vollumfänglich
		Ähnlichkeitskriterien für Problemklassen; Ähnlichkeitskriterien für Lösungsklassen	Ontologien für Problem- und Lösungsklassen		für eingeschränkte Problemklassen		Vollumfänglich
			Synthese von Lösungsansätzen			Bei gegebener Problem-/Lösungsontologie	Vollumfänglich
	Inspiration	Intuition entwickeln	Formalisierung von Intuition				Nicht in den nächsten 50 Jahren
		Theorie von Parallelwelten	Quantenphilosophie				Nicht in den nächsten 50 Jahren

Dimension	Ausbaustufen	Major Challenge	Kerntechnologie	Kurz 2027	Mittel 2035	Lang 2045ff	
Evolution Die Fähigkeit sich zur Betriebszeit durch Betriebserfahrungen zu optimieren oder Veränderungen des Kontextes anzupassen	Externally controlled evolution	Self-Aware Fähigkeit für alle Veränderungsrelevanten Beobachtungen	Secure Crowdsourcing	Vollumfänglich			
		Identifikation von Optimierungs- und Weiterentwicklungsmöglichkeiten	Exception-driven automated Causality Analysis on Big Data	eingeschränkter Komplexitätsraum; interaktiv			
			Ontologien für Problem- und Lösungsklassen	für eingeschränkte Problemklassen; Interaktiv			
		virtuelle Freigabe	vollständige Digital Twins		Eingeschränkt / partielle virt. Freigabe		
			SIL-Simulations- und -Validierungsumgebungen	Eingeschränkt; Interaktiv	Vollumfänglich		
			Testfallgenerierung	Eingeschränkt; Interaktiv			
			Architekturen für virtuelle Freigaben	Vollumfänglich	Vollumfänglich		
			Zertifizierung der Prozesse für die virtuelle Freigabe	Eingeschränkte Problemklassen			
			Secure Individual OTA Updates	On-Line Integritätssicherung	Vollumfänglich		
			Zusätzlich:				
	Guided Evolution		Self-Aware Fähigkeit für alle Veränderungsrelevanten Beobachtungen	Secure Crowdsourcing	Vollumfänglich		
			Identifikation von Optimierungs- und Weiterentwicklungsmöglichkeiten	Exception-driven automated Causality Analysis on Big Data		eingeschränkter Komplexitätsraum; automatisiert	
				Ontologien für Problem- und Lösungsklassen		für eingeschränkte Problemklassen; Automatisiert	
		Virtuelle Freigabe	Vollständige Digital Twins				
			SIL-Simulations- und -validierungsumgebungen		Eingeschränkt; Automatisiert		
			Testfallgenerierung		Eingeschränkt; Automatisiert		
			Architekturen für virtuelle Freigaben		Vollumfänglich		
			Zertifizierung der Prozesse für die virtuelle Freigabe		Eingeschränkte Problemklassen		
	Cooperative Evolution		Zusätzlich:				
		Weitergabe von Lernerfahrungen an andere	Ontologien für Problem- und Lösungsklassen	für einfache Kontexte	Eingeschränkt	Vollumfänglich	
			Bewertung der Generalisierbarkeit von Optimierungs- und Weiterentwicklungsmöglichkeiten	für einfache Kontexte	Eingeschränkt	Vollumfänglich	
		Evolution unter Berücksichtigung der Einbindbarkeit in ein Kollektiv	Cooperative Responsibility		Eingeschränkt	Vollumfänglich	
		Gemeinsame Hypothesenbildung			Eingeschränkt	Vollumfänglich	
		Gemeinsames Design von Experimenten zur Hypothesenüberprüfung			Eingeschränkt	Vollumfänglich	
		Gemeinsame Durchführung von Experimenten zur Hypothesenbildung			Eingeschränkt	Vollumfänglich	
		Gemeinsame Auswertung der Experimente			Eingeschränkt	Vollumfänglich	
		Gemeinsame Angleichung der Hypothesen und ggfs. Erweiterung der Ontologien			Eingeschränkt	Vollumfänglich	

Dimension	Ausbaustufen	Major Challenge	Kerntechnologie	Kurz 2027	Mittel 2035	Lang 2045ff
	Self Evolution	Zusätzlich:				
		On-line Self-Aware Fähigkeit für alle Veränderungsrelevanten Beobachtungen				Vollumfänglich
		On-line Identifikation von Optimierungs- und Weiterentwicklungsmöglichkeiten				Vollumfänglich
						Vollumfänglich
		On-line Virtuelle Freigabe	Vollständige Digital Twins			Vollumfänglich
			SIL-Simulations- und -validierungsumgebungen			Vollumfänglich
			Testfallgenerierung			Vollumfänglich
			Architekturen für virtuelle Freigaben			
		Zertifizierung der Prozesse für die virtuelle Freigabe				Vollumfänglich

3 Kontext

3.1 Zusammenfassung



In der Komplexitätsdimension Kontext messen wir die durch das System zu beherrschende Komplexität der Umgebung des Egosystems. Zur Umgebung des Egosystems zählen alle Systeme (technische Systeme, Menschen) und deren Eigenschaften, deren Kenntnis für die Durchsetzung der Ziele des Egosystems relevant sind (ohne deren Kenntnis eine Zielerreichung nicht möglich ist). Desgleichen zählen hierzu alle relevanten physikalischen Phänomene.

Gemeinsam ist diesen Facetten des Kontextes, dass diese entweder unkontrollierbare physikalische Phänomene beinhalten, die möglicherweise kritisch für die Zielerreichung des Egosystems sind, oder andere Systeme (Menschen oder technische Systeme), die jeweils ihre eigene Zielsetzungen verfolgen, welche möglicherweise konfligierend zu den Zielsetzungen des Egosystems sind. Objekte und Systeme im Kontext können somit dem Egosystem „freundlich gesonnen“ sein, d. h., auch potenziell mit ihm kooperieren; sie können antagonistisch wirken, d. h., gegen die Ziele des Egosystems wirken; oder sie sind neutral gegenüber dem Egosystem und dem von ihm verfolgten Zielen. Grundsätzlich unterscheiden wir, welches Wissen über diese Aspekte des Kontextes dem Egosystem bekannt sind: Im einfachsten Fall - determiniert - sind dem System die relevanten Umgebungsartefakte bekannt und für die Prädiktion der Weiterentwicklung dieser Umgebungsartefakte liegen genügend genaue (stochastische) Modelle vor - dies wird in der Regel nur für sehr restriktive Umgebungskontexte zutreffen. In den meisten Fällen sind Artefakte des Kontextes nur partiell bekannt. Die Roadmap differenziert hier weiter, ob die Artefakte bekannt sind oder zusätzlich noch Dynamikmodelle vorliegen. Schließlich kann es sein, dass das System sich in einem vollständig unbekanntem Kontext befindet (und sich diesen dann erst selbst erschließen muss). Wir präzisieren diese sehr grobe Skala getrennt für physikalische Umgebungsaspekte und Menschen oder technische Systeme in der Umgebung des Systems.

Ausbaustufen Kenntnis physikalische Umgebung

1. Phänotypen: Die für das System relevanten Phänotypen der physikalischen Umgebung sind determiniert/partiell bekannt/unbekannt.
2. Topologie: Die aktuelle Topologie der für das System bekannten physikalischen Umgebung ist bekannt/partiell bekannt/unbekannt. Die Position der Phänotypen in dieser Topologie ist bekannt/partiell bekannt/unbekannt.
3. Dynamik: (Stochastische) Modelle der Weiterentwicklung der Phänotypen sind bekannt/partiell bekannt/unbekannt.
4. Prädiktion: Mittelfristige Modelle der Änderungen der Charakteristika der Phänotypen sind bekannt/partiell bekannt/unbekannt.

Ausbaustufen Kenntnis Menschen und technische Systeme in der Systemumgebung

1. Lage und Pose: Die Position des Systems/Menschen und seine Pose (Orientierung und Haltung) sind bekannt/partiell bekannt/unbekannt.
2. Fähigkeiten und Fertigkeiten: Die Fähigkeiten und Fertigkeiten des Systems/Menschen sind bekannt/partiell bekannt/unbekannt.
3. Pläne und Absichten: Die Pläne und Absichten des Systems/Menschen sind bekannt/partiell bekannt/unbekannt.
4. Gruppeninteraktion: Die Interaktionen zwischen Gruppenmitgliedern und deren Auswirkungen auf Pläne und Absichten der beteiligten Systeme und Menschen sind bekannt/partiell bekannt/unbekannt.
5. Selbstevolution: Zusätzlich zu den Fähigkeiten der Stufe Gruppeninteraktion: Die Fähigkeiten zur Selbstevolution von Systemen und Menschen im Kontext sind bekannt/partiell bekannt/unbekannt.

3.2 Detaillierte Beschreibung

3.2.1 Analyse

Dimensionen des Kontextes

Eine der größten Herausforderungen im Systementwurf ergibt sich beim Übergang von geschlossenen zu offenen Systemen. Offene Systeme interagieren mit ihrem Kontext, und ohne ein tiefes Verständnis des Kontextes eines Systems kann kein Systementwurf gelingen. Bereits bei der „einfachsten“ in dieser Roadmap betrachteten Evolutionsstufe, den individuellen Cyber-Physical Systems (CPS), stellt eine hinreichend vollständige Charakterisierung und Wahrnehmung des relevanten System-Kontextes eine enorme Herausforderung dar. Diese wächst und verändert sich entlang der hier betrachteten Hierarchiestufen nicht nur quantitativ, sondern auch qualitativ; insbesondere unterliegt der relevante Kontext selbst sich ständig ändernden Bedingungen, so das klassische Prinzipien der Systemanalyse, welche auf eine a-priori zur Entwurfszeit erfolgende Kontextanalyse aufbauen, zum Scheitern verurteilt sind. Darüber hinaus wächst, wie nachfolgend dargestellt, die Anzahl der relevanten Dimensionen eines Kontextes.

Individuelle Cyber Physical Systems

In der im Kontext dieser Roadmap verwendeten Interpretation des Begriffs Cyber-Physical Systems sind voll funktionsfähige teilautonome Systeme subsumiert, welche eigenständig oder unter menschlicher Kontrolle die Umsetzung von sich möglicherweise ändernden Zielen verfolgen. Sie

müssen daher über die Fähigkeiten der Umgebungswahrnehmung, der Interpretation der wahrgenommenen Umgebung in Bezug auf die Einhaltung oder Bedrohung der aktuellen Ziele, der Selbsteinschätzung hinsichtlich der Fähigkeiten (Capabilities) zur Beeinflussung der Eigendynamik im wahrgenommenen Kontext, der Prädiktion des wahrgenommenen Kontextes, zur Prädiktion der Eigendynamik durch Aktivierung bzw. Deaktivierung von Capabilities in einem sich weiter entwickelnden Systemkontext sowie zur Planung und Durchführung von sich aus der Analyse ergebenden Manövern verfügen. Prominente Beispiele solcher Systeme sind hochautonome Fahrzeuge, Schiffe, Flugzeuge, Roboter, etc. Für diese Klassen von Systemen sind im verwendeten Logo dieses Abschnittes an Hand des Beispiels hochautonomer Systeme drei Kategorien als besonders relevant identifiziert:

- physikalische Umgebung
- Menschen
- andere CPS bzw. andere technische Systeme
- Vernetzung mit der Cloud

Da zahlreiche laufende Projekte sich bereits mit der Frage der hinreichenden Charakterisierung und Interpretation dieser vier Dimensionen befassen, nicht zuletzt im Kontext vom hochautomatisierten Fahren, sei an dieser Stelle nur eine kurze Charakterisierung gegeben.

Physikalische Umgebung

Hierzu zählen alle für das CPS und dessen Fähigkeit zur Perception und Prädiktion der Weiterentwicklung einer aktuell wahrgenommenen Information relevanten Aspekte. Wir diskutieren dies am Beispiel autonomes Fahren.

Diese lassen sich unterteilen in (weitgehend) statische Aspekte, wie etwa Straßenführung und Beschilderung, deren Wahrnehmung durch hochauflösende Karten und hochauflösende Positionsbestimmung unterstützt wird, und dynamischen Aspekten, wie etwa die Erfassung der umgebenden Verkehrsteilnehmer, der aktuellen Witterungsbedingungen, dem aktuellen Straßenzustand, und für die Sicherheit der Perception relevanten Faktoren wie Blendung, Metallplanken, metallischer, reflektierender Belag einer Plane eines vorausfahrenden Lastwagen, etc.

Mensch

Hier sind alle Arten der Interaktion des Menschen mit dem CPS zu erfassen, wie sie klassisch in einer Stakeholder-Analyse berücksichtigt werden. Neben der Rolle des Menschen als „Human-in-the-Loop“, also als derjenige, der insbesondere Vorgaben zu Zielen des CPS ändern kann und je nach Grad der Autonomie in Perception, Analyse, und Handlung mit einbezogen ist, sind – im Beispiel Autonomes Fahren – Menschen in der Rolle als Verkehrsteilnehmer (Fußgänger, Fahrradfahrer, Führer eines anderen Fahrzeuges), eines Wartungstechnikers, eines Entwicklers, eines Zulassers etc. einzubeziehen. Zahlreiche Projekte beschäftigen sich mit der Frage einer hinreichend guten Erfassung der Vorkenntnisse, des Zustandes, und der Intentionen des Menschen in diesen unterschiedlichen Rollen, z. B. zur Vorhersage der Dynamik von Fußgängern, oder in der Fahrzeugführung, oder der Instandhaltung etc.).

Andere CPS bzw. technische Systeme

Im Gegensatz zur Hierarchieebene „Gruppen von CPS“ kann auf der hier betrachteten untersten Ebene der individuellen Systeme von anderen CPS nur in dem Maße ein kooperatives Verhalten erwartet werden, in dem dies durch übergeordnete Regelwerke – wie z.B. die Straßenverkehrsordnung – geregelt ist. Somit teilt sich das Egosystem einen Verkehrsraum mit anderen CPS, welche alle eigene Zielsetzungen verfolgen, die potenziell den Zielsetzungen des Egosystems entgegenstehen (Beispiel: Erreichen einer Autobahnausfahrt von der linken Fahrbahn aus bei dichtem Ver-

kehr auf der rechten Fahrspur). Auf der nicht-kollaborativen Ebene, wie sie im Moment auch bei der Einführung von hochautonomen Fahrzeugen verfolgt wird, ist die Erkennung der Intention und Dynamik anderer CPS-Systeme damit entweder auf vereinbarte Regelsysteme („Blinker setzen“) oder zwischenmenschliche Kommunikation zwischen Menschen als Human-in-the-Loop beschränkt. Als Zwischenschritt zur voll-kollaborativen CPS-Stufe können zunehmend reichere Formen der Kommunikation zwischen CPS-Systemen verwendet werden, um durch Austausch von Lagebildern auf der Basis von drahtloser Kommunikation (cooperative situational awareness) oder weitergehend Austausch von Intentionen bis hin zu regelnden Eingriffen über Verkehrsinfrastruktur, zunehmend bessere Strategien für die Erreichung der aktuellen Ziele des Ego-Fahrzeugs zu bestimmen.

Gruppen von CPS

„Gruppen von CPS“ bilden sich dynamisch aus räumlich benachbarten CPS um eine von allen CPS für den Zeitraum der Existenz der Gruppe geteilte Zielsetzung zu erfüllen, zumindest bis definierte Abbruchbedingungen erreicht sind. Beispiele für solche Gruppen sind Platoons, Verbände von Industrierobotern in einer Fertigungsstraße, Systeme zur Schaffung von energetischen Nachbarschaften, Systeme zur Überwachung von Patientenstatus in einer intensive care unit, usw.

Der Kontext solcher Systeme muss sowohl Phasen wie Formierung und Auflösung einer Gruppe unterstützen, sowie den Kontext für die als Gruppe geteilten Zielsetzungen bereitstellen.

Für die Phase der Formierung ergeben sich neben den für CPS Systemen erfassten Kontexten die folgenden Dimensionen:

- Technische Konformität anderer CPS aus der Umgebung des Ego-Fahrzeugs in Bezug auf die Fähigkeit zur Formierung einer Gruppe
- Bewertung der Relevanz anderer technisch konformer CPS aus der Umgebung in Bezug auf die Erreichung der Zielsetzung einer Gruppe
- Rechtliche Bewertung der Integration relevanter technisch konformer anderer CPS in Bezug auf die aus der Formierung einer Gruppe resultierenden rechtlichen Fragestellungen, etwa in Bezug auf Haftung
- Mechanismen zur Durchführung von sichereren (im Sinne von Secure) Vertragsabschlüssen zur Formierung einer Gruppe (sicherere Authentifizierung, Absicherung der gemeinsamen Zielsetzungen, Identifikation von Bedingungen zur Auflösung einer Gruppe)
- Mechanismen zur Integration in eine koordinierte Führung der Gruppe (zentral oder dezentral)

Während der Existenz der Gruppe setzen wir voraus, dass diese

- Mechanismen zur Herstellung einer shared situational awareness
- Mechanismen zur Überwachung der Integrität des Systemzustandes
- Mechanismen zur koordinierten Führung der Gruppe

verfügen, so dass in dieser Phase die Dimensionen des Kontextes der Gruppe sich nicht von den Dimensionen des Kontextes der Teilnehmer der Gruppe unterscheiden.

Für die Auflösung einer Gruppe gelten Anforderungen an den Kontext analog zum Aufbau einer Gruppe.

(Heterogene) Kollektive von CPS

Während wir zunehmend höher aggregierte Systeme entlang der Vernetzungshierarchie von Abbildung 3 (Seite 17) betrachten, ergeben sich fundamentale Wechsel in der Perspektive und daraus resultierende Auswirkungen der Führung solcher Systeme auf Handlungsfelder, die den technischen Rahmen weit sprengen und unmittelbare Folgen für die Lebensbedingungen von Menschen, für das Klima, für die Angreifbarkeit einer Gesellschaft als Ganze haben. Damit weitet sich auch der Kontext der Systeme auf Zusammenhänge aus, die jenseits des klassischen Diskursbereiches technischer Disziplinen liegen: sie verlangen zwingend eine Einbeziehung von gesamtwirtschaftlichen, insbesondere volkswirtschaftlichen Betrachtungen, von sozialwissenschaftlichen Betrachtungen, rechtlichen Betrachtungen bis hin zu ethischen Betrachtungen. Zwar bleiben grundsätzlich die auf der untersten Ebene betrachteten Dimensionen des Kontextes relevant. Jedoch entstehen über diese hinaus Wechselwirkungen zwischen System und Kontext, die für eine ganzheitliche Betrachtung der Einführung solcher Systeme essentieller Bestandteil der Systemanalyse sein müssen.

Wir wählen in diesem Abschnitt eine exemplarische Vorgehensweise, um die hier sich ergebenden zusätzlichen Dimensionen des Kontextes einzuführen, und betrachten zunächst Verkehrsflussführungssysteme als Beispiel für ein (homogenes) Kollektiv.

Als Google seine Optimierungsalgorithmen so umstellte, dass die Verkehrsfluss Situation in Seitenstraßen stärker einbezogen wurde, schuf Google ein unmittelbar anschauliches Beispiel für konfigurierende Zielsetzungen in der Optimierung des Verkehrsflusses, welche in diesem Fall ohne gesellschaftliche Konsensbildung alleine aus der Optimierung der Kundeninteressen von Google vollzogen wurde. Klagen von Stadtteilen über die dadurch induzierte Veränderung der Wohnqualität in vorher ruhigen Stadtvierteln sind anhängig.

Dieses einfache Beispiel macht unmittelbar die Vielschichtigkeit der für die Verkehrsflussoptimierung relevanten Verfahren deutlich. So untersucht die DLR die Auswirkungen und Wechselwirkungen nicht nur auf lokale Schadstoffkonzentrationen, sondern auf das Klima als Ganzes. Zum erfassenden Kontext gehören zwingend Modelle des Mobilitätsverhaltens der Bevölkerung in den betrachteten Regionen: mit welchen Anreizen, unter welchen Voraussetzungen, bilden öffentliche Verkehrsmittel, Möglichkeiten zur Reduktion des Individualverkehrs? Wie wirken sich car-sharing Modelle aus? Bis zu welchem Zeitpunkt muss welche Durchdringungsrate von Elektromobilität erreicht werden, um Klimaziele der Bundesregierung einzuhalten? Diese Kategorien von Fragen zeigen auf, wie zwingend verlässliche Modelle von Nutzerverhalten unter unterschiedlichen Szenarien benötigt werden.

Für Kollektive von CPS bedeutet dies somit, dass die folgenden Dimensionen des Kontextes zu betrachten sind:

- Modelle des kollektiven Nutzerverhaltens, unter Berücksichtigung von Incentives und alternativen Nutzungsformen (green energy, green mobility, ...)
- Modelle der Interaktion von Kollektiven auf das Klima, Emissionen, ... unter Einbeziehung von Modellen des kollektiven Nutzerverhaltens
- Modelle für Human-in-the Loop Kontrolle von Kollektiven auf unterschiedlichen Aggregations-ebenen
- Modelle für Bedrohungsszenarien von Kollektiven unter Einbeziehung der Gesamtschäden auf Gesellschaft, Klima, Ressourcennutzung, ...
- Studien zur Auswirkung der Einführung des Kollektives von CPS und von neu bestimmten bzw.

- veränderten Führungsparametern auf die Lebensqualität der Bevölkerung
- Integrierte heterogene Ko-Simulationsmodelle für die Prädiktion des emergenten Systemverhaltens des Gesamtsystemverbundes
- Einschlägige Regelwerke und gesetzliche Vorgaben

Offensichtlich sind diese Facetten des Kontextes selbst dynamisch und einer ständigen Evolution unterzogen. Kollektive von CPS sind somit zwingend evolutionäre Systeme, die ständig die Weiterentwicklung des Kontextes erfassen müssen, und deren Strategien ständig an den sich ändernden Kontext angepasst werden müssen. Dieses setzt voraus, dass auf staatlicher Seite regulierend eingegriffen wird um eine allen gesellschaftlichen Dimensionen gerecht werdende Realisierung von Kollektiven von CPS dadurch sicher zu stellen, dass durch Regelwerke festgeschrieben wird, welche Facetten des Kontextes solcher Systeme in welcher Form zu berücksichtigen sind.

Betrachten wir nun abschließend die Erweiterung zu Heterogenen Kollektiven von CPS, zunächst wieder exemplarisch an Hand des Beispiels Smart City. Verbunden mit dieser Aggregationsstufe ist als konstitutives Element, das die Optimierung des Systems nicht nur aus einer Domäne (z. B. Verkehrsflussoptimierung) erfolgt, sondern unmittelbar (d.h. nicht nur über gemeinsame Kontextmodelle gekoppelt), die Wechselwirkungen von Führungsstrategien zwischen unterschiedlichen Domänen (Bsp. energetische Nachbarschaften und Optimierung des Verkehrsflusses in Nachbarschaften) betrachtet werden. Aus Sicht der Kontextmodellierung bedeutet dies zunächst nur, dass alle relevanten Kontexte der Teildomäne additiv betrachtet werden müssen. Die Herausforderung besteht nun allerdings darin, dass z. B. Modelle des Nutzerverhaltens benötigt werden, welche Incentives und Trends beider Domänen so genau abgebildet, dass sich verstärkende Synergieeffekte positiver und negativer Art sichtbar werden. Aus mathematischer Sicht beinhaltet das die Notwendigkeit, nicht-lineares Verhalten und Chaostheorie im Wechselspiel der Incentives zu berücksichtigen. Entsprechend komplex ist die Erstellung von Prädiktionsmodellen. Aus Sicht der Bedrohungsanalyse zeigt schon ein bekanntes populärwissenschaftliches Buch („Blackout“), wie Wechselwirkungen zwischen einzelnen Domänen zu einer chaotischen Eskalation des Gesamtsystems führen. Wissenschaftlicher dargestellt findet man die Darstellung noch komplexerer – und teilweise sogar noch unbekannter – Wechselwirkungen zwischen Domänen im Material des Intergovernmental Panel on Climate Change (<https://www.ipcc.ch/>).

3.2.2 Capabilities

Damit Systeme – wir verwenden diesen Ausdruck hier generisch für alle vier in dieser Roadmap betrachteten Aggregationsstufen – mit ihrem Kontext interagieren können, benötigen sie – je nach Komplexität des Kontextes - fundamentale Fähigkeiten (englisch: capabilities), die wir in fünf Gruppen kategorisieren:

- **Erkennen**
Die Fähigkeit, den für die aktuelle Mission und aktuelle Zielsetzung des Systems relevanten Teil des Kontextes mit einer dafür genügenden Konfidenz und Präzision zu „erkennen“, sich also ein digitales Bild dieses Kontextes aufzubauen, in dem die relevanten Artefakte, deren Kenngrößen und deren Relationen repräsentiert sind. Das Ergebnis des Erkenntnisvorgangs des Egosystems ist ein digitales Weltbild des Kontextes des Ego Systems.
- **Vorhersagen**
Die Fähigkeit, die Weiterentwicklung des Kontextes für einen von der jeweiligen Aggregations-ebene abhängigen Zeithorizont hinreichend genau auf der Basis des digitalen Weltbildes des aktuellen Zustandes des Kontextes bestimmen zu können, mit einer Konfidenz und Präzision, welche sich aus der Kritikalität der Verletzung der aktuellen Zielsetzungen des Egosystems

- **systematisch ableitet.**
- **Erkunden**
Die Fähigkeit, Hypothesen über nicht genügend genau erkennbare oder präzifizierbare Aspekte des Kontextes zu bilden, diese Hypothesen systematisch zu testen und auf der Basis dieser Beobachtungen Erweiterungen der Modellbildung für das digitale Weltbild und für die Prädiktion der Evolution des Kontextes so zu bilden, dass sie für die aktuellen Zielsetzungen mit genügender Präzision und genügender Konfidenz Erkennung und Vorhersage erlauben.
- **Interagieren**
Die Fähigkeit mit dem Kontext so zu interagieren, dass die aktuellen Zielsetzungen maximal erreicht werden.
- **Gestalten**
Die Fähigkeit, den Kontext so zu verändern, dass eine erfolgreiche Interaktion mit dem Kontext möglich wird

Die folgende Tabelle veranschaulicht diese Begriffsbildungen anhand von Beispielen:

	Individuelle CPS	Gruppen von CPS	Kollektive von CPS	Heterogene Kollektive von CPS
Erkennen	Fußgänger	Kompatibler LKW zur Integration in Platoon	Erkennung von Cyber-Attacken auf Umstellwerk	
Vorhersagen	Prädiktionsmodelle für Fußgänger	Verbesserung von Optimierungszielen durch Integration in Platoon	Durch intelligente beim Fahren buchbare Parkplätze in Parkhäusern lässt sich die Verkehrsdichte im Innenstadtbereich um x reduzieren	Veränderung des Lastbedarfs im regionalen Energienetzen bei 80% Durchdringung des Individualverkehrs mit e-Fahrzeugen
Erkunden	End-to-End learning of driving strategies	Sind Beschleunigungswerte des LKW kompatibel zum Platoon?	Wie wirkt sich eine vollständige Umstellung des ÖPNV in Stadt X auf die CO2 Konzentration aus? Welche Akzeptanz hat das Schaffen einer autofreien Innenstadt?	Lernen von Lastprofilen von Kunden in energetischen Nachbarschaften Insbesondere von e-Fahrzeugen
Interagieren	Sichere Fahrzeugführung	Integration in Platoon	Unterbinden der Weiterverbreitung durch Isolation von durch Computer-Viren befallener Systeme; kollektive Sicherstellung von Versorgung bei Abschaltung betroffener Umspannwerke.	Dezentrale Steuerung energetischer Nachbarschaften unter Berücksichtigung individueller Lastprofile
Gestalten	Projektion von virtuellen Zebrastreifen für Fußgänger zur Vermittlung der Intention des Fahrzeuges	Vermittlungsplattform zur optimalen Formierung von Platoons unterschiedlicher Logistikanbieter in staugefährdeten Autobahnen	Separate Fahrspuren mit physischer Abschirmung zur Komplexitätsreduktion von Erkennen und Vorhersagen für autonome Busse	Durch Vorgabe von Ladezeiten von Elektrofahrzeugen lässt sich in der aktuellen Wetterlage eine Versorgungssicherheit herstellen

Die vorgeschlagenen Kategorien von Capabilities spiegeln unterschiedlicher Arten von Wissen des Systems über den Kontext wieder, die wir in vier Komplexitätsstufen einordnen:

- Komplexitätsstufe A: Bekannte Artefakte – Dynamik determiniert
- Komplexitätsstufe B: Bekannte Artefakte – Dynamik nicht-determiniert
- Komplexitätsstufe C: Partiiell Bekannte Artefakte - Dynamik nicht determiniert
- Komplexitätsstufe D: Unbekannte Artefakte

Komplexitätsstufe A

ist dadurch charakterisiert, dass für die jeweiligen aktuellen Zielsetzungen des Systems alle für die Erreichung dieser Zielsetzungen notwendigen Bestandteile des Kontextes in dem Sinne bekannt sind, dass

- Alle Klassen von Artefakten des Kontextes des Systems bekannt sind
- Alle die Attribute dieser Artefakte, soweit sie für die Erreichung der jeweiligen aktuelle Zielsetzung relevant sind, bekannt sind
- Insoweit es sich bei diesen Artefakten um sich selbst dynamisch weiterentwickelnde Systeme handelt: dass alle für die jeweils aktuellen Zielsetzungen des Egosystems relevanten dynamischen Aspekte der Kontext-Artefakte durch input-deterministische oder probabilistische Modelle beschreibbar sind UND diese Modelle dem Egosystem bekannt sind.

Komplexitätsstufe B

ist dadurch charakterisiert, dass für die jeweiligen aktuellen Zielsetzungen des Systems ALLE für die Erreichung dieser Zielsetzungen notwendigen Bestandteile des Kontextes in dem Sinne bekannt sind, dass

- alle Klassen von Artefakten des Kontextes des Systems bekannt sind
- alle die Attribute dieser Artefakte, soweit sie für die Erreichung der jeweiligen aktuelle Zielsetzung relevant sind, bekannt sind
- insoweit es sich bei diesen Artefakten um sich selbst dynamisch weiterentwickelnde Systeme handelt: dass alle für die jeweils aktuellen Zielsetzungen des Egosystems relevanten dynamischen Aspekte der Kontext-Artefakte durch nicht-deterministische Modelle beschreibbar sind UND diese Modelle dem Egosystem bekannt sind.

Somit ist die Stufe deswegen komplexer, weil zwar eine hinreichend genaue Umgebungswahrnehmung möglich ist, allerdings die Prädiktion der Weiterentwicklung des Kontextes im Gegensatz zu Stufe A worst-case Annahmen vornehmen muss, die alle durch nicht-deterministische Modelle der Dynamik der Umgebungskomponenten möglichen Evolutionen umfassen.

Komplexitätsstufe C

ist dadurch charakterisiert, dass für die jeweiligen aktuellen Zielsetzungen des Systems manche für die Erreichung dieser Zielsetzungen notwendigen Bestandteile des Kontextes in dem Sinne bekannt sind, dass

- manche Klassen von Artefakten des Kontextes des Systems bekannt sind
- manche Attribute dieser Artefakte, soweit sie für die Erreichung der jeweiligen aktuelle Zielsetzung relevant sind, bekannt sind
- insoweit es sich bei diesen Artefakten um sich selbst weiterentwickelnde Systeme handelt: dass alle für die jeweils aktuellen Zielsetzungen des Egosystems relevanten dynamischen Aspekte der Kontext-Artefakte durch nicht-deterministische Modelle beschreibbar sind UND diese Modelle dem Egosystem bekannt sind

Komplexitätsstufe D

ist dadurch charakterisiert, dass das System in einer nicht bekannten Umgebung mit nicht bekannter Dynamik eingesetzt ist. Bsp: Katastrophenbehandlung nach Atomangriff

Diese Komplexitätsstufen bilden eine in Abbildung 2 gezeigte Halbordnung, welche diese totale Ordnung verfeinert.

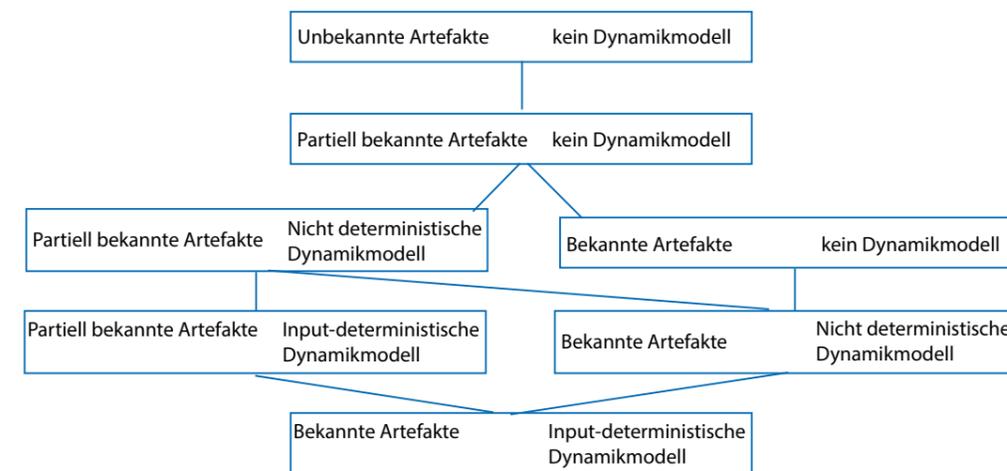


Abbildung 4: Halbordnung der Komplexitätsdimension Kontext

Abbildung 4 zeigt Zwischenstufen auf, in denen für die dynamischen Artefakte, welche dem System bekannt sind, auch determinierte Modelle bekannt sind, bzw. zwar die Artefakte bekannt sind, aber keinerlei Kenntnis über deren Dynamik vorliegt. Offensichtlich ist die Halbordnung von Abbildung 1 beliebig nach dem Grad der Partialität der Kenntnis der relevanten Umgebungsartefakte verfeinerbar. Da wir uns hier auf qualitative Betrachtungen beschränken, verzichten wir auf die Darstellung dieser Zwischenstufen, welche für die Praxis der Systementwicklung dagegen hochgradig relevant sind. Tatsächlich ist für fast alle komplexeren hochautonomen Systeme davon auszugehen, dass NIE eine vollständige Charakterisierung aller für die Erreichung der Ziele relevanten Artefakte gelingen wird, sondern stets im Feld Erfahrungen über weitere relevante Artefakte gewonnen werden.

Der Raum der Technischen Herausforderungen in der Interaktion des Systems mit dem Kontext ist damit durch mehrere Dimensionen aufgespannt:

- die jeweilige Aggregationsstufe
- die jeweiligen Aspekte des Kontextes auf dieser Aggregationsstufe
- dem Grad an Vorwissen über diesen Aspekt des Kontextes, wie er durch die obige Halbordnung beschrieben wird

Grundsätzlich werden dabei die Capabilities der Kategorien „Erkunden“ und „Gestalten“ benötigt, um hohe Komplexitätsgrade des Kontextes so zu reduzieren, dass „Erkennung“, „Prädiktion“, und „Interaktion“ möglich ist. Dies gelingt umso leichter, wie wesentliche Aspekte des Kontextes des Systems diesem gegenüber eher „freundlich“ oder eher „böswillig“ eingestellt sind, also sich eher so verhalten, dass sie das System in der Durchsetzung seiner Ziele „wenn möglich“ unterstützen, oder im Gegenteil versuchen, das System an der Durchsetzung seiner Ziele wenn immer möglich zu behindern. Diese informellen Begriffe lassen sich im Rahmen der Spieltheorie formal definieren. Der Begriff „wenn möglich“ spielt hierbei auf Interessenkonflikte an; eine der Instanzen der Capability „Interagieren“ stellen Verhandlungen dar, die darauf zielen, temporäre Koalitionen

zu bilden und damit in dieser Phase eine breitere Unterstützung für die Durchsetzung der Ziele des Egosystems zu erreichen.

3.2.3 Herausforderungen und Forschungsfragen

Grundsätzlich lassen sich aus der in Abschnitt 3.2 dargestellten Systematik Forschungsherausforderungen und Forschungsfragen für jede Kategorie von Systemen ableiten. Wir veranschaulichen dieses an wenigen Beispielen.

Forschungsherausforderungen

Beispiel 1: Wir betrachten das homogene Kollektiv von CPS „Verkehrssystem“

In der Kategorie **Gestalten** stellen sich in diesem Beispiel unter anderem die folgende Forschungsfragen:

- Welche Fahrverbote von Dieselfahrzeugen erzielen welche Verbesserung der Luftqualität?
- Welche Durchsetzbarkeit haben diese Maßnahmen?
- Welche Incentives tragen zur Kooperationsbereitschaft bei, statt der Nutzung von Dieselfahrzeugen auf Elektrofahrzeuge, Fahrrad, oder ÖPNV umzusteigen?
- Welche Maßnahmen sind für die Erzielung der Klimaziele im Mobilitätssektor bis 2030 wie zu priorisieren (Grad an Durchdringung von e-Fahrzeugen, e-people mover als Kompensation von Fahrverboten für Individualverkehr in Innenstädten, last-end-logistic nur über einen provider pro Stadt, nur mit Elektrofahrzeugen,)

Beispiel 2: Krebsbehandlung durch eine Group of HCPS: Team von Ärzten und Medizinischen Geräten und Pflegepersonal. Hier stellen sich unter anderem die folgenden Forschungsfragen:

In der Kategorie **Erkennen**

- Wie stellen wir sicher, dass alle für die Identifikation von Lungenkrebs relevante Artefakte in einem KI-basierten Diagnosesystem berücksichtigt worden sind?

In der Kategorie **Vorhersagen**

- Wie stellen wir sicher, dass eine KI basierte Verlaufsprognose des diagnostizierten Lungenkrebs den individuellen Gegebenheiten eines konkreten Patienten Rechnung trägt (z. B. emotionale Einstellung, Grad der Unterstützung durch intakte Familienbeziehung, etc.)

In der Kategorie **Interagieren**

- Wie kann das KI System dem behandelnden Arzt begründen, warum genau diese Therapieempfehlung gegeben wurde?

In der Kategorie **Erkunden**

- Forschung zur Identifikation weiterer relevanter Wirkungsfaktoren für die Entwicklung von Lungenkrebs
- Verwendung von Krebsregistern zur Detektion weiterer potenziell relevanter Merkmale

In der Kategorie **Gestalten**

- Welche Regularien müssen für die Überwachung von Integritätsbedingungen für Trainings- und Testdaten für Erkennung und Prognose von Lungenkrebs etabliert werden?
- Wie können diese vor böswilligen Manipulationen geschützt werden?
- Wie sehen Zulassungsprozesse für KI-basierte Medizinprodukte aus?

- Wie kann sichergestellt werden, dass bei erkannten schweren Fehldiagnosen von KI Systemen die dafür verwendete Lernsoftware in Ihrer Nutzung so lange untersagt wird, bis eine rezer-tifizierte Version erstellt und verteilt worden ist, welche diese Fehlklassifikation nicht mehr vornimmt

Weitere Forschungsfragen

Kategorie **Certify**

- Wie kann die Sicherheit und Integrität und Qualität der aus der Cloud übertragenen Daten so abgesichert werden, dass diese eine Basis für eine gegebenenfalls on-line Zertifizierung der Sicherheit des Gesamtsystems bieten kann?
- Wie sehen dafür geeignete on-line Verifikationsverfahren aus?
- Wie können abgesicherte Konfidenzaussagen bei der Verwendung von Lernenden Verfahren in der Umgebungswahrnehmung hergeleitet werden?

Kategorie **Erfassen**

- Für die Erfassung von Umgebungsparametern sind trotz massiver Forschungsanstrengungen noch weitere fundamentale Forschungsfragen zu lösen, wie etwa
 - Wie kann die Fehleranfälligkeit der Wahrnehmung aller für die aktuelle Fahrzeugführung relevanter Aspekte der Umgebung als „ausreichend klein“ garantiert werden?

Kategorie **Interagieren**

- Wie können Prozesse eines community learnings gestaltet werden, um durch Erfahrungen im Feld die inhärent unvollständigen Taxonomien und Charakterisierungen des Verhaltens anderer Verkehrsteilnehmer zu vervollständigen?
- Absicherung des Austauschs solcher Informationen gegen Cyber-Angriffe, insbesondere sichere Authentifizierung von Kommunikationspartnern und Sicherstellung der Integrität der verwendeten Lagebilder
- Einbeziehung von normierten Qualitätszusagen im Austausch von Lagebildern in Bezug auf Genauigkeit in Zeit und Raum

Kategorie **Vorhersagen**

- Eine besondere Herausforderung ergibt sich bei „mixed traffic“: Wie kann eine wechselseitige Einschätzung der Intentionen zwischen vollautonomen und Fahrern von teilautonomen Fahrzeugen oder Fußgängern erfolgen?
- Individuelle Akzeptanz hochautonomer Systeme: Wie sind hochautonome Systeme zu gestalten, dass die dort verwendeten Prinzipien zur Entscheidungsfindung für unterschiedliche Typen von Nutzern einsehbar sind und zu einem Vertrauen in die hochautonome Führung des Systems führen?
- Gesellschaftliche Akzeptanz: Wie wird der Prozess der Einführung hochautonomer Systeme in einem gesellschaftlichen Diskurs über Potenziale und Risiken begleitet?
- Wenn der Mensch als Handelnder im Loop ersetzt wird durch das technische System: Welchen ethischen und rechtlichen Prinzipien muss die Gestaltung solcher Systeme genügen?

Notwendige **Standardisierungen**:

- Domänenspezifische und domänenübergreifende Standards der auszutauschenden Situationswahrnehmungen und Intentionen

3.2.4 Benötigte Technologien und Zeithorizont

Die folgende Tabelle gibt für jede Ausbaustufe dieser Komplexitätsdimension die größten Herausforderungen sowie die wichtigsten benötigten Technologien zu ihrer Realisierung an und ordnet diese zeitlich ein. Diese Analyse ist die Grundlage für die in Kapitel 7 dargestellte Roadmap.

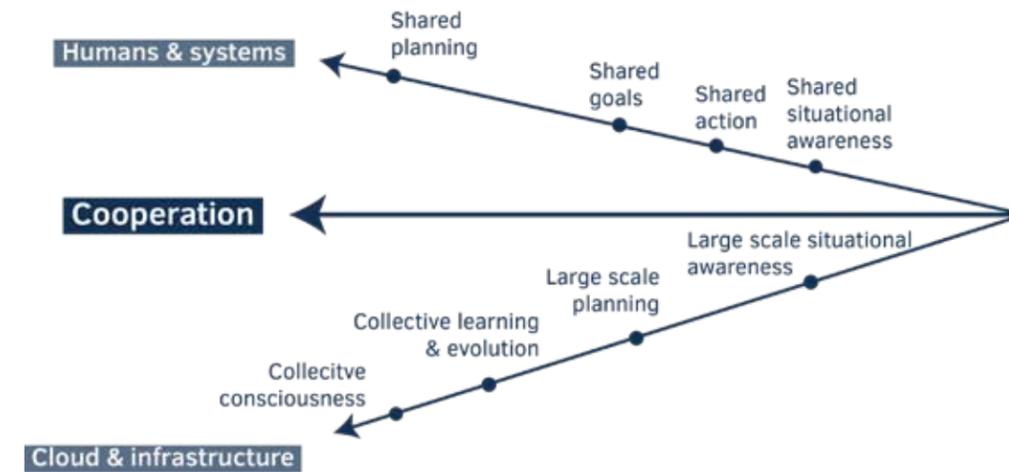
Dimension	Ausbaustufen	Major Challenge	Kerntechnologie	Kurz 2027	Mittel 2035	Lang 2045ff	
Physical Environment Die Fähigkeit des Systems mit verschiedenen Eigenschaften des physikalischen Kontext angemessen umgehen zu können	Phaenotypes	Confidence in Perception Chain	Genügend genaue Modelle von Sensoren		Partiell in bekannten Kontexten	Vollumfänglich	
			Sensor-Fusion, inklusive Auflösen von Widersprüchen in der Wahrnehmung versch. Sensoren		Partiell in bekannten Kontexten	Vollumfänglich	
		Completeness of Identification of relevant Phaenotypes	vgl. Technologien/Ausbaustufen Evolution	in bekannten Kontexten	In partiell bekannten Kontexten	In unbekanntem Kontexten	
	Topology	Zusätzlich:					
		Konfidenz der Lokalisation von Phaenotypen im Lagebild	High Accuracy Digital Maps	Vollumfänglich			
			High-confidence infrastructure-based perception	In bekannten Kontexten	In partiell bekannten Kontexten	In unbekanntem Kontexten	
		Shared Situational Awareness	Ontologie der Phaenotypen und der Topologie	In bekannten Kontexten	In partiell bekannten Kontexten	In unbekanntem Kontexten	
	Dynamics	Zusätzlich:					
		Erkennung von Dynamikeigenschaften der Phaenotypen	Lokale Wahrnehmung relevanter Witterungsveränderungen	Vollumfänglich			
			Lokale Wahrnehmung der Veränderungen der Emissionen	Vollumfänglich			
			Relative Änderung der Lokalisation durch Eigenbewegung	Vollumfänglich			
		Erkennung von Änderungen der Topologie gegenüber der digitale Karte	Durch extreme Witterungsverhältnisse / Umweltbedingungen (Erdbeben, Überschwemmungen)	In bekannten Kontexten	In partiell bekannten Kontexten	In unbekanntem Kontexten	
			Durch „normale“ Ereignisse (Baustellen,...)	In bekannten Kontexten	In partiell bekannten Kontexten	In unbekanntem Kontexten	
		Verhalten des Eigensystems in einem gegebenen (physikalischen) Kontext					
	Predictability	Zusätzlich:					
		Genügend genaue Modelle für die lokale Echtzeit-Vorhersage von Witterungsbedingungen	Shared Situational Awareness	In bekannten Kontexten	In partiell bekannten Kontexten	In unbekanntem Kontexten	
		Genügend genaue Modelle für die lokale Echtzeit-Vorhersage von Änderungen der Topologie	Infrastructure-based Prediction	In bekannten Kontexten	In partiell bekannten Kontexten	In unbekanntem Kontexten	
		Genügend genaue Modelle für die Vorhersage der Beeinflussung des physikalischen Kontexts durch das Egosystem (Phaenotypen wie z.B. Emissionen und Topologie)	(Genügend genaue) Modellbildung	In bekannten Kontexten	In partiell bekannten Kontexten	In unbekanntem Kontexten	

Dimension	Ausbaustufen	Major Challenge	Kerntechnologie	Kurz 2027	Mittel 2035	Lang 2045ff	
Humans and Systems Die Fähigkeit des Systems auf verschiedenen Ebenen mit Menschen und Systemen im Kontext angemessen zu interagieren	Location and Pose (inkl. Mimik, Gestik und deren technischen Äquivalente, wie Sounds/Klang/Lichtgebung...)	Confidence of Perception	Siehe oben				
		Ontologie/Klassifikation der Handlungsfähigkeiten von Menschen und technischen Systemen	Crowd-Sourcing und Data-Analytics	In bekannten Kontexten	In partiell bekannten Kontexten	In unbekanntem Kontexten	
		Confidence of Interpretation of Intention of Pose	Generierung von Modellen der Interpretation... (Kulturabhängig)	In bekannten Kontexten	In partiell bekannten Kontexten	In unbekanntem Kontexten	
			Zeitreihenanalysen	In bekannten Kontexten	In partiell bekannten Kontexten	In unbekanntem Kontexten	
			Intelligent Wearables	In bekannten Kontexten	In partiell bekannten Kontexten	In unbekanntem Kontexten	
	Capability and Skills	Zusätzlich:					
		High Confidence Perception der Wahrnehmungs- und Adaptivitätsfähigkeiten von Menschen und technischen Systemen	Ontologie / Klassifikation der (Handlungs-,) Wahrnehmungs- und Adaptivitätsfähigkeiten von Menschen und technischen Systemen, einschließlich deren Dynamik			In bekannten Kontexten	In partiell oder unbekanntem Kontexten
			Human-State Detection	In bekannten, eingeschränkten Kontexten	In bekannten Kontexten	In partiell oder unbekanntem Kontexten	
			Echtzeit-Erkennung der Zugehörigkeit eines Systems/ eines Menschen zu einer solchen Klasse mit hoher Konfidenz	In bekannten, eingeschränkten Kontexten	In bekannten Kontexten	In partiell oder unbekanntem Kontexten	
			High Confidence Prediction of Dynamics of Human/ System based on identified Capabilities and Skills		In bekannten Kontexten	In partiell bekannten Kontexten	In unbekanntem Kontexten
	Plans and Intent	Zusätzlich:					
			Ontologie von Plänen und Intentionen	In bekannten Kontexten	In partiell bekannten Kontexten	In unbekanntem Kontexten	
		Individual Detection of Plan and Intention	Individualisierung von statistischen Modellen	In bekannten Kontexten	In partiell bekannten Kontexten	In unbekanntem Kontexten	
			Brain-Computer-Interfaces		In bekannten Kontexten	In partiellen und unbekanntem Kontexten	
			Shared Planning	In bekannten Kontexten	In partiell bekannten Kontexten	In unbekanntem Kontexten	
		Optimal koordinierte Dynamik zwischen Egosystem und anderen Systemen/Menschen	Antizipation der Auswirkungen der Handlungen des Egosystems auf Pläne und Intentionen anderer Systeme und Menschen		In bekannten Kontexten	In partiellen und unbekanntem Kontexten	
			Synthese von Plänen zur optimal koordinierten Dynamik		In bekannten Kontexten	In partiellen und unbekanntem Kontexten	
		Vermeidung von Oszillationen in der wechselseitigen Beeinflussung von Plänen des Egosystems und Plänen anderer Systeme und Menschen			In bekannten Kontexten	In partiellen und unbekanntem Kontexten	
		Konfliktlösungsstrategien	Incentives zur Reduktion eigener Ziele	In bekannten Kontexten	In partiell bekannten Kontexten	In unbekanntem Kontexten	
	Erkennung von malicious actors	Anomalie-Detection	In bekannten eingeschränkten Kontexten	In bekannten Kontexten	In partiell bekannten und unbekanntem Kontexten		

Dimension	Ausbaustufen	Major Challenge	Kerntechnologie	Kurz 2027	Mittel 2035	Lang 2045ff
	Group Interaction	Zusätzlich:				
		Erkennen von Schwarmverhalten	Verhaltensmodelle von Schwärmen	In bekannten eingeschränkten Kontexten	In bekannten Kontexten	In partiell bekannten und unbekanntem Kontexten
			Perzeptionsmodelle für Schwarmverhalten	In bekannten eingeschränkten Kontexten	In bekannten Kontexten	In partiell bekannten und unbekanntem Kontexten
		Auswirkungen von Schwarmintelligenz auf Dynamik des Gesamtsystems und des Kontexts	Einbeziehung von Dynamikmodellen von Schwärmen in Lageentwicklung		In bekannten eingeschränkten Kontexten	In partiell bekannten und unbekanntem Kontexten
		Erkennung von malicious swarms	Anomalie-Detection	In bekannten eingeschränkten Kontexten	In bekannten Kontexten	In partiell bekannten und unbekanntem Kontexten
	Self Evolution Capabilities	Zusätzlich:				
		Erkennung der Selbst-Evolutionsfähigkeiten und -stufen von umgebenden Menschen und Systemen				Nicht in den nächsten 50 Jahren
		Antizipation der Dynamik in Extremsituationen				Nicht in den nächsten 50 Jahren

4 Kooperation

4.1 Zusammenfassung



Dieser Abschnitt beschreibt die erwartete Entwicklung der Kooperation zwischen Systemen oder Gruppen von Systemen sowie mit Menschen. Sie umfasst Systemkollektive wie ein Platoon, Systemverbände einer Domäne wie Mobilität oder auch Verbände verschiedenen Domänen wie z.B. Mobilität und Energie. Dabei gehen wir davon aus, dass die Kooperation spontan erfolgen kann und die Kooperation situativ angepasst wird. Dazu betrachten wir zum einen die Kooperationsvoraussetzungen und zum anderen die Kooperationsformen auf verschiedenen Kooperationsebenen.

Immer dann, wenn die eigene Stärke eines Systems nicht ausreicht, um in einem gegebenen Kontext seine Ziele durchzusetzen, kann das Egosystem andere Systeme oder Menschen der Systemumgebung anfragen, für einen abgestimmten Zeitraum das Egosystem in der Umsetzung seiner Ziele zu unterstützen. Wir unterscheiden die folgenden Ausbaustufen von Kooperationsformen, deren Gemeinsamkeit darin besteht, dass die potenziellen Kooperationspartner das Egosystem tendenziell unterstützen.

Ausbaustufen Kooperation mit anderen Systemen und/oder Menschen

1. **Gemeinsames Situationsbewusstsein:** Die Kooperationspartner tauschen alles Wissen über den jeweiligen Systemkontext aus, soweit dieses für die Ziele des Egosystems relevant ist.
2. **Geteilte Aktion:** Die Kooperationspartner führen koordiniert Manöver zur Erreichung der Ziele des Egosystems durch. **Geteilte Ziele:** Die Kooperationspartner stimmen sich ab, welche Ziele gemeinsam verfolgt werden.
3. **Geteilte Pläne:** Die Kooperationspartner stimmen ihre Pläne und Strategien zur Erreichung dieser Ziele ab.

In der Umsetzung seiner Ziele wird das Egosystem in der Regel auch durch über die Cloud bereitgestellte Informationen und Fähigkeiten unterstützt. Ein zentrales Instrument dafür sind Digital Twins, welche auf der Basis von im Feld über Infrastruktur oder Teilnehmer eines Kollektivs wahrgenommenen Daten nicht nur Informationen zu einem globalen Lagebild zusammenfügen können, sondern auf der Basis von

Prädiktionen der Weiterentwicklung des globalen Lagebildes optimale Pläne zur Erreichung der Ziele des Kollektivs ermitteln und an das Kollektiv übertragen können, wie dies etwa im Verkehrsoptimierungsbeispiel in Kapitel 2 veranschaulicht wurde. Auf der Basis der Rückmeldungen aus dem Feld kann im Digital Twin auch ein Lernvorgang für das Kollektiv realisiert werden, der z. B. zu Verfahren in der Objektidentifikation und Prädiktion mit höherer Konfidenz führen kann, die dann zu over-the-air updates der entsprechenden Teilsysteme der Teilnehmer des Kollektivs führen. Schließlich können darüber für Kollektive die Auswirkungen von Plänen in Bezug auf mögliche Zielkonflikte mit übergeordneten gesellschaftlichen oder staatlichen Zielsetzungen untersucht werden, etwa die Auswirkungen in Bezug auf die Erreichung der vereinbarten Klimaziele.

Ausbaustufen Kooperation mit Cloud und Infrastruktur

1. Situationsbewusstsein in großem Maßstab: Die Partner des Kollektivs teilen – direkt oder über die Cloud – ein gemeinsames globales Lagebild.
2. Großplanung: Auf der Basis eines globalen Lagebildes und der Fähigkeit der Prädiktion der Weiterentwicklung des Lagebildes werden für das Kollektiv optimale Pläne zur Erreichung seiner Ziele erstellt.
3. Kollektives Lernen und Evolution: Aus Erfahrungen der Partner des Kollektivs in der Beherrschung von Perzeptions- und Führungsaufgaben werden verbesserte Versionen der dafür verwendeten Teilsysteme gelernt und über over-the-air updates den Partnern des Kollektivs zur Verfügung gestellt.
4. Kollektives Bewusstsein: Die Auswirkungen von Großplanungen in Bezug auf übergeordnete gesellschaftliche, ethische, staatliche Zielsetzungen werden in die Auswahl von Plänen so eingeschlossen, dass diese nicht verletzt werden.

4.2 Detaillierte Beschreibung

4.2.1 Analyse

Analyse der Veränderung von technischen Kooperationen

Voraussetzungen für Kooperation

Damit Kooperation gelingen kann bedarf es einer Reihe von Voraussetzungen, die je nach Kooperationsform unterschiedlich ausgeprägt sind. Es stellen sich folgende Fragen:

- Kommunikation:
 - Ist eine für die Kooperation geeignete Kommunikationsinfrastruktur vorhanden? (Topologie, Reaktivität, Bandbreite, Sicherheit, u. a.)
- Gemeinsame Sprache (Syntax, Semantik, Ontologie):
 - Sind die geeigneten Kooperationsprotokolle mit den kooperationspezifischen Sprachelementen in Syntax und Semantik vorhanden?
- Gemeinsame Situation bzw. Kontext:
 - Haben die Kooperationspartner eine gemeinsame Sicht auf den Kontext und die Situation in der kooperiert werden soll und damit die gleiche Basis zur Entscheidung?
- Gemeinsame Menge von Fähigkeiten:
 - Haben die Kooperationspartner gemeinsames Wissen über die Fähigkeiten des Partners und können so gemeinsames Handeln abstimmen?
- Gemeinsame Ziele und Strategie:

- Haben die Kooperationspartner gemeinsames Wissen über die Ziele der jeweils anderen und die Kooperationsziele (Optimierungsgrößen) und die Strategien (Handlungsoptionen zur Zielerreichung), um in eine Abstimmung über die jeweiligen Beiträge der Partner zu den abgestimmten Zielen einzutreten?
- Gemeinsames Belohnungssystem:
- Gibt es ein gemeinsames Verrechnungssystem zwischen den Partnern, um Nutzen für alle herstellen zu können?
 - Beispiel: Warum sollte im Platoon einer an der Spitze fahren?

Die folgende Tabelle gibt einen Überblick über die Konkretisierung dieser Konzepte entlang der Vernetzungshierarchie:

CPS	Individuelle Kooperation bedeutet: spontane situative Kooperation zweier CPS. Kooperationsraum bilateral	Einfädeln lassen. Oder infrastrukturbasiertes Fahren
Gruppe von CPS	Gruppen Kooperation bedeutet: mehrere CPS bilden strategisch eine Gruppe, um gemeinsame Ziele zu erreichen. Kooperationsraum: Homogen, eine Ebene multilateral	Platoon zur Reduktion von Verbrauch und Emission
Kollektiv von CPS	Kooperation homogener Kollektive (eine Domäne, z. B. Mobilität) bedeutet: CPS und Gruppen von CPS bilden strategisch ein Kollektiv um ein übergeordnetes Ziel wie Emission, Durchsatz, Ressourcenoptimierung, Sicherheit u.a.m. zu erreichen. Kooperationsraum homogen mehrere hierarchische Ebenen multilateral	Verkehrslenkung
Heterogenes Kollektiv von CPS	Kooperation heterogener Kollektive bedeutet, CPS, Gruppen und homogene Kollektive aus verschiedenen Domänen (Mobilität, Energie, Gesundheit, ...) bilden ein strategisches, heterogenes Kollektiv, um übergeordnete Ziele wie Sicherheit, Emission, u.a.m. zu erreichen. Kooperationsraum: heterogen mehrere hierarchische multilaterale Ebenen,	CO2 neutrale Smart City

Formen der Kooperation

Kooperation zwischen Partnern kann auf verschiedenen Ebenen in verschiedenen Formen ablaufen. Betrachtet man zunächst die direkte Kooperation von technischen Systemen untereinander oder mit Menschen so ergeben sich die folgenden Formen:

- **Informationsaustausch** ist die einfachste Form: z.B. liefert ein Fahrzeug im Sinne eines erweiterten Sensors Information über die Verkehrssituation, die von anderen Fahrzeugen nicht eingesehen werden kann (Stau voraus, Glatteis, ...). Dies kann direkt Car-to-Car oder indirekt über einen Cloud Service geschehen. Ggf. gibt es keine direkte Beziehung zwischen Provider und Consumer.
- Kernfragen sind:
 - Sicherstellen der Gültigkeit (Aktualität und Verlässlichkeit) der Informationen.

- Übertragbarkeit der Information von einem System in ein anderes (Syntax und Semantik).
- **Kooperatives Handeln** erfordert einen gemeinsamen Handlungsraum (Kooperationskontext), eine Sicht über die jeweiligen Fähigkeiten der Partner, eine gemeinsame Intention und ein abgestimmtes Handlungsmuster.
 - z. B. einfädeln lassen auf der Autobahn oder Querstraße. Da es sich nur um ein einzelnes Manöver handelt ist die Planungskomponente rudimentär.
- **Kooperative Abstimmung** von Strategie und Zielen bedeutet Optimierung kollektiver Ziele zu Gunsten individueller Ziele und kollektives Handeln zugunsten individuellen Handelns. Damit stellt sich die Frage inwieweit ein System seine individuellen Ziele und Strategien den kollektiven Zielen unterordnen kann und was es dafür im Ausgleich bekommt? Bsp. Warum sollte in einem Platoon ein Fahrzeug an der Spitze fahren? Es spart keine Energie und ist auch nicht schneller und kann auch nicht flexibel auf einzelne Situationen reagieren.
- **Kooperatives Planen** setzt ein gemeinsames Bild der Aufgabe voraus, gemeinsame Handlungsmuster, die Fähigkeit Sequenzen von Handlungsmustern zur Erfüllung der Aufgabe zu bilden, nach bestimmten Kriterien zu bewerten und auszuwählen sowie diese Handlungsmuster in kooperatives Handeln aufzulösen, dem die Partner zustimmen können.
 - z.B. Aufgabe Fertigen eines komplexen Bauteils mit Maschinen mit verschiedenen Sets von Fähigkeiten oder die Arbeitsteilung bei der Logistik von komplexen Warenströmen mit unterschiedlichen Fahrzeugen.

Bezieht man die Cloud sowie die Infrastruktur in die Kooperation ein, so erweitern sich die möglichen Formen der Kooperation und es ergeben sich weitere Möglichkeiten:

- **großflächiger Informationsaustausch und Großflächiges kooperatives Planen**
Durch die Einbeziehung der Cloud und der Infrastruktur können gemeinsame Lagebilder und gemeinsame Pläne auch zwischen geographisch entfernten (individuellen, Gruppen oder Kollektiven von) Systemen erstellt und ausgetauscht werden.
- **kooperatives Lernen und Evolution**
bedeutet, dass ein System von Erfahrungen lernen kann, die ein anderes System in seinem operationalen Kontext gemacht hat. Dies kann auf allen Ebenen der Kooperation geschehen. Kernfrage ist hier die Frage der Übertragbarkeit und der Auf/Abwärtskompatibilität und die Frage in wieweit Systeme die individuell gelernt haben, noch kooperationsfähig sind.
- **kooperatives Bewusstsein:** Die Auswirkungen von Großplanungen in Bezug auf übergeordnete gesellschaftliche, ethische, staatliche Zielsetzungen werden in die Auswahl von Plänen so eingeschlossen, dass diese Ziele nicht verletzt werden.

Wiewohl bei all diesen Kooperationsformen davon ausgegangen wird, dass potenzielle Kooperationspartner dem Egosystem tendenziell wohlgesonnen sind, so ist zwingend auch der Umgang mit **antagonistische Handlungen** vorzusehen. Diese beschreiben das Verhalten eines Systems unter der Annahme, dass Anteile der Systemteilnehmer sich unkooperativ oder gezielt schädlich verhalten. Antagonistisches Verhalten muss in allen Prozessschritten, Aggregations- und technischen Ebenen erwartet und behandelt werden können. Antagonistische Akteure verfügen dabei über ähnliche Capabilities auf den jeweiligen Entwicklungsstufen wie das Egosystem.

Aus den obigen Betrachtungen ergeben sich folgende **Kooperationsebenen**:

Wahrnehmungsebene: Systeme kooperieren auf dieser Ebene, um aus den Einzelwahrnehmungen der Teilsysteme ein konsistentes Bild des relevanten Kontextes zu erzeugen und zu aktualisieren.

Dabei sind die Kategorien ‚bekannte determinierte‘, ‚bekannte nicht determinierte‘, ‚teilweise bekannte‘ und ‚unbekannte Wahrnehmungen‘ zu unterscheiden.

Beispiel: Fahrzeuge als Sensoren tragen dazu bei ein vollständiges Bild einer Verkehrssituation zu erzeugen.

Aktionsebene: Systeme kooperieren auf dieser Ebene, um koordiniert einen Plan umzusetzen.

Beispiel: Eine Schar von Fertigungs-Robotern erfüllt koordiniert eine Menge von Fertigungsaufträgen möglichst optimal.

Planungs-/ Entscheidungsebene: Systeme kooperieren auf dieser Ebene, um einen Plan zur Erfüllung einer Mission in der Weise abzustimmen, dass jedes Teilsystem seine Fähigkeiten optimal einsetzen kann, seine Mission erfüllen kann und gemeinsame Ressourcen optimal genutzt werden.

Beispiel: Eine Schar von Fertigungsrobotern plant gemeinsam die Abarbeitung der Fertigungsaufträge.

Strategieebene: Systeme kooperieren auf dieser Ebene, um die Ansätze, mit denen die Ziele der Kooperation erreicht werden, miteinander auszuhandeln. Dabei müssen ggf. Konflikte aufgelöst werden.

Beispiel: Bestimmen des ersten Fahrzeugs in einem Platoon (der einzige der keinen Vorteil hat).

Missionsebene: Systeme kooperieren auf dieser Ebene, um übergeordnete Ziele zu erfüllen, auch wenn das bedeutet, dass sie ihre eigenen Ziele nur suboptimal erreichen können und dafür ggf. entschädigt werden müssen.

Beispiel: Ein eiliger Transportauftrag bekommt im Verkehrssystem Vorrang und ist bereit für diesen Quality-of-Service betroffene Verkehrsteilnehmer zu entschädigen.

Value Governance Ebene: Systeme kooperieren auf dieser Ebene um sicherzustellen, dass die Entscheidungen und Aktionen der Kooperationsgemeinschaft den ethischen, juristischen und kulturell akzeptierten Vereinbarungen entsprechen. Zudem die Fähigkeit zum Aushandeln verschiedener Wertesysteme in einem Verbund.

Wir instanzieren diese Konzepte entlang der Aggregationshierarchie:

Individuelle CPS:

	Systemspezifische Beschreibung der Kooperationsebenen	
Wahrnehmungsebene	Spontane, bilaterale Kooperation von Einzelsystemen, um aus den Einzelwahrnehmungen der Teilsysteme ein konsistentes Bild des relevanten Kontextes zu erzeugen und zu aktualisieren.	Fahrzeug liefert Verkehrs-, Wetter-, Straßenzustandsdaten u.a.m. an eine zentrale Stelle ohne genau zu wissen wie sie verwendet oder wozu sie verwendet werden.
Aktionsebene	Spontane, bilaterale Kooperation von Einzelsystemen, um koordiniert einen Plan durch lokale Aktionen umzusetzen.	Fahrzeug passt seine Fahrweise individuell an den ausgehandelten QoS für Verkehrsfluss an

Planung- und Entscheidungsebene	Spontane, bilaterale Kooperation von Einzelsystemen um einen Plan so abzustimmen, dass jedes Teilsystem seine Fähigkeiten optimal einsetzen kann, seine Mission erfüllen kann und gemeinsame Ressourcen optimal genutzt werden.	Infrastrukturbasiertes Fahren verteilt die Entscheidung über Manöverplanung auf Infrastruktur mit globalem Kontextwissen und dem CPS mit lokalem Kontextwissen.
Strategieebene	Spontane, bilaterale Kooperation von Einzelsystemen, um die Ansätze und Prioritäten mit dem die Ziele der Kooperation erreicht werden mit einander auszuhandeln.	Aushandeln des QoS in einer Mobility Situation
Missionsebene	Spontane, bilaterale Kooperation von Einzelsystemen zur Optimierung lokaler und globaler Ressourcen	Mission ist: Suche einen Parkplatz in der Nähe. Sie wird durch Infrastruktur in der Weise kooperativ koordiniert, dass nicht nur alle freien Parkplätze in der Nähe angezeigt werden sondern die Fahrzeuge koordiniert zu einzelnen freien Plätzen dirigiert werden.
Value Governance	Spontane, bi- oder multi-laterale Kooperation von Einzelsystemen zur Einhaltung von Governance Regeln, ggfs. unter Einbindung einer zentralen Instanz in der Infrastruktur.	Einhalten von rechtlichen, moralischen und kulturellen Regeln durch Abstimmung von Strategie und Manöverplanung mit einer zentralen Instanz in der Infrastruktur.

Gruppen von CPS:

	Systemspezifische Beschreibung der Kooperationsebenen	
Wahrnehmungsebene	Temporäre, multilaterale Kooperation von Einzelsystemen und Gruppen, um aus den Einzelwahrnehmungen der Teilsysteme ein konsistentes Bild des relevanten Kontextes zu erzeugen und zu aktualisieren.	Jedes Fahrzeug liefert „seinen“ Teil der Kontextwahrnehmung kontinuierlich an alle Teilnehmer eines Platoons; aus all diesen Teilbildern entsteht damit ein komplettes, valides Bild der aktuellen Situation, aus dem ein Lagebild abgeleitet werden kann.
Aktionsebene	Temporäre, multilaterale Kooperation von Einzelsystemen und Gruppen, um koordiniert einen Plan durch lokale Aktionen umzusetzen.	Koordination der Aktionen der Einzelfahrzeuge beim Spurwechsel eines Platoons. Dabei entsteht die Situation, dass die führenden Fahrzeuge die Spur schon gewechselt haben, während die mittleren Fahrzeuge sich zwischen zwei Spuren befinden und die hinteren Fahrzeuge noch auf der ursprünglichen Spur sind.
Planung- und Entscheidungsebene	Temporäre, multilaterale Kooperation von Einzelsystemen und Gruppen, um einen Plan oder die Erfüllung einer Mission in der Weise abzustimmen, dass jedes Teilsystem seine Fähigkeiten optimal einsetzen kann, seine Mission erfüllen kann und gemeinsame Ressourcen optimal genutzt werden	Platoon plant und entscheidet was das angemessene Manöver in einer gegebenen Situation ist. Dabei werden die aktuellen Fähigkeiten der Einzelfahrzeuge berücksichtigt.

Strategieebene	Temporäre, multilaterale Kooperation von Einzelsystemen und Gruppen, um die Ansätze mit denen die Ziele der Kooperation erreicht werden, miteinander auszuhandeln	Ein Platoon entscheidet welcher Fahrstrategie (sicher, effizient, schnell, Ressourcen schonend, emissionsarm, ..) das Platoon folgen will. Dabei können einzelne Fahrzeuge unterschiedliche Präferenzen haben, die ausgehandelt werden müssen. Davon kann auch die Entscheidung eines Einzelfahrzeugs abhängen sich einem Platoon anzuschließen.
Missionsebene	Temporäre, multilaterale Kooperation von Einzelsystemen und Gruppen, um Interessenausgleiche in der Kooperation bei ungleicher Verteilung von Lasten zu erzielen.	Aushandeln welches Fahrzeug das Führungsfahrzeug ist. (Führungsfahrzeug hat selbst den geringsten Nutzen)
Value Governance	Value-Awareness in temporäre, multilaterale Kooperation von Einzelsystemen und Gruppen.	Einhalten von rechtlichen, moralischen und kulturellen Regeln durch Abstimmung von Strategie und Manöverplanung mit einer zentralen Instanz in der Infrastruktur. Ggf. auch nur transparent machen der Entscheidungen.

Kollektive von CPS:

	Systemspezifische Beschreibung der Kooperationsebenen	
Wahrnehmungsebene	Dauerhafte, multilaterale, hierarchische Kooperation von Einzelsystemen, Gruppen und Kollektiven um aus den Einzelwahrnehmungen der Teilsysteme ein konsistentes Bild des relevanten Kontextes zu erzeugen und zu aktualisieren.	Reibwertkarte
Aktionsebene	Dauerhafte, multilaterale, hierarchische Kooperation von Einzelsystemen, Gruppen und Kollektiven, um koordiniert einen Plan durch lokale Aktionen umzusetzen.	Verkehrslenkung um Emissionen in einem Verkehrsabschnitt gering zu halten.
Planung- und Entscheidungsebene	Dauerhafte, multilaterale, hierarchische Kooperation von Einzelsystemen und Gruppen und Kollektiven, um einen Plan zur Erfüllung einer Mission in der Weise abzustimmen, dass jedes Teilsystem seine Fähigkeiten optimal einsetzen kann, seine Mission erfüllen kann und gemeinsame Ressourcen optimal genutzt werden	tbd
Strategieebene	Dauerhafte, multilaterale, hierarchische Kooperation von Einzelsystemen, Gruppen und Kollektiven um die Ansätze mit dem die Ziele der Kooperation erreicht werden, miteinander auszuhandeln.	Aushandeln, ob langsam Fahren oder Verbote einzelner Fahrzeugtypen oder ein anderer Ansatz die Interessen und die Emissionslage angemessener berücksichtigt.
Missionsebene	Dauerhafte, multilaterale, hierarchische Kooperation von Einzelsystemen, Gruppen und Kollektiven um die Missionen der Teilsysteme im Sinn der Gesamtmission aufeinander abzustimmen.	Aushandeln, welche Gruppe welchen Beitrag zur Erhöhung des Durchsatzes in einem Verkehrsraum liefern kann.

Value Governance	Value-Awareness in dauerhaften, multilateralen hierarchischen Kooperationen von Einzelsystemen, Gruppen und Kollektiven.	Einhalten von rechtlichen, moralischen und kulturellen Regeln durch Abstimmung von Strategie und Manöverplanung mit einer zentralen Instanz in der Infrastruktur. Ggf. auch nur transparent machen der Entscheidungen.
------------------	--------------------------------------------------------------------------------------------------------------------------	------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

Heterogene Kollektive von CPS:

	Systemspezifische Beschreibung der Kooperationsebenen	
Wahrnehmungsebene	Dauerhafte, multilaterale, hierarchische, heterogene domänenübergreifende Kooperation von Einzelsystemen, Gruppen und Kollektiven um aus den Einzelwahrnehmungen der Teilsysteme ein konsistentes Bild des relevanten Kontextes zu erzeugen und zu aktualisieren.	Energie- oder Emission- Karte, Staukarte aus Handy-Bewegungen.
Aktionsebene	Dauerhafte, multilaterale, hierarchische, heterogene Kooperation von Einzelsystemen, Gruppen und Kollektiven um koordiniert einen Plan durch lokale Aktionen umzusetzen.	Verkehrslenkung, um Emissionen in einem Verkehrsabschnitt gering zu halten.
Planung- und Entscheidungsebene	Dauerhafte, multilaterale, hierarchische, heterogene Kooperation von Einzelsystemen und Gruppen und Kollektiven um einen Plan Erfüllung einer Mission in der Weise abzustimmen, dass jedes Teilsystem seine Fähigkeiten optimal einsetzen kann, seine Mission erfüllen kann und gemeinsame Ressourcen optimal genutzt werden	Kooperative Planung von diversen Energie- und Verkehrssystemen um bestmögliche Emissionswerte zu erhalten.
Strategieebene	Dauerhafte, multilaterale, hierarchische, heterogene Kooperation von Einzelsystemen, Gruppen und Kollektiven um die Ansätze, mit denen die Ziele der Kooperation erreicht werden, mit einander auszuhandeln.	Aushandeln, welche Domäne welchen Beitrag zur Zielerreichung leisten muss unter Berücksichtigung roter Linien (Mindesterhalt von QoS in jeder Domäne)
Missionsebene	Dauerhafte, multilaterale, hierarchische heterogene Kooperation von Einzelsystemen, Gruppen und Kollektiven um die Missionen der Teilsysteme im Sinn der Gesamtmission aufeinander abzustimmen.	Aushandeln, welche Gruppe oder Domäne welchen Beitrag zur Erreichung eines Healthy Smart City-Zieles liefern kann.
Value Governance	Value-Awareness in dauerhaften, multilateralen, hierarchischen, heterogenen Kooperationen von Einzelsystemen, Gruppen und Kollektiven	Einhalten von rechtlichen, moralischen und kulturellen Regeln durch Abstimmung von Strategie und Manöverplanung mit einer zentralen Instanz in der Infrastruktur. Ggf. auch nur transparent machen der Entscheidungen.

4.2.2 Capabilities

Die Grundmechanismen kooperierender Systeme unterscheiden sich grundsätzlich nicht von Systemen mit evolutionärer und autonomer Handlungsdynamik wie in Kapitel 3.2 beschrieben. Diese sind Sense, Analyse, Predict, Reflect, Decide, Certify und Act. Zusätzlich benötigten koope-

rierende Systeme auf allen Aggregationsstufen und allen Kooperationsebenen verschiedene Ausprägungen von kooperativen Fähigkeiten wie:

- **Kooperative Kommunikationsfähigkeit** bedeutet einen gemeinsamen Kanal, gemeinsame Protokolle und einen gemeinsamen Begriffsraum (Ontologie) zu haben. ggf. gehört hierzu auch die Fähigkeit der Transduktion (Übersetzen mit Hilfe einer Meta-Sprache).
- **Kooperative Transaktionsfähigkeit** bedeutet eine Menge von Interaktionen zu einer Transaktion so zu bündeln, dass sie nach dem AKID Prinzip ablaufen. Atomar (ganz oder gar nicht, ggf. Rollback), Konsistent (hinterlässt einen systemspezifischen konsistenten Zustand nach der Transaktion), Isoliert (ist von nebenläufigen Transaktionen nicht beeinflusst), Dauerhaft (der Veränderte Systemzustand ist in allen Systemteilen repräsentiert) (Rollback von Physikalischen Aktivitäten praktisch nicht möglich) aber Bsp.: Rollback von SW-Upgrades in evolutionären System ist möglich. Oder Rollback das das System in einen integren Zustand versetzt, der dem Zustand vor der Transaktion entspricht.
- **Kooperative Kontext-awareness** bedeutet, dass kooperierende Systeme ihren Kontext beobachten, interpretieren, um ein gemeinsames Lagebild zu entwickeln, was die Basis für das kooperative Entscheiden und Handeln ist.
- **Kooperative Selbst-Adaption** bedeutet, dass Systeme sich kooperativ selbst limitieren, heilen, reorganisieren und ggf. auch erweitern können.
- **Kooperative Lernfähigkeit** bedeutet, dass ein System sich Unbekanntes erschließen und seine Handlungsoptionen angemessen anpassen kann. Es bedeutet auch, dass ein System, je mehr es lernt, seine Lernfähigkeit selbst verbessert. Hier wird Lernen zur Kooperationsaufgabe im Verbund.
- **Kooperative Autonomiefähigkeit**: Die Fähigkeit von kooperierenden Systemen in komplexen Kontexten angemessen und kooperativ zu handeln, ohne steuernde Einwirkung von Dritten.
- **Kooperative Anschlussfähigkeit (Composability)** bedeutet in dem Sinn anschlussfähig an eine Gruppe oder Kollektiv zu sein, dass die Wirkfähigkeit eines Einzelsystems Teil einer umfassenderen Wirkkette sein kann, ohne die eigene Systemintegrität aufzugeben. (Holon: Ein Ganzes, dass ein Teil eines Ganzen ist.)
Beispiel: Fahrzeug in einem Platoon, oder Schwarm von Drohnen,...
- **Kooperative Vertrauenswürdigkeit** bedeutet, dass ein System in einem kooperativen Systemverbund Vertrauen aufbauen und erhalten kann um von anderen Systemen als vertrauenswürdiger Kooperationspartner wahrgenommen zu werden, der seine Zusagen wie vereinbart einhält.
- **Kooperative Verhandlungsfähigkeit** bedeutet, dass ein System die Freiheitsgrade seiner Ziel- und Handlungsoptionen kennt, ihren Wert einschätzen kann und darüber mit anderen Systemen in einem kooperativen Verbund in eine verbindliche Verhandlung eintreten kann.
- **Kooperative Konfliktfähigkeit** bedeutet, dass ein System seine Grenzen und Zielsetzungen kennt, deren Überschreitung oder Verletzung kooperativ dialogisch verhandeln kann. Das beinhaltet die Fähigkeit zum Kompromiss sowie den Ausschlag von Kooperationsbeziehungen mit antagonistischen oder nicht vertrauenswürdigen Partnern.

4.2.3 Herausforderungen und Forschungsfragen

Problemanalyse – Systemanalyse

- Welche Kooperationsformen gibt es, und welche Kooperationsmuster brauchen wir dafür? Was ist die geeignete Kommunikationsinfrastruktur dafür?
- Wie werden kooperationsfähige Systeme formal beschrieben, die spontan mit anderen Systeme-

men eine Kooperation eingehen können?

- Wie werden Kooperationskontexte formal beschrieben und wie sehen Synchronisationsmuster zwischen den Partnern aus?
- Wie werden Ziele und Strategie formal beschrieben und wie sehen Abstimmungsprozeduren dafür aus?
- In wie weit sind die Wissens- und die Verhaltensspezifikationen übertragbar oder wie heterogen können Systeme in einem Kollektiv sein?
- Wie kann die Gültigkeit von Information und Abstimmung über dem gesamten Kooperationszeitraum valide gehalten werden?
- Wie kann die Kohärenz der Systeme in einem Kollektiv von Systemen hergestellt und aufrechterhalten werden
- Erstellung geeigneter (formaler) Ontologien
 - für die verschiedenen Diskurse: Raum/Zeit/etc.
 - für die Kooperation(Verhandlungen/Entscheidungen)
 - in der Art der Believe-Design-Intention (BDI), Incentives, Regeln, Assumptions/Garanties
 - für die Exploration gemeinsamer Kooperationen
 - gemeinsame Interessen, Nutzen, Ontology-Matching

Systementwurf und Realisierung

- Kompositionelle Architekturen für Kooperationen in ihren verschiedenen Formen.
- Mechanismen zum lokalen Begrenzen von Zielen einzelner Partner.
- Mechanismen zum verteilten Abgleich von Kooperationskontextinformation in einem Kooperationsintervall.
- Mechanismen zum verteilten Abgleich von Zielen und Strategien inklusive der lokalen Degradation.
- Mechanismen zum Aushandeln von Verrechnungseinheiten, um Nutzengefälle im Kooperationsverbund auszugleichen. Mechanismen/Modelle zur Gründung, Betrieb und Auflösung von Kooperation
- (tbd.: Zusammenführen mit entsprechenden Texten aus „Kontext“ (von dort hierhin verschieben und zusammenführen)
- Skalierung von Kooperationen
 - Kooperation von Kooperationen: Interessensausgleich zwischen den Kooperation, Kooperationsvertreter/Sprecher
 - Vertrauen, Kooperationswillen, Mandate: Gruppen von Kooperationspartnern verhandeln Kooperation untereinander; dies kann dazu führen, dass individuelle Ziele von Gruppenmitgliedern weniger stark berücksichtigt werden müssen; dies wiederum kann zu Austritt aus der Gruppe und ggfs. Neuformierungen führen (Beispiel aus Soziologie: Brexit)
- Mechanismen zur (on-line) Verifikation von Kooperationen (oder deren Angebote)
 - Safety & Security (Wer garantiert was, Schutzmechanismen,...)
 - sonstige: Wirtschaftlichkeit, Erfolgswahrscheinlichkeit, Güte der Kooperation, etc.
 - (Anmerkung: Weitere Einbindung von Sozialwissenschaftlern anstreben)

4.2.4 Benötigte Technologien und Zeithorizont

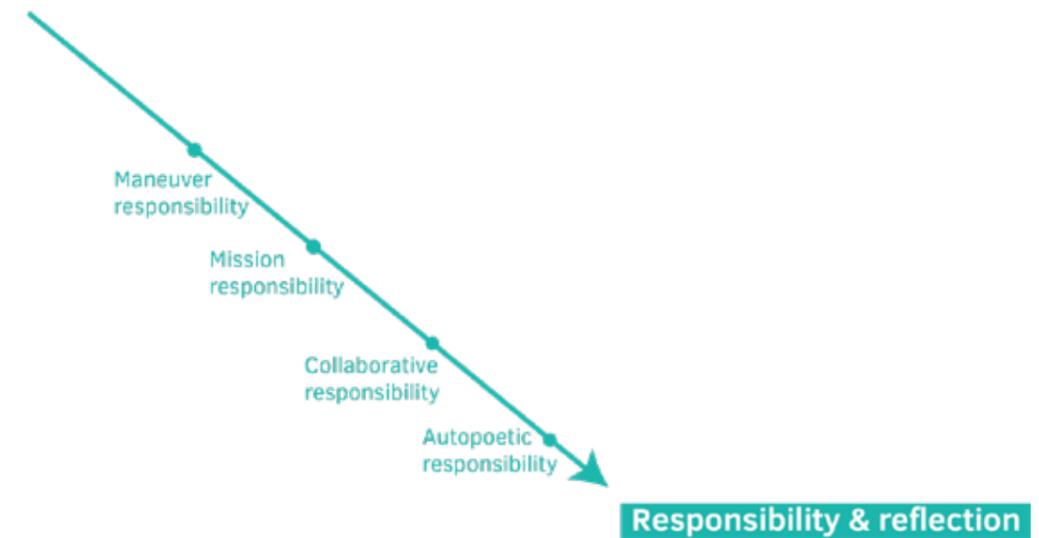
Die folgende Tabelle gibt für jede Ausbaustufe dieser Komplexitätsdimension die größten Herausforderungen sowie die wichtigsten benötigten Technologien zu ihrer Realisierung an und ordnet diese zeitlich ein. Diese Analyse ist die Grundlage für die in Kapitel 7 dargestellte Roadmap. (Tabelle siehe folgende Seite.)

Dimension	Ausbaustufen	Major Challenge	Kerntechnologie	Kurz 2027	Mittel 2035	Lang 2045ff	
Cloud and Infrastructure Die Fähigkeit Large Scale, d. h., über die eigene Einheit hinaus mit Cloud und Infrastruktur zu kooperieren	Large Scale Situational Awareness	Gemeinsame Ontologien	Zusammenführung von Domänenspezifischen Ontologien (für Pläne) bei heterogenen Kollektiven	Vollumfänglich			
		Sicherstellung der Integrität des globalen Lagebildes		Vollumfänglich			
			Crowd-Sourcing	Vollumfänglich			
			Real-time secure communication (5G)	Vollumfänglich			
			Echtzeitfähige cloud services	Vollumfänglich			
			Annotation der Wahrnehmungskonfidenzen	Vollumfänglich			
	Large Scale Planing	Ontologie für Pläne	Zusammenführung von Domänenspezifischen Ontologien bei heterogenen Kollektiven	Vollumfänglich			
		Ontologie für Skills und Capabilities	Kenntnis der Capabilities und Skills der beteiligten Systeme	Vollumfänglich			
			Verteilte Strategie-Synthese in hierarchischen Spielen mit unvollständiger Information	Vollumfänglich			
		Sicherstellung der Integrität der Pläne		Vollumfänglich			
	Collective Learning and Evolution	Zusätzlich:					
		Gemeinsame Hypothesenbildung				Eingeschränkt	Vollumfänglich
		Gemeinsames Design von Experimenten zur Hypothesenüberprüfung				Eingeschränkt	Vollumfänglich
		Gemeinsame Durchführung von Experimenten zur Hypothesenbildung				Eingeschränkt	Vollumfänglich
		Gemeinsame Auswertung der Experimente				Eingeschränkt	Vollumfänglich
		Gemeinsame Angleichung der Hypothesen und ggfs. Erweiterung der Ontologien				Eingeschränkt	Vollumfänglich
		Weitergabe von Lernerfahrungen an andere	Ontologien für Problem- und Lösungsklassen	Für einfache Kontexte	Eingeschränkt	Vollumfänglich	
			Bewertung der Generalisierbarkeit von Optimierung- und Weiterentwicklungsmöglichkeiten	Für einfache Kontexte	Eingeschränkt	Vollumfänglich	
		Evolution unter Berücksichtigung der Einbindbarkeit in ein Kollektiv	Collaborative Responsibility		Eingeschränkt	Vollumfänglich	
	Collective Consciousness	Zusätzlich					
Einbeziehung von Responsibility in der Generierung von verteilten Plänen		Formalisierung von ethischen, moralischen, gesellschaftlichen Prinzipien				Vollumfänglich	
Mechanismen zum Abgleich von ethischen, moralischen, gesellschaftlichen Prinzipien						Vollumfänglich	

Dimension	Ausbaustufen	Major Challenge	Kerntechnologie	Kurz 2027	Mittel 2035	Lang 2045ff
Humans and Systems Die Fähigkeit des Systems auf verschiedenen Ebenen mit anderen Systemen oder Menschen zu kooperieren	Shared Situational Awareness	Ontologie der relevanten Kontext-Artefakt	Zusammenführung von domänenspezifischen Ontologien bei heterogenen Kollektiven	Vollumfänglich		
		Secure real-time communication	5G basierte Kommunikationsprotokolle	Vollumfänglich		
		Meta-Modell der Repräsentation des digitalen Welt-Bildes einschließlich Konfidenzen der identifizierten Objekte (incl. deren Lage, Pose, ...)	Setzt voraus, dass Perzeptionen mit Konfidenz annotiert sind; darüber hinaus: Standardisierungsprozess	Vollumfänglich		
		Dialogfähigkeit der Systemwahrnehmung in einer für den Kenntnisstand und Arbeitssituation des Menschen geeigneten Repräsentation	Kognitive Modelle	Für eingeschränkte Objektklassen	Vollumfänglich	
			HCI, VR, Augmented Reality,...; Human-Brain-Interface	Für eingeschränkte Objektklassen	Vollumfänglich	
	Shared Actions	Synchronisationsmechanismen für kooperative Aktionen		Vollumfänglich		
		Wechselseitige Kenntnis der Capabilities und Skills	Ontologie von Capabilities und Skills	Vollumfänglich		
		Distributed Agreement Protocols zur Auswahl gemeinsamer Handlungen		Für reine Technik-Kommunikation	Für Mensch-Maschine-Kooperation	
	Shared Goals	Formalisierbare Beschreibung der für die Kooperation notwendigen Aspekte des mentalen Modells des Menschen bzw. des technischen Systems		Für reine Technik-Kommunikation	Für Mensch-Maschine-Kooperation	
		Verhandlungsstrategien zur Abstimmung gemeinsam verfolgter Ziele		Für reine Technik-Kommunikation	Für Mensch-Maschine-Kooperation	
		Kommunikations- und Repräsentationsmechanismen für Ziele	Incentives	Für reine Technik-Kommunikation	Für Mensch-Maschine-Kooperation	
		Fähigkeit, Ziele zu relaxieren		Für reine Technik-Kommunikation	Für Mensch-Maschine-Kooperation	
			Konfliktresolutionsstrategie	Für reine Technik-Kommunikation	Für Mensch-Maschine-Kooperation	
	Shared Planning	Automatische Generierung von Vorschlägen für kooperative Pläne zur Erreichung der Ziele des Egosystems	Ontologie von Handlungen	Für reine Technik-Kommunikation	Für Mensch-Maschine-Kooperation	
			Verteilte Syntheseverfahren für Pläne unter unvollständiger Information	Für reine Technik-Kommunikation	Für Mensch-Maschine-Kooperation	

5 Verantwortlichkeit und Reflexion

5.1 Zusammenfassung



Dieser Abschnitt der Roadmap stellt die Fragen: wofür können Maschinen verantwortlich sein? Was dürfen oder können Maschinen entscheiden? Werden Maschinen in einer zukünftigen „Human Machine Society“ Partner von Menschen sein oder gar über sie entscheiden? Damit wird implizit auch die Frage gestellt: Wieviel Autonomie wollen wir Maschinen zugestehen?

Das zunehmende Maß an Autonomie von Systemen erfordert die Festlegung eines Konzepts der Systemverantwortung: Wenn nicht mehr der Mensch die endgültige Kontrolle hat, wer übernimmt dann die Verantwortung für die Aktionen eines Systems? Im Allgemeinen beschreibt Verantwortung eine Beziehung zwischen einem Subjekt (in unserem Kontext ein autonomes System) und einem Objekt, die von den Handlungen des Subjekts beeinflusst wird, und bezeichnet die Fähigkeit des Subjekts, seine Handlungen so zu bewerten und auszuwählen, dass sie konform sind zu bestehenden juristischen, ethischen und moralischen Normen. Normative Fähigkeiten sind Voraussetzung für Verantwortung. Sie bezeichnen die Fähigkeit des Systems über die Einhaltung seiner eigenen Handlungen, über andere Systeme, über Normen, ethische Grundsätze, ökologische Auswirkungen und gesellschaftliche Auswirkungen nachzudenken und erfordern insbesondere die digitale Darstellung solcher Normen verinnerlicht im System. Beispiele für solche Fähigkeiten sind die Bereitstellung maschineninterpretierbarer Versionen von Verkehrsgesetzen in hochautonomen Autos (vgl. auch die Diskussion über Dilemmasituationen¹ sowie ethische Richtlinien für den Aufbau autonomer Systeme²). Ein System handelt verantwortungsbewusst, wenn es immer zu rechtfertigen vermag, dass sein Handeln den anerkannten Normen entspricht. Solche Rechtfertigungen für alle Arten von Systemfähigkeiten müssen auf allen Ebenen der Systemhierarchie erstellt werden.

¹ Automatisiertes und Vernetztes Fahren. BMVI. Ethik-Kommission (Hrsg.). 2017

² siehe u. a. folgende Initiative: IEEE SA Standards Association. The IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems

Wir schlagen vor, in dieser Komplexitätsdimension die Ausbaustufen nach der zunehmenden Komplexität der Aufgaben zu ordnen, für die in diesem Sinne Verantwortung übernommen wird.

Ausbaustufen Verantwortung

1. Funktionale Verantwortung: Das System besitzt die Fähigkeit, die Verantwortung für die in der Umsetzung einer Funktion durchgeführten Handlungen zu übernehmen.
2. Missionsverantwortung: Das System besitzt die Fähigkeit, die Verantwortung für die in der Umsetzung einer Mission durchgeführten Handlungen zu übernehmen.
3. Mit-Verantwortung: Das System besitzt die Fähigkeit, die Verantwortung für die gemeinsam vorgenommenen Handlungen zu übernehmen.
4. Autopoietische Verantwortung: Das System kann sich nicht nur einen unbekanntem Kontext selbst erschließen und darin alleine oder kooperativ handeln, es besitzt darüber hinaus die Fähigkeit, für diese Handlungen Verantwortung zu übernehmen.

5.2 Detaillierte Beschreibung

5.2.1 Analyse der Veränderungen von Verantwortung

Entscheidungen können verschiedene Tragweite haben, deshalb unterscheiden wir

- **Funktionale Entscheidungen** (einzelne zeitlich begrenzte Aktionen: wie z. B. automatische Notbremsung). Die zugehörige Art der Verantwortung ist die funktionale Verantwortung.
- **Manöver- oder Missionsentscheidungen** als Reaktion auf komplexe Situationen mit verschiedenen Alternativen (wie z. B. „Kollisionsvermeidung durch Spurwechsel“ oder „Suche dir einen Parkplatz in der Nähe“, jeweils als Kette von Einzelfunktionen). Hier werden Ketten von Aktionen nach einem Plan abgearbeitet. Der Plan kann sich während der Aktionskette ändern. Entsprechend ist die hierzu relevante Ausbaustufe die Missions-Verantwortung.
- **Kollaborationsentscheidungen** bei denen Ziele und Nutzen im Rahmen einer gemeinsamen Mission gegeneinander abgewogen werden. (z.B. Platooning). Die entsprechende Ausbaustufe ist die Mit-Verantwortung.
- **Ethische Entscheidungen** bei denen durch daraus abgeleiteten Handlungen Einfluss auf Menschen und Gesellschaft genommen wird. Dazu zählen auch politische Entscheidungen, oder Konflikte mit, zwischen und an Menschen. (z.B. medizinische Apparaturen). Diese Art der Entscheidungen entspricht der höchsten Ausbaustufe der Dimension Verantwortung, der autopoietischen Verantwortung.

Eine andere Gruppe von Entscheidungen entsteht, wenn Systeme selbst entscheiden wie sie sich weiterentwickeln und was sie lernen und damit ihre Rolle als Subjekt in Beziehung zur Welt selbstständig ändern. Dies beeinflusst potenziell alle obigen Klassen von Entscheidungen. Diese induzieren die Frage: Welche Rollenänderung wollen wir zulassen und wie stellen wir sicher, dass diese im Einklang mit den in unserem Kulturraum geltenden Normen sind?

5.2.2 Capabilities

Um Verantwortung wahrnehmen zu können muss ein System eine Selbstwahrnehmung besitzen und wissen, dass es eine Fremdwahrnehmung in anderen Systemen gibt, die das Egosystem betrifft und die durch Handlungen des Egosystems beeinflusst wird. Dazu bedarf es einer Reihe von reflexiven Fähigkeiten.

- **Verantwortungsrelevante Wahrnehmung** ist die Fähigkeit, den Kontext und den eigenen Zustand beobachten und intern repräsentieren zu können.
- **Verantwortungsrelevante Analyse** ist die Fähigkeit, ein internes Lagebild und ein Lagebild im Kontext bezogen auf die normativen Regeln und ihre Anwendung auf die Objekte im Kontext bewerten zu können. Diese Bewertung findet mit verschiedenen zeitlichen Handlungsskalen, wie Aktion, Manöver, Mission, statt. Die Bewertung der eigenen Integrität schließt das Erkennen der Selbstbeschränkungen durch Fehler oder Ressourcenbegrenzungen ein. Die Bewertung der Verantwortungslage bezieht sich sowohl auf die aktuelle als auch solche hypothetischen Situationen, die eine Handlung des Systems unter Antizipation des Verhaltens des Kontexts erzeugt.
- **Verantwortungsrelevante Hypothesenbildung** ist die Fähigkeit, eine zukünftige, normativ akzeptierte Lage als Ziel zu spezifizieren, die das System durch seine Handlungen in der Wechselwirkung mit dem Kontext herstellen will und diese ggf. anzupassen. Dazu ist es notwendig, dass Verhalten der Objekte im Kontext und den eigenen Zustand zu antizipieren.
- **Verantwortungsrelevante Planung** ist die Fähigkeit, durch eine Kette von Aktionen eine normativ akzeptierte Ziel-Lage herzustellen. Das beinhaltet die Fähigkeit, dazu auch ggf. Pläne zu verändern und anzupassen.
- **Verantwortungsrelevante Experimente durchführen** ist die Fähigkeit, den Plan auf Basis der aktuellen Lage und der Ziel-Lage virtuell so durchzuführen, dass die Ziel-Lage durch eine verantwortungsrelevante Analyse bewertet werden kann.
- **Verantwortungsrelevante Nachvollziehbarkeit** ist die Fähigkeit, das Handeln des Systems vor dem Hintergrund der normativen Kriterien transparent, nachvollziehbar und erklärbar zu machen. Dazu gehört auch verworfene Handlungsoptionen begründen zu können und der Welt nachvollziehbare Informationen bereitzustellen.

Voraussetzung für all diese Fähigkeiten ist eine geeignete Modellbildung, welche geeignete Repräsentationen von Normen, Werte-Hierarchien, Verhaltensmustern und Kriterien zur Analyse und Bewertung des Erreichten unter normativen Gesichtspunkten erlaubt.

5.2.3 Herausforderungen und Forschungsfragen

Problemanalyse – Systemanalyse

Um Verantwortung wahrnehmen zu können, müssen einige Voraussetzungen gegeben und im System abgebildet sein.

- **Vertrauen:** Welchen Interessen folgt die Entscheidung? Wie wird Vertrauen in einem technischen System repräsentiert?
- **Ethik:** Welchem Kanon von Werten folgt die Entscheidung?
- **Privatheit:** Welche Information darf für Entscheidungen genutzt werden?
- **Transparenz:** Wie nachvollziehbar sind Entscheidungen? Dies beinhaltet unter anderem auch die Eigenschaften „Reproduzierbarkeit“ von Entscheidungen und „Begründbarkeit“ von Entscheidungen.
- **Kontrolle:** Wie können Entscheidungen kontrolliert, bzw. beeinflusst werden?
- **Geschäftsfähigkeit:** Wie weisen Maschinen ihre Geschäftsfähigkeit nach, vor allem wenn sie autonom etwas dazu lernen?
- **Haftung:** Wie können Maschinen im Schadensfall haftbar gemacht werden? (z. B.: Ausgleich

von durch die Maschine verursachten Schäden durch eine für die Maschine abgeschlossene Versicherung.)

- **Würde:** Haben Maschinen eine Art Würde und wie unterscheidet sich diese von der menschlichen Würde? Wie wird Würde in einem technischen System repräsentiert?
- **Demokratie:** Wie werden Interessenkonflikte ausgetragen und umgesetzt?

Systementwurf und Verifikation

Welche Mechanismen müssen in Systemen realisiert sein, um die verschiedenen Klassen von Entscheidungen treffen zu können mit den Eigenschaften, die für verantwortungsvolle Entscheidungen notwendig sind?

Wir brauchen analog zu dem Thema Sicherheit „by Design“, dem wir heute folgen,

- Ansätze, die die Eigenschaften „By Design“ sicherstellen. Das heißt, z. B. Demokratie by design. Dazu brauchen wir entsprechende
 - Beschreibungssprachen zur Spezifikation der Eigenschaften
 - Ontologien
- Architekturen (interoperabel, kompositional, self-aware)
 - Da die Aspekte von Land zu Land oder Kultur unterschiedlich sind und sich über der Zeit auch weiterentwickeln, müssen diese die notwendige Varianz abbilden und evolutionierbar sein.

Damit Systeme bzgl Konformität zu Normen verifiziert werden können, müssen die Ansätze formal sein. Dies erlaubt eine automatisierbare Verifikation, die angesichts der Komplexität und der Dynamik der Veränderung der Systeme alternativlos ist. Die Verifikation kann in einem solch dynamischen Umfeld nicht mehr statisch sein, sondern muss on-line im System oder in einer virtuellen Umgebung parallel zur Systemevolution erfolgen.

Validierung

Da Systeme sich immer schneller und individueller ausprägen werden, kann eine Validierung nicht mehr durch eine Zentralinstanz oder Behörde erfolgen, sondern es muss eine automatisierbare „Meta Freigabe“ erfolgen, die die Freigabe-Argumentationskette, die Systeme, ihre Eigenschaften, Nachweise und Freigabevoraussetzungen formal beschreibt, in der Art wie heute juristische Claims beschrieben sind.

Forschungsfragen

- Modelle, Theorien und Ansätze zur Beschreibung der Systemeigenschaften wie Vertrauen, Gesundheit, Ethik, Würde, Kultur, Demokratie, Haftung, Kontrolle u.a.m.
- Kompositionelle, interoperable und self-aware Architekturen
- Konflikt-Lösungsmechanismen in Maschinen
- Konzepte von sozial verantwortlichen Maschinen
- Konzepte für verantwortliches Delegieren zwischen Maschinen
- Konzepte für verantwortliches Lernen
- Konzepte für die Teilung von Verantwortung zwischen Mensch und Maschine

5.2.4 Benötigte Technologien und Zeithorizont

Die folgende Tabelle gibt für jede Ausbaustufe dieser Komplexitätsdimension die größten Herausforderungen sowie die wichtigsten benötigten Technologien zu ihrer Realisierung an und ordnet diese zeitlich ein. Diese Analyse ist die Grundlage für die in Kapitel 7 dargestellte Roadmap. (Tabelle siehe folgende Seite.)

Dimension	Ausbaustufen	Major Challenge	Kerntechnologie	Kurz 2027	Mittel 2035	Lang 2045ff	
Responsibility and Reflection Die Fähigkeit über die Außenwirkung des System zu Reflektieren und Verantwortung für die Auswirkungen des Handelns zu übernehmen.	Manöver - Responsibility	Modell des Kontexts (inkl. operationalem Kontext mit Physik, andere Systeme, Menschen; darüber hinaus auch normatives Verhalten, ethische Prinzipien gesetzliche Regelungen)	Modellierung von normativen Verhaltens, ethischer Prinzipien und gesetzlicher Regelungen	Nur operationaler Kontext (siehe Kontext:Humans), nur jeweils eine Domäne	Nur operationaler Kontext, mehrere Domänen	Unter Einbeziehung von ethischen Prinzipien	
			Prinzipien zur Integration von Kontext-Modellen aus unterschiedlichen Domänen	Nur operationaler Kontext, nur jeweils eine Domäne	Nur operationaler Kontext, mehrere Domänen	Unter Einbeziehung von ethischen Prinzipien	
		Prädiktion, Bewertung und Priorisierung der Auswirkungen des Handelns des Egosystems auf den Kontext		Nur operationaler Kontext, nur jeweils eine Domäne	Nur operationaler Kontext, mehrere Domänen	Unter Einbeziehung von ethischen Prinzipien	
		Führung des Egosystems unter Vermeidung von negativen Auswirkungen auf den Kontext	Bestimmung eines Handlungsraums für das Egosystem, so dass das Manöver unter Respektierung von hohen Kontext-Anforderungen	Nur operationaler Kontext, nur jeweils eine Domäne	Nur operationaler Kontext, mehrere Domänen	Unter Einbeziehung von ethischen Prinzipien	
			Architekturkonzepte für Responsibility	Nur operationaler Kontext, nur jeweils eine Domäne	Nur operationaler Kontext, mehrere Domänen	Unter Einbeziehung von ethischen Prinzipien	
	Mission - Responsibility	Modell des Kontexts (incl. operationalem Kontext mit Physik, andere Systeme, Menschen; darüber hinaus auch normatives Verhalten, ethische Prinzipien gesetzliche Regelungen)	Modellierung von normativen Verhaltens, ethischer Prinzipien und gesetzlicher Regelungen	Nur operationaler Kontext (siehe Kontext Humans), nur jeweils eine Domäne	Nur operationaler Kontext, mehrere Domänen	Unter Einbeziehung von ethischen Prinzipien	
			Prinzipien zur Integration von Kontext-Modellen aus unterschiedlichen Domänen	Nur operationaler Kontext, nur jeweils eine Domäne	Nur operationaler Kontext, mehrere Domänen	Unter Einbeziehung von ethischen Prinzipien	
		Prädiktion, Bewertung und Priorisierung der Auswirkungen des Handelns des Egosystems auf den Kontext		Nur operationaler Kontext, nur jeweils eine Domäne	Nur operationaler Kontext, mehrere Domänen	Unter Einbeziehung von ethischen Prinzipien	
		Führung des Egosystems unter Vermeidung von negativen Auswirkungen auf den Kontext	Bestimmung eines Handlungsraums für das Egosystem, so dass das Manöver unter Respektierung von hohen Kontext-Anforderungen	Nur operationaler Kontext, nur jeweils eine Domäne	Nur operationaler Kontext, mehrere Domänen	Unter Einbeziehung von ethischen Prinzipien	
	Collaborative - Responsibility	Zusätzlich:					
		Verteilte Abstimmung über möglichen verantwortbaren Handlungsraum	Distributed Real-Time Agreement Protocols	Nur operationaler Kontext, nur jeweils eine Domäne	Nur operationaler Kontext, mehrere Domänen	Unter Einbeziehung von ethischen Prinzipien	
		Verteilte Konfliktlösungsstrategien (bei zu restriktiven Handlungsräumen) s. oben		Nur operationaler Kontext, nur jeweils eine Domäne	Nur operationaler Kontext, mehrere Domänen	Unter Einbeziehung von ethischen Prinzipien	
		Kooperative Führung (vgl. Dimension Kooperation)		Nur operationaler Kontext, nur jeweils eine Domäne	Nur operationaler Kontext, mehrere Domänen	Unter Einbeziehung von ethischen Prinzipien	
	Autopoietische - Responsibility		wie oben, aber für autopoietische Systeme			Nicht in den nächsten 50 Jahren	

6 Integrität und Zertifizierung

6.1 Zusammenfassung



Um das übergeordnete Ziel der Sicherung der Qualität trotz wachsender Komplexität zu erreichen, sind Maßnahmen zur Sicherung der Integrität und der nachprüfbaren Absicherung der Umsetzung der Qualitätsvorgaben zu ergreifen. Die Komplexitätsdimension der Systemintegrität widmet sich der Herausforderung, durch Konstruktion sicherzustellen, dass alle Quellen und Akteure für die Entscheidungsfindung konsistent und vertrauenswürdig sind. Die Gewährleistung der Systemintegrität geht über die Fähigkeit zur Selbstbeobachtung und -reflexion zur Erkennung von Eingriffen hinaus, indem die Fähigkeit zur Wiederherstellung der Systemintegrität nach Verstößen sichergestellt wird. Auch hier hängt die Komplexität solcher Selbstreparaturaktionen vom Grad der Autonomie ab, mit der dies durchgeführt werden kann, und von der Komplexität der Objekte und Elemente, deren Integrität für die Systemfähigkeiten kritisch ist. Dies geht auch über die klassischen Ansätze zur Gewährleistung von Information und/oder Integrität des Zustandes hinaus, da die zunehmende Stärke von Systemen mit zunehmenden Herausforderungen für die Gewährleistung von Integrität verbunden ist, um die Leistungsfähigkeit der Evolution und/oder der Lernfähigkeiten zu steigern. Beispielsweise darf die Systemevolution nicht die Integrität der Systemkonfiguration verletzen und es muss gezeigt werden, dass Systemstrategien, die durch unbeaufsichtigtes Lernen erhalten wurden, mindestens das gleiche Maß an Qualität und Sicherheit erreicht, wie vorhandene Ausweichstrategien, die mit der Systemverschlechterung fertig werden. Die folgenden Ausbaustufen zur Gewährleistung der Systemintegrität sind nötig:

1. Informationsintegrität: Die Fähigkeit des Systems, sicherzustellen, dass alle Informationen, die in Planung, Entscheidungsfindung und Handeln des Egosystems eingehen, nicht kompromittiert, sondern vertrauenswürdig sind, und die Fähigkeit zur Wiederherstellung haben.
2. Integrität des Zustandes: Zusätzlich: Die Fähigkeit, sicherzustellen, dass der Zustand des Egosystems konsistent und nicht gefährdet ist sowie automatisch von gefährdeten Zuständen in ungefährdete Zustände zu gelangen.
3. Planintegrität: Zusätzlich: Die Fähigkeit, sicherzustellen, dass die Pläne des Ich-Systems nicht gefährdet und konsistent sind sowie die Fähigkeit, die Integrität von Plänen wiederherzustellen.
4. Evolutionsintegrität: Zusätzlich: Die Fähigkeit des Systems, die Informationsintegrität, die Zustandsintegrität und die Planintegrität während der Evolution des Systems aufrechtzuerhalten (differenziert nach den Ausbaustufen für Evolution).
5. Lernintegrität: Zusätzlich: Die Fähigkeit des Systems zur Aufrechterhaltung der Informationsintegrität, der Integrität des Zustandes und der Planintegrität für zunehmend leistungsfähigere Lernebenen (differenziert nach den Ausbaustufen für Intelligenz).

Technischer Fortschritt und die daraus resultierenden wirtschaftlichen Potenziale wie auch die Potenziale zur Lösung gesellschaftlicher Probleme werden dazu führen, dass die Komplexität dieser Systeme in den Dimensionen Autonomie, Intelligenz, Evolution und Kooperation exponentiell wachsen werden und zunehmend komplexeren Kontexten ausgesetzt sind. Die entscheidenden Herausforderungen zur Sicherung von „Quality Made in Germany“ ist die hier betrachtete Dimension der Nachvollziehbarkeit der Qualitätssicherung solcher Systeme: Sicherheit (sowohl im Sinne von Safety wie auch Security) und Verfügbarkeit solcher Systeme sowie die Respektierung gegebener etablierter gesellschaftlicher, rechtlicher und ethischer Rahmenbedingungen sind in die Qualitätsbewertung einzubeziehen. Letzteres setzt normative Reflexionsfähigkeiten voraus, die in der Komplexitätsdimension zu Verantwortlichkeit dargestellt wurden.

Auf europäischer Ebene wird dies bereits gefordert in den ICT Standardisation Priorities for the Digital Single Market¹ sowie in verschiedenen Forschungsagenden, wie z. B. in der von drei europäischen Industrie-Verbänden herausgegebenen Strategic Research Agenda² der Joint Undertaking

¹ ICT Standardisation Priorities for the Digital Single Market. European Commission (Hrsg.), 2016

² AENEAS, ARTEMIS-IA. EPoSS (Hrsg.). Electronic Components and Systems Strategic Research Agenda (ECS SRA). Gairle (NL). 2019
 Seite 156: „Dependability and Trustability are fundamental components of any innovation in the digital economy. It is undeniable that novel products and services like personal healthcare monitoring, connected cars or smart homes bring strong benefits for the society, provided that dependability and trustability are taken care of. If this cannot be ensured, there is a significant risk that these innovations will not be accepted by society due to missing consumer confidence.“
 Seite 156: „Safety aspects have a major impact in case of public knowledge of accidents due to technical failure.“
 Seite 157: „The most important characteristics for businesses in the future will be the aspect that they are perceived as trusted companies. Only as trusted organisations, they can maintain a long-term relationship to their customers. New “trusted products” represent a great opportunity for European companies, for example with the development of a “Trusted IoT” label.“

ECSEL. Die VDA Leitinitiative³ koordiniert gemeinsame FuE-Aktivitäten der Automobilindustrie zur Qualitätssicherung hochautomatisierter Fahrzeuge. In Rahmen der Plattform Lernende Systeme werden gegenwärtig Risikostufen für KI-basierte Systeme entwickelt, um dem jeweiligen Risiko angemessene Maßnahmen zur Qualitätssicherung vorzuschlagen⁴. Die Auswertung von mehr als 80 durch Expertenkommissionen veröffentlichten Vorschläge zur verantwortlichen Einführung von künstlicher Intelligenz empfehlen den Nachweis von Transparenz von KI-Systemen, deren Entscheidungen oder Bewertungen signifikante Auswirkungen auf Privatsphäre, Fairness, Gerechtigkeit oder Sicherheit haben⁵.

Vor diesem Hintergrund schlagen wir Ausbaustufen in den Fähigkeiten zur Absicherung vor, die je nach Risikoklasse des Egosystems angestrebt werden sollen. Die Festlegung solcher Risikoklassen und daraus pro Risikoklasse abgeleiteten Maßnahmen sind nicht Gegenstand dieser Roadmap. Hierzu sind weitere Diskussionen sowohl in der Plattform Lernende Systeme wie auch in den relevanten Industrieverbänden erforderlich, welche letztlich zu entsprechenden Standards führen. Der Begriff der Zertifizierung wird hier für den Nachweis einer erfolgreichen Qualitätsabsicherung gegenüber solchen Richtlinien und gegebenenfalls Standards verwendet.

1. Off-line Zertifizierung: Die Zertifizierung erfolgt außerhalb des Egosystems gemäß den Regeln der zuständigen regulativen Behörde.
2. On-line Zertifizierung: Die Zertifizierung erfolgt im Egosystem selbst. Das beinhaltet, dass die Maßnahmen, die im Egosystem zur Selbstzertifizierung verwendet werden, durch die zuständige regulative Behörde (off-line) zertifiziert wurden.
3. On-line Zertifizierung und Handlungsbegründung: Zusätzlich: Das Egosystem ist in der Lage, alle qualitätsrelevanten Entscheidungen und Handlungen zu begründen und diese Begründungen in die on-line Zertifizierung einbeziehen zu können.
4. Zertifizierung und Begründung der Einhaltung der Regeln: Zusätzlich: Das Egosystem kann die Auswirkungen seines Handelns auf vorgegebene normative Regelwerke reflektieren und deren Einhaltung on-line zertifizieren. Das beinhaltet, dass die Maßnahmen, die im Egosystem zur Selbstzertifizierung verwendet werden, durch die zuständige regulative Behörde (off-line) zertifiziert wurden.
5. Zertifizierung und Begründung der Einhaltung ethischer Grundsätze: Zusätzlich: Das Egosystem kann die Auswirkungen seines Handelns und die Einhaltung gesellschaftlich geforderter ethischer Prinzipien reflektieren und deren Einhaltung on-line zertifizieren. Das beinhaltet, dass die Maßnahmen, die im Egosystem zur Selbstzertifizierung verwendet werden, durch die zuständige regulative Behörde (off-line) zertifiziert wurden.

³ Siehe etwa:

- Projekt PEGASUS. <https://www.pegasusprojekt.de/> (Zugriffsdatum: 28.10.2019)
- Projekt VVMMethoden - Verifikations- und Validierungsmethoden automatisierter Fahrzeuge Level 4 und 5. <http://www.tuvpt.de/index.php?id=vvmethoden> (Zugriffsdatum: 28.10.2019)
- Aktionsplan Forschung für autonomes Fahren. Ein übergreifender Forschungsrahmen von BMBF, BMWi und BMVI. BMBF, BMWi, BMVI (Hrsg.). 2019

⁴ Als Vorläufer hierzu dient das Positionspapier zur Bewertung von Risiken in der Verwendung von KI in medizinischen Anwendungen: Sichere KI-Systeme in der Medizin - Positionspapier. Plattform Lernende Systeme (Hrsg.). Voraussichtlich 2019

⁵ Ethische Grundsätze wie bereits in der Einleitung zitiert.

6.2 Detaillierte Beschreibung: Integrität

6.2.1 Analyse

Integrität stellt die Frage, ob ein System in einem Zustand ist, in dem es die intendierte Funktion mit der geforderten Qualität (wie Safety, Reaktivität, ...) verantwortlich erfüllen kann. Kann ein System seinen Zustand korrekt erkennen? Kann es sich im Falle von Fehlern soweit wieder heilen, dass die intendierte Funktion erbracht werden kann oder zumindest einen akzeptablen Zustand herbeiführen kann? Diese Fähigkeit mit Störungen aller Art umzugehen wird auch als Resilienz bezeichnet. In diesem Abschnitt wird die Frage nach dem Gesundheitszustand vor dem Hintergrund der Veränderungen, die in den vorangegangenen Kapiteln antizipiert wurden, gestellt. Sie geht damit in ihrer Reichweite weit über bisher bekannte Konzepte von Self-Awareness und Self-Healing hinaus.

Die Analyse der Integrität von Verhalten beleuchtet die Problemstellung, die sich aus der Integritätsfrage ergibt. Die Frage ist: Hat das System die intendierten Eigenschaften (Funktion, Sicherheit, Ressourcenverbrauch u.a.m.) auch nach langer Laufzeit und auch in ggfs. unvorhergesehenen Kontexten immer noch? Können wir diese Eigenschaften zur Laufzeit beobachten und ggf. im System Maßnahmen einleiten, um wieder einen integren Zustand herzustellen? Besondere Aspekte dabei sind:

- **Berücksichtigung von dynamischem Verhalten:** Autonome (lernende) Systeme sind in der Lage, sich auch in Situationen, für die keine Spezifikation verfügbar ist, sinnvoll zu verhalten. Aufgrund der open-world Eigenschaft ist die Spezifikation eines solchen Systems immer unvollständig, und autonome Systeme müssen diese Eigenschaft somit notwendig beinhalten. Somit kann die Korrektheit eines Verhaltens nur im Kontext der aktuellen (spezifizierten oder nicht spezifizierten) Situation beurteilt werden. Dies ist durch das System sicherzustellen.
- **Integrität zur Laufzeit:** Aufgrund der unvollständigen Spezifikation sowie aufgrund ihrer Komplexität, können autonome Systeme nicht vollständig zur Design-Zeit validiert und getestet werden. Solche Systeme können unvorhergesehene Situationen erleben und sie können sich durch Evolution (Lernen) in durch den Designer nicht antizipierte Richtungen weiterentwickeln. Somit muss für solche Systeme eine Integritätsprüfung zur Laufzeit möglich sein, um Resilienz zu erreichen.
- **Integrität des Lernens:** Schon auf geringeren Autonomie-Stufen (wie beispielsweise in aktuellen ADAS Systemen) werden Verfahren des maschinellen Lernens - u.a. Deep Learning - verwendet um bestimmte Funktionen zu realisieren, die mit traditionellen Algorithmen nicht oder nur sehr aufwändig zu realisieren wären (z. B. semantische Segmentierung von Kamerabildern). Aktuelle Verfahren zur Sicherstellung der Integrität sowie zu Sicherheitsnachweisen sind kaum oder gar nicht in der Lage, solche auf Neuronalen Netzen basierende Verfahren einzubeziehen. Somit entsteht hier ein großer Forschungsbedarf.

Integrität und Struktur- und Aktionsdynamik

... bedeutet: das System hat einen Zustand, der ihm erlaubt die intendierten Systemfunktionen und -eigenschaften in seinem operationalen Kontext zu erbringen. Dabei gehen wir davon aus, dass dieses Set von Eigenschaften definiert und beobachtbar ist und im Beobachtungsintervall stabil ist. Ist der integre Zustand verletzt, erkennt das System dies und kann geeignete Maßnahmen einleiten, um den maximalen „Quality of Service“ wiederherzustellen. Dabei stellen sich folgende Fragen:

- **Strukturdynamik**
 - Wie definieren wir strukturelle Integrität, wann ist ein System im definierten Sinn integer?
 - Wie erkennt das System seine neuen Systemgrenzen und seinen neuen Kontext?
 - Wie verändert sich die Definition von struktureller Integrität in der neuen Systemverbundkonfiguration?
- **Aktionsdynamik**
 - Wie beschreiben wir die veränderte Aktionsdynamik durch Feature- und Strukturdynamik?
 - Wie erfassen wir emergentes Verhalten und beurteilen es?
 - Wie beobachtet ein System (-Verbund) seine jeweils gültigen Systemeigenschaften?
 - Wie werden Strukturübergänge gestaltet und orchestriert?
 - Wie heilen sich nicht integre Systeme?
- **Qualitative Fragen (sowohl Struktur- als auch Aktionsdynamik)**
 - Wie stellen wir sicher, dass ein neues System in Systemverbund diesen nicht korrumpiert?
 - Wie beobachten wir die Systemeigenschaften zur Laufzeit?
 - Wie weisen wir Eigenschaften wie Sicherheit oder Vertrauenswürdigkeit in dynamischen Systemverbänden nach?

Integrität und Evolution

... bedeutet, das System hat Funktions- und Eigenschaftendynamik, bei der sich die Eigenschaften des Systems zur Laufzeit verändern, insbesondere erweitern. Dabei stellen sich folgende Fragen:

- Passen die Veränderungen zum Rest des Systems (Kompatibilität)?
- Ist die neue Systemkonfiguration integer (im Sinne der oben beschriebenen Systemintegrität)?
- Sind die Beobachtungsfunktionen für die neue Konfiguration integer?
- Sind die Heilungsmaßnahmen zur Wiederherstellung der Integrität für die neue Konfiguration integer?

Integrität und Intelligenz

Intelligenz bedeutet, dass das System während seiner Lebenszeit lernt. Lernen bedeutet das Wissen über den Kontext erweitert sich und sowohl Ziele als auch Handlungsmuster des Systems passen sich immer differenzierter an den erweiterten Kontext an. Dadurch sollen zusätzliche Optionen für das intendierte Verhalten ermöglicht werden.

Integrität bedeutet bezogen auf das Gelernte:

- Ist der Wissensraum integer erweitert bezogen auf das was schon an Wissen da ist?
- Ist der Verhaltensraum integer erweitert?
- Ist das Erlernte intendierte Verhalten spezifiziert und damit beobachtbar?
- Ist die Auswirkung der Erweiterungen auf die Systemeigenschaften integer?
- Sind die Beobachtungsfunktionen für die Integrität des neuen Systems erweitert?
- Sind die Heilungsmechanismen für die Wiederherstellung der Integrität für das neue System integer?

Bezogen auf die Integrität des Lernvorgangs selbst:

- Integrität der neu formierten Hypothesen mit dem existierenden System
- Integrität des „Design of Experiment“ zur Stimulierung des Objekts (Kontext)
- Integrität des Beobachtungssystems
- Integrität der Interpretation des Beobachteten
- Integrität der Weiterentwicklung der Hypothese

Integrität und Autonomie

Autonomie bedeutet: das System kann in Kontexten verschiedener Komplexität ohne Interaktion von außen angemessen agieren. Wir diskutieren für die vorher eingeführten Stufen von Autonomie die Auswirkung auf Integrität

- **Funktionale Autonomie:** Autonomes Handeln wird hier in einem geschlossenen Kontext durchgeführt, der physikalisch formal durch einen Algorithmus beschrieben werden kann, z.B. Automatisches Einparken. Integrität kann von einem vordefinierten situativ parametrisierten Modell beobachtet und ggf. geheilt werden.
- **Missionsorientierte Autonomie** erfordert eine Analyse des Kontextes einer Planung und ggf. eine Anpassung des Plans bei auftretenden Schwierigkeiten. z.B. suche einen Parkplatz in der Nähe. Integrität bedeutet Integrität des Kontextwissens in Bezug auf Vollständigkeit, Widerspruchsfreiheit und Korrektheit zu prüfen, so dass ein angemessener Plan abgearbeitet werden kann und die Annahmen des Plans beobachtet werden können. Heilung bedeutet, dass wenn die Annahmen des Plans nicht eintreffen, aus der neuen Situation heraus einen neuen Plan zu entwickeln und dabei nicht in Endlosschleifen zu geraten.
- **Kooperative Autonomie** erfordert alle Eigenschaften, die bereits unter dem Abschnitt Kooperation bzgl. Integrität genannt worden sind. Aus der Autonomieeigenschaft ergeben sich keine spezifischen Integritätsbedingungen, weil im Abschnitt Integrität von Kooperation Autonomie schon vorausgesetzt wurde.
- **Autopoietische⁶ Autonomie** bedeutet nach Maturana U., dass ein System seine relevante Umwelt durch Experimente, eigene Beobachtung und angemessene interne Repräsentation des Modells der Umwelt in sich selbst erzeugt und auf dieser Basis in der Welt handelt. Ein Integritätsbegriff ist dafür noch nicht bekannt. Aus jetziger Sicht ist er mit der Integrität der Intelligenz bezogen auf den Lernvorgang selbst verwandt; dies soll hier als Arbeitshypothese genügen.

Integrität im Kontext

... bedeutet: das System nimmt den relevanten Kontext für die intendierte Systemfunktion wahr und wechselwirkt in zielorientierter Weise mit dem Kontext. Das System ist integer, wenn beide Bedingungen erfüllt sind.

- **Physikalisch/logisch geschlossene Kontexte** sind durch die Wechselwirkungen mit den Kontext-Elementen, die wir beobachten und die wir beeinflussen können, vollständig beschrieben

⁶ Maturana, H.R./Varela, F.J. The cognitive process. Autopoiesis and cognition: The realization of the living. Springer Science & Business Media. 1980. Seite 3: "An autopoietic machine is a machine organized (defined as a unity) as a network of processes of production (transformation and destruction) of components which: (i) through their interactions and transformations continuously regenerate and realize the network of processes (relations) that produced them; and (ii) constitute it (the machine) as a concrete unity in space in which they (the components) exist by specifying the topological domain of its realization as such a network."

und können deshalb in ihrem Verhalten vorhergesagt werden. Wenn die Vorhersage nicht mit dem tatsächlich beobachteten Verhalten übereinstimmt, ist die Integrität verletzt, die Ursache dafür kann analysiert werden und ggf. Maßnahmen zur Heilung eingeleitet werden.

- **Offener bekannter Kontext** bedeutet, alle Klassen von Objekten und Relationen zwischen den Objekten mit ihren Eigenschaften (auch dynamische) sind im Prinzip bekannt (wie z.B. Verkehrssituationen) allerdings durch die Menge der Kombinationen nicht vollständig spezifizierbar. Neue Elemente können einer Klasse zugeordnet werden, auch wenn sie in ihrer Bedeutung noch keinem Handlungsmuster zugeordnet werden können.
 - Können wir alle Objekte mit genügender Sicherheit beobachten und klassifizieren?
 - Können wir alle unbekannt Objekte von den bekannten mit genügender Sicherheit erkennen und klassifizieren und daraus einen Lernvorgang (auch off-line) anstoßen?
 - Können wir Objekte situativ richtig miteinander in Beziehung setzen, um daraus eine Situation zu charakterisieren?
 - Können wir unbekannt Situationen (Relationsmuster) von bekannten (beherrschten) mit genügend Sicherheit unterscheiden und daraus einen Lernvorgang (auch off-line) anstoßen?
 - Können wir das Systemverhalten in einer Situation beobachten und auf Angemessenheit hin bewerten? Ggf. einen Lernvorgang anstoßen
- **Offene unbekannte Kontexte** bedeutet, wir bewegen uns in Gelände mit unbekannt Klassen von Objekten und Beziehungen zwischen den Objekten, was zu unbekannt Situationen führt. Integrität kann im Unbekannt nicht definiert werden. Wichtig ist, dass ein System erkennt, dass es sich in unbekanntem Gelände befindet und in einen Notfallmodus oder Explorationsmodus schaltet, in dem es minimalen Schaden im Kontext und für sich selbst anrichtet und maximalen Erkenntnisfortschritt macht, um im Sinne des im Abschnitt Kontext dargestellten Explorationszyklusses das Unbekannte so zu erforschen, dass es mit Abschluss des Lernzyklus hinreichend bekannt ist.

Integrität und Kollaboration

stellt sicher, dass von spontan kooperierenden Systemen, ähnlich wie bei Integrität im Verhalten, der Verbund von Systemen zu jedem Zeitpunkt einen Zustand hat, der ihm erlaubt, die intendierten Systemfunktionen und – Eigenschaften in seinem operationalen Kontext zu erbringen. Hier werden nicht nochmals die Aspekte der Integrität von Verhalten wiederholt, sondern es wird auf die Aspekte eingegangen, die sich aus der spontanen Kooperation ergeben und die dafür notwendigen Integritätsbedingungen. Dabei wird die Integrität von Kooperationsvoraussetzungen und Kooperationsarten betrachtet.

- **Gemeinsame Integritätsmodelle:** Integrität ist heute typisch eine Eigenschaft von Einzelsystemen. Wenn System-Kooperationen betrachtet werden, muss die Absicherung und das Management von Integrität auf den kooperierenden Systemverbund ausgeweitet werden.
- **IT-Sicherheit für Integrität:** Um zu kooperieren müssen Systeme sich öffnen; dadurch entstehen neue Möglichkeiten für Security-Attacken.
- **Generierung von Vertrauen:** Wenn der Integritätsbegriff über Systemgrenzen hinaus erweitert wird, müssen entsprechende Informationen zwischen den kooperierenden Systemen ausgetauscht werden. Diese Informationen müssen vertrauenswürdig sein, um zur Sicherstellung der Integrität beitragen zu können. Da Kooperationsbeziehungen dynamisch -- auch mit bisher unbekannt Systemen - entstehen können, müssen geeignete Maßnahmen zur Etablierung dieses Vertrauens vorhanden sein.
- **Integrität der Kooperationsvoraussetzungen**

- Kommunikation - Kanal und Basisprotokoll
- Gemeinsame Sprache - Ontologie, Syntax und Semantik
- Gemeinsames Situationsverständnis - haben die Partner dieselbe Sicht auf die Kooperationssituation, damit Entscheidungen über kooperative Aktionen valide sind
- Gemeinsames Set von Fähigkeiten - gibt es ein Set von gemeinsamen oder komplementären Fähigkeiten, damit überhaupt kooperiert werden kann
- Gemeinsames Kooperationsstrategie und -ziele - gibt es gemeinsame Handlungskonzepte und verhandelbare Kooperationsziele
- Gemeinsames Belohnungssystem – gibt es gemeinsame Incentives, damit sich Kooperation für die Partner lohnt.
- Integrität auf verschiedenen Kooperationsebenen
 - Kooperativer Austausch von Information - Integrität der ausgetauschten Information sowie Identifikation von Fake-Information, die den Verbund zu falschem Handeln bringt, um eigene Vorteile zu erzielen.
 - Kooperatives Handeln - Integrität der gemeinsamen Aktionen im Operativen Kontext z.B. ein Fahrzeug vereinbart, das andere einfädeln zu lassen, aber es verringert seine Geschwindigkeit nicht und ist damit auf Kollisionskurs.
 - Kooperatives Planen und Entscheiden - Integrität von vereinbarten Plänen und Entscheidungen zur Vereinbarungszeit und zur Durchführungszeit.
 - Kooperatives vereinbaren von Zielen – Integrität der Erreichbarkeit von vereinbarten Zielen.
 - Kooperatives Lernen – Integrität von gelerntem Wissen und Handlungsoptionen sowohl im Verbund als auch im Einzelsystem.

Integrität und Verantwortung

bedeutet, das System kann seine Entscheidungen auf einer legitimen Basis durchführen und rechtfertigen und ist damit vertrauenswürdig. D.h. Entscheidungen transparent, auf einer kollektiven akzeptierten Basis (Werte und Konventionen), und zum Wohle und zur Sicherheit aller zu treffen, ohne die Selbstbestimmung der Anderen in unangemessener Weise einzuschränken. Die Kernelemente der Integrität des Aspekts Verantwortung sind Vertrauen, Ethik, Würde, Privatheit, Transparenz, Kontrolle, Geschäftsfähigkeit, Haftung und Demokratie im Sinne von Gewaltenteilung (inklusive Transparenz)

Dabei stellen sich folgende Fragen:

- Integrität der **Entscheidungsbasis**, inklusive der Frage welche Information für die Entscheidung verwendet wurde.
- Integrität der **Entscheidungen**: Nachvollziehbarkeit der Argumentation für eine Entscheidung, inklusive der Frage, welche Werte für die Entscheidungsfindung ausschlaggebend waren und durch welche Interessen die Entscheidung geprägt ist.
- Integrität der **Sicherheitsbewertung von Entscheidungen?**
- Integrität der **Allgemeinwohlbewertung von Entscheidungen?** (Welche Interessen waren für die Entscheidung ausschlaggebend und wie wurden sie abgewogen.)
- Integrität der **Bewertung der Einschränkung** der Selbstbestimmung von anderen?
- Integrität der **Grenzen der Entscheidbarkeit** von Situationen. (Dilemma Situationen)
- Integrität der **Wahrnehmung der Situation**, in der entschieden wurde?
- Integrität über Entscheidungen von **Veränderungen der Entscheidungsbasis?**
- Integrität über die Entscheidungen, das System (Systemverbund) durch Evolution, Kooperation oder Lernen **zu verändern**.

6.2.2 Capabilities

Die im vergangenen Abschnitt dargestellten vielfältigen Integritätstests sowie die zugehörigen Fähigkeiten zur Wiederherstellung der Integrität im Falle von Verletzungen können mittels der folgenden Capabilities vereinfachend zusammengefasst werden. Diese sind jeweils aufeinander aufbauend, d.h. ein System, das eine bestimmte Capability besitzt, benötigt auch alle vorangehenden Fähigkeiten:

Informationsintegrität

Die Fähigkeit des Systems, sicherzustellen, dass alle Informationen, die in Planung, Entscheidungsfindung und Handeln des Egosystems eingehen, nicht kompromittiert, sondern vertrauenswürdig sind, und die Fähigkeit zur Wiederherstellung haben.

Integrität des Zustandes

Zusätzlich zur Informationsintegrität: Die Fähigkeit, sicherzustellen, dass der Zustand des Egosystems konsistent und nicht gefährdet ist, und automatisch von gefährdeten Zuständen in ungefährdete Zustände zu gelangen.

Planintegrität

Zusätzlich zur Integrität des Zustandes: Die Fähigkeit, sicherzustellen, dass die Pläne des Egosystems nicht gefährdet und konsistent sind sowie die Fähigkeit, die Integrität von Plänen wiederherzustellen.

Evolutionsintegrität

Zusätzlich zur Planintegrität: Die Fähigkeit des Systems, die Informationsintegrität, die Zustandsintegrität und die Planintegrität während der Evolution des Systems aufrechtzuerhalten (differenziert nach den Ausbaustufen für Evolution).

Lernintegrität

Zusätzlich zur Evolutionsintegrität: Die Fähigkeit des Systems zur Aufrechterhaltung der Informationsintegrität, der Integrität des Zustandes und der Planintegrität für zunehmend leistungsfähigere Lernebenen (differenziert nach den Ausbaustufen für Intelligenz).

6.2.3 Technische Herausforderungen und Forschungsfragen

Problemanalyse – Systemanalyse

Bedeutet, die Kernelemente der Integrität und ihre Eigenschaften in verteilten, strukturdynamischen, evolutionären, intelligenten autonomen Systemen, im offenen dynamischen Kontext zu benennen.

- **Beobachten**: Die Eigenschaften, aus denen sich die Integritätsbedingungen der verschiedenen Integritätsaspekte ergeben, zuverlässig zu beobachten. Dazu braucht das System „Self-Awareness“-Eigenschaften, die sich auf die Einzelsysteme und die Systemverbünde auf allen ihren Funktionsebenen beziehen.
- **Bewerten**: Aus dem komplexen Netzwerk von Integritätsindizes muss eine Lagebeurteilung des System of Systems abgeleitet werden, die ein angemessenes Handeln zur Wiederherstellung der Integrität erlaubt.

- **Handeln:** Durch welche Maßnahmen kann die „Heilung des Systems“ so durchgeführt werden, dass einerseits ein maximaler Quality of Service erhalten wird und andererseits die Handlungsoptionen für die Heilung weiterer Störungen maximal sind. Dazu sind Self-Development-Eigenschaften im Systemverbund notwendig.
- **Rechtfertigen:** Durch welche Maßnahmen und können Transparenz und Nachweise bereitgestellt werden, um in einem Streitfall eventuelle Konflikte zu lösen, bzw. im Rahmen der Zertifizierung genutzt zu werden?

Systementwurf und Realisierung

Bedeutet, die Lösungskonzepte und ihre Eigenschaften für Systemintegrität in verteilten, strukturdynamischen, evolutionären, intelligenten autonomen Systemen, im offenen dynamischen Kontext zu benennen

- System-Self-Awareness:
 - Modelle, Architektur und Mechanismen für verteiltes Selbst-beobachten
 - Modelle, Architektur und Mechanismen für verteilte Selbst-Erklärung
 - Modelle, Architektur und Mechanismen für verteiltes Selbstverstehen
- System-Self-Developing:
 - Modelle, Architekturen und Mechanismen für verteilte Selbst-Begrenzung
 - Modelle, Architekturen und Mechanismen für verteilte Selbst-Rekonfiguration
 - Modelle, Architekturen und Mechanismen für verteilte Selbst-Heilung
- Kontext Awareness
 - Modelle, Architektur und Mechanismen für die Beobachtung und Klassifizierung von Unbekanntem im Kontext.
 - Modelle, Architekturen und Mechanismen für nutzungsorientierte Repräsentation von erlerntem Wissen über den Kontext
- Selbst Lernen
 - Modelle, Architektur und Mechanismen für die Entwicklung von Hypothesen und deren Weiterentwicklung.
 - Modelle, Architektur und Mechanismen für die Entwicklung von Experimenten und deren Durchführung
 - Modelle, Architektur und Mechanismen für die Interpretation von Beobachtungen aus Experimenten vor dem Hintergrund von Hypothesen

Systemverifikation und -Validierung

bedeutet die Verifikations-Konzepte und ihre Eigenschaften zu benennen

- Modelle, Architektur und Mechanismen für verteilte (Selbst-)Validierung
- Mechanismen zur sicheren vertrauenswürdigen Speicherung von nachweispflichtigen Informationen

Connected Systems

- verteilte Integritätsmodelle: Methoden und Werkzeuge zur Sicherstellung der Integrität über Systemgrenzen hinweg sowohl zur Design-Zeit wie auch zur Laufzeit.
- Herstellen von Vertrauen: Methoden zum Austausch von Informationen zwischen Systemen mit denen die Vertrauenswürdigkeit von Systemen ermittelt werden kann.
- Security for Integrity: Methoden zur Evaluation des eigenen Zustands in Bezug auf IT-Sicherheit, Attacken, etc. sowie Methoden zur Wiederherstellung sicherer Zustände nach Attacken.

Hochautomatisierte Systeme

- Integritätstests: Methoden und Werkzeuge zur Sicherstellung von funktionaler, struktureller und semantischer Integrität sowie zur Überprüfung entsprechender Bedingungen auch zur Laufzeit sowie auch in Anwesenheit von Security-Attacken.
- Kontext-Integrität: Methoden zur Ermittlung der Integrität des Kontexts und der daraus wahrgenommenen Informationen, inklusive der Integrität von Informationen aus der Cloud bzw. von der Infrastruktur, zur Verbesserung der IT-Sicherheit.
- Behandlung von Unsicherheit: In der Kontextwahrnehmung sowie in der Vertrauenswürdigkeit extern erhaltener Informationen (aus der Cloud; von der Infrastruktur).
- On-line Verifikation: Methoden und Verfahren zur on-line Verifikation des Systemzustands (auch im Fehlerfall).
- On-line Validierung der Verfügbarkeit von (System-)Ressourcen.
- Dynamisches Verhalten: Methoden und Werkzeuge zur Sicherstellung der Korrektheit des intendierten Verhaltens in der gegenwärtigen (ggfs. auch nicht-spezifizierten) Situation.
- (Deep) Learning: Methoden und Werkzeuge zur Sicherstellung der Integrität von Lernverfahren.
- Laufzeit Integrität: Methoden und Werkzeuge zur Sicherstellung der Integrität zur Laufzeit auch für veränderte/neue Systemkomponenten (z.B. durch Lernen, durch Updates, durch kreative Aktionen des Systems, etc.). Angestrebte Vision: „Der intelligente On-Board Integritäts-Ingenieur“.

6.3 Detaillierte Beschreibung: Absicherung

6.3.1 Analyse

Der Begriff der Zertifizierung wird hier für den Nachweis einer erfolgreichen Qualitätsabsicherung gegenüber entsprechender Standards, Richtlinien und Regularien verwendet. Zertifizierbarkeit ist eine wesentliche Eigenschaft zukünftiger Systeme und entscheidender Baustein zur Schaffung von Vertrauen. Die Tragweite der durch solche neuartigen Systeme durchgeführten Entscheidungen und Handlungen als auch die in solchen Systemen verarbeitete Datenmenge nimmt rasant zu, so dass Qualitätseigenschaften wie funktionale Sicherheit, IT-Sicherheit und Verfügbarkeit sowie der Nachweis dieser Eigenschaften in ihrer Bedeutung deutlich über das ohnehin schon hohe Maß traditioneller Systeme steigen. Gleichzeitig wird der Nachweis der Konformität zu gesellschaftlichen, ethischen und moralischen Normen und Standards – soweit sie über die Konformität zu rechtlichen Rahmenbedingungen hinausgehen – erstmalig Teil der Qualitätsnachweise für solche Systeme. Daher kann Zertifizierbarkeit als wichtiger „Enabler“ verstanden werden, welcher großen Einfluss auf den wirtschaftlichen und gesellschaftlichen Erfolg neuartiger Systeme hat. Dies gilt insbesondere für kritische Domänen, in welchen ein Nachweis der Sicherheit unabdingbar ist und ein System entsprechend ohne einen solchen Nachweis nicht in den Markt eingeführt werden kann.

Etablierte Nachweis-/Zertifizierungsmethoden und -standards gehen dabei bisher von einem kompletten Verständnis eines Systems und seiner Umgebung aus („Closed World Assumption“). Dies ist bei den betrachteten evolutionären Systemen zunehmend nicht mehr gegeben („Open-World Assumption“): Systeme erhalten in der Zukunft mehr und mehr die Fähigkeit sich anzupassen, sich weiterzuentwickeln und mit anderen, vorher ggfs. unbekanntenen Systemen sinnvoll zu kooperieren. Für die Absicherung und Zertifizierbarkeit ergeben sich daraus zwei wesentliche Änderungen:

- Unvollständige Systemspezifikationen: Aufgrund der Open-World-Assumption sowie der Evolutionsfähigkeit von Systemen wird es in Zukunft weder möglich sein, die Anwendungsszenarien (ODDs: Operational Design Domain) noch das System selbst vollständig zu spezifizieren. Dadurch können auch die darauf basierenden Nachweise von Qualitätseigenschaften zur Entwicklungszeit nur unvollständig durchgeführt werden und müssen zumindest zum Teil zur Laufzeit erfolgen.
- Korrektheit im Kontext: Aufgrund der inhärent unvollständigen Spezifikation des Kontexts sowie der Anforderung, dass solche Systeme auch in unbekanntem Situationen sinnvoll agieren müssen (und sei es „nur“ als „Rückfall auf eine abgesicherte, eingeschränkte Funktionalität“), kann nicht mehr über eine generelle Korrektheit von Aktionen argumentiert werden. Eine Aktion kann immer nur in Bezug auf den aktuellen Kontext bzw. die aktuelle Situation, in der diese Aktion ausgeführt wird, korrekt oder nicht korrekt sein. Auch hierzu ist somit eine Qualitätsbewertung von Aktionen zur Laufzeit notwendig.

Ändert sich somit ein System (durch Evolution) oder wird es in neuen, bisher nicht in der Absicherung berücksichtigten Kontexten eingesetzt, so muss dies im Hinblick auf die zu zertifizierenden Eigenschaften zur Laufzeit bewertet werden. Und wird diese Änderung durch das System selbst herbeigeführt, so muss auch die Bewertung und Re-Zertifizierung durch das System erfolgen – oder zumindest initiiert werden. Um dies realisieren zu können benötigen Systeme spezifische Fähigkeiten zur Überprüfung ihrer Integrität (vgl. vorangehender Abschnitt), zur Analyse der zu überprüfenden Qualitätseigenschaften sowie zur Validierung dieser Eigenschaften.

Welche dieser Fähigkeiten wie umfassend jeweils vorhanden sein müssen, hängt von verschiedenen Komplexitätsgraden ab:

- Dies ist zunächst der Vernetzungsgrad der CPS (vgl. Kapitel 1). Während die für die Zertifizierung nachzuweisenden Eigenschaften für individuelle CPS sich typisch nur auf das System selbst und seinen Kontext beziehen, sind für höhere Vernetzungsgrade oft Eigenschaften nachzuweisen, die nur im Zusammenspiel der CPS einer Gruppe oder eines Kollektivs gelten und daher oft aufwändiger und ggfs. auch nur durch verteilte Verfahren nachweisbar sind.
- Der zweite beeinflussende Faktor ist die Komplexität des Kontexts (vgl. Kapitel 2). Als für die hier gewählte Darstellung ausreichende Generalisierung der in Kapitel 2 dargestellten Halbbordung der Kontext-Komplexität beschränken wir uns hier auf die Komplexitätsstufen bekannt-determiniert, bekannt-nichtdeterminiert und unbekannt-nichtdeterminiert.

Insgesamt ergeben sich somit Ausbaustufen für die Fähigkeiten zur Absicherung von Systemen, die in sich gestuft sein müssen nach sogenannten Risikoklassen eines Systems. Diese Risikoklassen ergeben sich prinzipiell aus

- den Fähigkeiten des Systems, also seiner Stärke, dem beherrschten Kontext, sowie den jeweiligen Ausbaustufen in allen Komplexitätsdimensionen
- einer dazu in Bezug stehenden Risikobewertung von Fehlverhalten des Systems.

Die konkrete Festlegung dieser Risikoklassen und der aus jeder Risikoklasse abgeleiteten Maßnahmen sind nicht Gegenstand dieser Roadmap. Hierzu sind weitere Diskussionen mit allen relevanten Stakeholdern – Herstellern, Industrieverbänden, Regulierungsbehörden und –institutionen – notwendig, welche letztlich zu entsprechenden Standards und Regularien führen werden, gegen die eine konkrete Zertifizierung ausgeführt werden wird.

6.3.2 Capabilities

Um Zertifizierung in oben beschriebenen Sinne durchführen zu können, sind die folgenden Capabilities nötig:

Off-line Zertifizierung

Die Zertifizierung erfolgt außerhalb des Egosystems gemäß den Regeln der zuständigen regulativen Behörde.

On-line Zertifizierung

Die Zertifizierung erfolgt im Egosystem selbst. Das beinhaltet, dass die Maßnahmen, die im Egosystem zur Selbstzertifizierung verwendet werden, durch die zuständige regulative Behörde (off-line) zertifiziert wurden. Entsprechend liegt das Gewicht der Begründung der Einhaltung relevanter Eigenschaften auf der Entwicklungszeit. Es werden lediglich einige klar definierte Überprüfungen in die Laufzeit geschoben, um Unsicherheiten bezüglich des Kontextes dynamisch aufzulösen und so unnötige worst-case Abschätzungen zu vermeiden.

On-line Zertifizierung und automatische Argumentation für eingeschränkte Kontexte:

Zusätzlich: Das Egosystem ist in der Lage, alle qualitätsrelevanten Eigenschaften, Entscheidungen und Handlungen zu begründen und diese Begründungen in die on-line Zertifizierung einbeziehen zu können. Der Kontext ist bekannt, entsprechend ist eine Vorabanalyse und Vorbereitung der Argumentation zur Entwicklungszeit möglich, das System hat aber die Fähigkeit auf Basis von Laufzeitanalysen die Argumentation dynamisch anzupassen.

Zertifizierung und automatische Argumentation für teilweise bekannte Kontexte; Begründung der Einhaltung von normativen Regelwerken und Gesetzen:

Zusätzlich: Das Egosystem kann die Auswirkungen seiner Eigenschaften und seines Handelns auf vorgegebene Regelwerke reflektieren und deren Einhaltung on-line zertifizieren. Wie auch bei den vorgenannten Punkten beinhaltet dies, dass die Maßnahmen, die im Egosystem zur Selbstzertifizierung verwendet werden, durch die zuständige regulative Behörde (off-line) zertifiziert wurden. Das System besitzt nun aber weiterführende Fähigkeiten zur Analyse und Argumentation, was den Einsatz in nur teilweise bekanntem (und durch die Regelwerke adressierten) Kontext ermöglicht.

On-line Zertifizierung und automatische Argumentation für unbekannte Kontexte; Begründung der Einhaltung ethischer Grundsätze:

Zusätzlich: Das Egosystem kann die Auswirkungen seines Handelns und die Einhaltung gesellschaftlich geforderter ethischer Prinzipien reflektieren und deren Einhaltung on-line zertifizieren. Wie auch bei den vorgenannten Punkten beinhaltet dies, dass die Maßnahmen, die im Egosystem zur Selbstzertifizierung verwendet werden, durch die zuständige regulative Behörde (off-line) zertifiziert wurden. Die Analyse- und Argumentationsfähigkeiten des Systems erlauben nun auch den Einsatz in unbekanntem Kontext, für welchen gegebenenfalls auch keine expliziten Regelwerke existieren.

6.3.3 Herausforderungen und Forschungsfragen

Zur Realisierung der Capabilities sind – neben den unter 6.2. beschriebenen Herausforderungen für die Sicherstellung der Systemintegrität – die folgenden Herausforderungen für die benötigten Analysefähigkeiten und Zertifizierungsfähigkeiten der Systeme zu lösen:

- **Predeterminierte Selbst-Beobachtung**
 - zur Entwicklungszeit vollständig antizipierte und definierte, ggfs. konfigurierbare Erkennung und ggfs. Vorhersage des internen Zustands
 - Das System ist in der Lage, sich selbst zu beobachten, d.h. Statusdaten bzgl. des eigenen "Gesundheitszustands" aus verschiedenen seiner Komponenten zu messen. Aufgrund der Fülle dieser Statusdaten ist es ggfs. nicht möglich alle solche Daten gleichzeitig zu messen. Welche Daten im konkreten Fall gemessen werden, ist konfigurierbar. Diese Fähigkeit hat ihren Schwerpunkt zur Designzeit, da hier alles durch den Entwickler predeterminiert wird. Der Fokus liegt auf einzelnen CPS. Der Kontext muss bekannt sein.
- **Intelligente nicht-determinierte Selbst-Beobachtung**
 - Erkennung, ggfs. Vorhersage und Erkundung des internen Zustands.
 - Das System ist in der Lage, sich selbst zu beobachten, d.h. Statusdaten bzgl. des eigenen "Gesundheitszustands" aus verschiedenen seiner Komponenten zu messen. Gegenüber dem predeterminierten Fall hat das System die Fähigkeit diese Selbstbeobachtung den aktuellen Bedürfnissen anzupassen und auch unvorhergesehene Ereignisse richtig einzuordnen. Entsprechend werden nun Laufzeitfähigkeiten benötigt. Der Fokus ist nach wie vor das einzelne CPS, allerdings kann nun auch mit unbekanntem Kontext umgegangen werden.
- **Predeterminierte und ggfs. konfigurierbare Beobachtung des externen Kontext**
 - Das System ist in der Lage den relevanten externen Kontext zu beobachten. Welche Kontextinformationen das im Einzelnen sind und wie sie durch das System erfasst und verarbeitet werden wird komplett zur Entwicklungszeit determiniert. Gerade in dynamischen, volatilen und komplexen Kontexten entstehen dadurch Einschränkungen. Es handelt sich demnach um eine Designzeit-Fähigkeit, welche einzelne CPS oder bekannte Gruppen/Kollektive adressiert und von bekanntem Kontext ausgeht.
- **Intelligente nicht-determinierte Beobachtung des externen Kontexts**
 - Das System ist in der Lage den relevanten externen Kontext zu beobachten. Insbesondere ist das System auch in der Lage unbekanntem und unvorhergesehenen Kontext zu deuten und zu verwenden. Bei der nicht-determinierten Beobachtung verschiebt sich der Schwerpunkt der Fähigkeit automatisch in die Laufzeit, das System muss nun die „Intelligenz“ besitzen vorher nicht bekannten Kontext korrekt zu verarbeiten. Diese Fähigkeit adressiert alle CPS und Kontextklassen, bekannt und unbekannt.
- **Intelligente und kollaborative nicht-determinierte Beobachtung des externen Kontext**
 - Das System ist in der Lage den relevanten externen Kontext zu beobachten. Insbesondere ist das System auch in der Lage unbekanntem und unvorhergesehenen Kontext zu deuten und zu verwenden. Das System ist nun ferner nicht mehr als einzelnes CPS zu verstehen, sondern auch als Gruppe, Kollektiv oder heterogenes Kollektiv. Relevanter Kontext definiert sich entsprechend dynamisch aus der Beschaffenheit und Zielsetzung der aktuellen Kollaboration. Entsprechend adressiert diese Fähigkeiten alle kollaborativen (und nicht einzelnen) CPS und sowohl bekannten als auch unbekanntem Kontext.
- **Predeterminierte Analysen, welche zur Laufzeit durchgeführt werden**
 - Das System ist in der Lage spezifische predeterminierte Analysen durchzuführen (z. B. zur Generierung von Evidenzen eines Safety Nachweises / Assurance Cases zur Laufzeit). Diese Fähigkeit hängt eng mit den Fähigkeiten zur Erfassung des Kontextes

zusammen, da die Analysen typischerweise Kontextinformationen verwenden. Diese Fähigkeit hat trotz des Laufzeitanteils ihren Schwerpunkt in der Designzeit und kann nur für bekannte Konstellationen von CPS und bekanntem Kontext verwendet werden.

- **Automatische Generierung von Analysefragen für einzelne CPS oder bekannte Gruppen von CPS**
 - Das System ist in der Lage, aus globalen Anforderungen (z.B.: einzelnen Fällen einer Fallunterscheidung im Safety-Nachweis) Analysefragen angepasst auf die eigene aktuelle Systemkonfiguration und Selbstbeobachtungsmöglichkeiten zu generieren (sowie diese dann auch zu testen und zu bewerten, siehe "Selbstbewertung der Analyseergebnisse"). Damit verschiebt sich der Schwerpunkt der Fähigkeit in die Laufzeit. Trotzdem geht die Fähigkeit von bekannten Konstellationen von CPS und bekanntem Kontext aus.
- **Automatische Generierung von Analysefragen für heterogene teilweise unbekannte Kollektive von CPS**
 - Siehe die vorige Fähigkeit – nun allerdings mit der Möglichkeit mit unbekanntem Konstellationen von CPS und unbekanntem Kontext umgehen zu können.
- **Selbstbewertung der Analyseergebnisse für einzelne CPS oder bekannte Gruppen von CPS**
 - Das System ist in der Lage, die gemessenen Daten zum eigenen Gesundheitszustand bzw. zur Kontext/Situationserkennung zu analysieren (nach vorgegebenen oder selbst generierten Analysefragen) und die Ergebnisse der Analysen zu bewerten (sowie entsprechende Maßnahmen wie "Freigabe" oder "Enter Safe-Fallback-Mode") durchzuführen. Diese Fähigkeit hat trotz des Laufzeitanteils ihren Schwerpunkt in der Designzeit und kann nur für bekannte Konstellationen von CPS und bekanntem Kontext verwendet werden.
- **Selbstbewertung der Analyseergebnisse für heterogene teilweise unbekannte Kollektive von CPS**
 - Siehe die vorige Fähigkeit – nun allerdings mit der Möglichkeit mit unbekanntem Konstellationen von CPS und unbekanntem Kontext umgehen zu können.
- **Fähigkeit zur prädiktiven Analyse.**
 - z.B. Erkennung von Absichten von Menschen/Systemen
 - Wie oben "Selbstbewertung von Analyseergebnissen", jedoch zusätzlich auch Analyse der Prädiktionsgüte (bzgl. (a) der zukünftigen Entwicklung der Situation bzw. (b) zukünftig auftretender Fehler/zukünftiger "mangelnder Gesundheit"). Diese Fähigkeit ist unabhängig von den Kategorien der Komplexitätsklassifikation.

- **Off-line Certification of on-line Certification**
 - Die im System bestehenden Verfahren (z.B. Implementierung von Monitoren, Implementierungen der V&V Tests für den on-line-Teil der Zertifizierung von Updates) müssen selbst zertifiziert sein. Da diese selbst nicht veränderbar sind, reicht eine off-line-Zertifizierung vor in-Verkehr-bringen des Systems aus. Diese Fähigkeit verwendet predeterminierte Kontextfähigkeiten und fokussiert sich auf einzelne CPS und bekannten Kontext.
- **Nachweisführung (mit Laufzeitunterstützung) für predeterminierte dynamische Kollaborationshierarchien und Systemkonfigurationen**
 - Die für die Bildung von Gruppen und (hierarchischen) Kollektiven genutzten Verfahren und Kommunikationsformen müssen validiert/zertifiziert werden. Gleiches gilt für sich über die Zeit verändernde Systemkonfigurationen (durch Ausfall von Komponenten

ten, durch Rekonfiguration, durch Updates/Evolution,...).

- Teile dieser Validation erfolgen on-line (“Laufzeitunterstützung”), der Schwerpunkt liegt durch die Predetermination auf der Designzeit, es werden bekannte Konstellationen von CPS und bekannter Kontext adressiert.
- **Laufzeitnachweis für bekannten determinierten Kontext**
 - On-line/On-board Komponente der Sicherheitsnachweise für Systeme, die sich in bekanntem determiniertem Kontext bewegen (vgl. “Kontext”). Assurance Cases können zur Designzeit (off-line) erstellt werden, sind jedoch modular angelegt und enthalten Annahmen bzgl. des Kontexts die zur Laufzeit kontinuierlich zu überwachen sind. Die zur Laufzeit nötigen Verfahren basieren wahrscheinlich auf Contracts, ConSerts, o.ä. Forschung und prototypische Anwendungen hierzu haben bereits begonnen. Weitere Forschung ist nötig um den Umgang mit Unsicherheiten bzgl. des Kontexts zu verbessern. Der Schwerpunkt der Fähigkeit liegt in der Entwicklungszeit und sie adressiert bekannte Konstellationen von CPS und bekanntem Kontext.
- **Laufzeitnachweis für bekannten nicht-determinierten Kontext**
 - Wie für bekannten determinierten Kontext, jedoch erhöhte Schwierigkeiten aufgrund von größerer Kontextkomplexität, zusätzlich benötigten Erkennungsfähigkeiten (siehe oben), Unwissen über (Evolution des) Kontexts und dessen Auswirkungen auf die Komplexität der Umwelt.
 - Forschung hierzu in wenigen Ansätzen vorhanden, ConSerts bieten zB die Möglichkeit bekannten nicht-determinierten Kontext über Varianten in den bedingten Zertifikaten und Laufzeitevidenzen abzubilden. Auch hier liegt ein Schwerpunkt der Fähigkeit in der Entwicklungszeit, da der Assurance Case (oder ein ähnliches Artefakt) das Rückgrat bildet und entsprechend der menschliche Ingenieur die wesentliche „Safety Engineering Intelligenz“ beisteuert. Gleichzeitig hat diese Fähigkeit nun aber auch einen ausgeprägten Laufzeitanteil, mit unter Umständen bereits recht komplexen und umfassenden Laufzeitmodellen. Trotzdem liegt der Fokus auf bekannten Konstellationen von CPS und bekanntem Kontext.
- **Laufzeitnachweis für unbekanntem undeterminierten Kontext**
 - Wie oben, hier für unbekanntem-nicht-determinierte Kontexte. Diese Fähigkeit ist stark abhängig vom Vorhandensein der entsprechenden Fähigkeiten im Bereich “Kontext”, “Kollaboration” und “Intelligenz”.

6.4 Benötigte Technologien und Zeithorizont

Die folgende Tabelle gibt für jede Ausbaustufe dieser Komplexitätsdimension die größten Herausforderungen sowie die wichtigsten benötigten Technologien zu ihrer Realisierung an und ordnet diese zeitlich ein. Diese Analyse ist die Grundlage für die in Kapitel 7 dargestellte Roadmap. (Tabelle siehe folgende Seite.)

Dimension	Ausbaustufen	Major Challenge	Kerntechnologie	Kurz 2027	Mittel 2035	Lang 2045ff	
Integrity Die Fähigkeit des Systems seinen inneren Zustand zu bewerten und Auswirkungen von systemischen Veränderungen auf den Systemzustand vorherzusagen.	Information Integrity		Anomalie Detection	Vollumfänglich			
		Self-Awareness of Information Integrity	Formalisierung von Integritätsbedingungen (an Daten und Zugriffsstrukturen) sowie aus logischen und physikalischen Gesetzmäßigkeiten abgeleitete Plausibilitätsbedingungen	Vollumfänglich			
			Automatische Echtzeit-Erkennung der Verletzung von Integritätsbedingungen	Bei Verwendung von hochleistungsfähiger Hardware	In ressourcen-restringierten Anwendungsdomänen		
			Architekturkonzepte for Self-Awareness of Information Integrity	Vollumfänglich			
		Self-Awareness of Distributed Information Integrity		Vollumfänglich			
		Self-Healing for Information Integrity	Transaktionskonzept zur Sicherstellung der Information-Integrity	Vollumfänglich			
			Architekturkonzepte für Self-Healing for Information Integrity	Vollumfänglich			
		Self-Healing for Distributed Information Integrity	Ledger-Technologies	Vollumfänglich			
	State Integrity	Wie oben, allerdings für State Integrity; zusätzlich:					
		Assuring State-Integrity under System-Transitions	On-Line Complete Digital Twin mit Real-Time Simulation Capabilities	Vollumfänglich			
			Formalisierung von State-Integrity Bedingungen	Vollumfänglich			
			Formalisierung von Distributed State Integrity Bedingungen	Vollumfänglich			
		Assuring State-Integrity under transitions in systems-of-systems	Distributed Agreement Protocols for Coordinated Transition Taking in SoS	Vollumfänglich			
		Self-Healing for State-Integrity	Anomalie Detection through monitoring of state invariants	In nicht ressourcen-restringierten Systemen und/oder in stark eingeschränkten Kontexten	In nicht ressourcen-restringierten Systemen für bekannte Kontexte		
			Redundanz-Konzepte, z.B. unter Ausnutzung von Digital Twins	In nicht ressourcen-restringierten Systemen und/oder in stark eingeschränkten Kontexten	In nicht ressourcen-restringierten Systemen für bekannte Kontexte		
			Anomalie Detection through monitoring of state invariants	In nicht ressourcen-restringierten Systemen und/oder in stark eingeschränkten Kontexten	In nicht ressourcen-restringierten Systemen für bekannte Kontexte		

Dimension	Ausbaustufen	Major Challenge	Kerntechnologie	Kurz 2027	Mittel 2035	Lang 2045ff	
		Self-Healing for Distributed State-Integrity	Identification of Reset-States of System violating Integrity-Condition, Compatible with Distributed State-Integrity Condition	In nicht ressourcen-restringierten Systemen und/oder in stark eingeschränkten Kontexten	In nicht ressourcen-restringierten Systemen für bekannte Kontexte		
			Immune-System: Detection and Isolation of Sub-Systems violating State-Integrity	In nicht ressourcen-restringierten Systemen und/oder in stark eingeschränkten Kontexten	In nicht ressourcen-restringierten Systemen für bekannte Kontexte		
	Plan Integrity	Wie oben, allerdings für Pläne					
		Assuring State Integrity until mission completeness	On-Line Digital Twin mit real-time Simulation Capabilities	In nicht ressourcen-restringierten Systemen und/oder in stark eingeschränkten Kontexten	In nicht ressourcen-restringierten Systemen für bekannte Kontexte		
			Formalisierung von Plan-Integrity Bedingungen				
		Assuring plan integrity under update of plans	Determining safe pre-emption points; avoiding inconsistencies between pre-empted and new plans	In nicht ressourcen-restringierten Systemen und/oder in stark eingeschränkten Kontexten	In nicht ressourcen-restringierten Systemen für bekannte Kontexte		
		Assuring State Integrity until mission completeness for shared plans	Digital Twin of System-of-Systems	In nicht ressourcen-restringierten Systemen und/oder in stark eingeschränkten Kontexten	In nicht ressourcen-restringierten Systemen für bekannte Kontexte		
			Real-time Simulation of Digital Twins for SoS	In nicht ressourcen-restringierten Systemen und/oder in stark eingeschränkten Kontexten	In nicht ressourcen-restringierten Systemen für bekannte Kontexte		
		Self-Healing for Plan Integrity	Detection of anticipated future State-Integrity Violations: Abort plan and switch to integer plan	Vollumfänglich			
	Self-Healing for Distributed Plan Integrity	Identification of System(s) violating plan integrity; re-negotiation of integer distributed plan	Vollumfänglich				
	Evolution Integrity	Wie oben, unter allen Evolutionsstufen					
		Adaption der Maßnahmen (Integritätsbedingungen definieren, prüfen und im Falle von Verletzungen heilen) für Externally Controlled Evolution			Wie für die oben dargestellten Stufen	Wie für die oben dargestellten Stufen	
		Adaption der Maßnahmen (Integritätsbedingungen definieren, prüfen und im Falle von Verletzungen heilen) für Guided Evolution	Automatische Generierung von Integritätsbedingungen und Maßnahme zur Erkennung von Verletzungen und Heilung				Vollumfänglich
		Adaption der Maßnahmen (Integritätsbedingungen definieren, prüfen und im Falle von Verletzungen heilen) für Collaborative Evolution	Automatische Generierung von Integritätsbedingungen und Maßnahme zur Erkennung von Verletzungen und Heilung				Vollumfänglich
		Adaption der Maßnahmen (Integritätsbedingungen definieren, prüfen und im Falle von Verletzungen heilen) für Self Evolution	Automatische Generierung von Integritätsbedingungen und Maßnahme zur Erkennung von Verletzungen und Heilung				Nicht in den nächsten 50 Jahren

Dimension	Ausbaustufen	Major Challenge	Kerntechnologie	Kurz 2027	Mittel 2035	Lang 2045ff
	Learning Integrity	Wie oben, unter allen Intelligenzstufen	Identification of anomalies and biases in learned data	For supervised learning in known, restricted contexts		For supervised learning in known contexts
			Identification of anomalies for learned plans based on plausibility constraints (siehe oben)	For supervised learning in known, restricted contexts	For supervised learning in known contexts	For supervised learning in known contexts
			All other cases			Nicht in den nächsten 50 Jahren
Certifiability Die Fähigkeit des Systems sich selbst zu zertifizieren, d.h. für den Betrieb in einem Operativen Kontext zuzulassen.	Off-line Certification	Not usable for complete validation/certification of systems of complexity classes „guided Evolution“ and higher; not usable for „unsupervised learning“ and higher classes; not usable for „autopoietic Autonomie				
		Generation of Certification Evidence that the Egosystem behaves correctly in all allowed contexts	Demonstration of Completeness of Certification Evidence	Für eingeschränkte Kontexte	In partiell bekannten Kontexten	Inhärent unmöglich für unbekannte Kontexte
			Simulationsbasierte Ansätze	Für eingeschränkte Kontexte	In partiell bekannten Kontexten	Inhärent unmöglich für unbekannte Kontexte
			Nachweis entsprechender Architektureigenschaften (Safety, Security, Reliability,...)	Für eingeschränkte Kontexte	In partiell bekannten Kontexten	Inhärent unmöglich für unbekannte Kontexte
		Agreement and formalization of relevant Contexts (incl. known-unknowns)	Determination of metrics for acceptable definitions of relevant context	Für eingeschränkte Kontexte	In partiell bekannten Kontexten	Inhärent unmöglich für unbekannte Kontexte
			Erstellung von Szenarien-(Test-)Katalogen, die den relevanten Kontext überdecken	Für eingeschränkte Kontexte	In partiell bekannten Kontexten	Inhärent unmöglich für unbekannte Kontexte
			Nachweisbar vollständige Situationsbeherrschung in Bezug auf relevante Kontexten	Für eingeschränkte Kontexte	In partiell bekannten Kontexten	Inhärent unmöglich für unbekannte Kontexte
		Additionally: Generation of Certification Evidence that the Egosystem ensures integrity		Für eingeschränkte Kontexte	In partiell bekannten Kontexten	Inhärent unmöglich für unbekannte Kontexte
		Additionally: Generation of Certification Evidence that the Egosystem ensures responsibility		Für eingeschränkte Kontexte	In partiell bekannten Kontexten	Inhärent unmöglich für unbekannte Kontexte
		Demonstration of compliance of processes (incl. development, safety, security, usability,...) to requirements of certification authorities		Für eingeschränkte Kontexte	In partiell bekannten Kontexten	Inhärent unmöglich für unbekannte Kontexte
		Brauchbarkeitsnachweis: Durch Nutzung des Systems (in der realen Welt oder in Simulation) in ausreichender Menge Nachweis eines genügend kleinen Restrisikos	Nur realistisch durchführbar in Closed-World-Systems (known Context)	Für eingeschränkte Kontexte	In partiell bekannten Kontexten	Inhärent unmöglich für unbekannte Kontexte
Generation of Justification and Explanation of safety relevant decisions and actions of Egosystem		Für eingeschränkte Kontexte	In partiell bekannten Kontexten	Inhärent unmöglich für unbekannte Kontexte		

Dimension	Ausbaustufen	Major Challenge	Kerntechnologie	Kurz 2027	Mittel 2035	Lang 2045ff
	On-line Certification	Automatic Generation of certification evidences (Design, Process, and Usability/Brauchbarkeit)	On-line Zertifizierung for turning unknown context into known context		Für eingeschränkte Veränderungen	Allgemeinere Klassen
			On-line Zerifizierung of safety of evolution steps for all complexity degrees of evolution		Für eingeschränkte Veränderungen	Allgemeinere Klassen
			On-line Zerifizierung of safety of learning steps for all complexity degrees of intelligence		Für eingeschränkte Veränderungen	Allgemeinere Klassen
			Certification on demand (nur noch die jeweils aktuelle Handlung im jeweils aktuellen Kontext)		Für eingeschränkte Veränderungen	Allgemeinere Klassen
			Basis hierfür: Theorie der Analyse und Synthese		Für eingeschränkte Veränderungen	Allgemeinere Klassen
			Off-line Zertifizierung der verwendeten Verfahren zur On-line Zertifizierung (für Autonomie, für Evolution und für Intelligenz)	Nachweis der Vollständigkeit der für off-line-Zertifizierung verwendeten Verfahren zur Sicherstellung der Systemicherheit in Kombination mit den On-line Zertifizierungsverfahren		Für eingeschränkte Veränderungen
	On-line Certification and Justification	Additionally to be included into certification evidence: For Self-learning or higher intelligence: Generation of justifications for all safety critical actions of Egosystem			Für eingeschränkte Veränderungen	Allgemeinere Klassen
		Additionally to be included into certification evidence: Explainability of learning based classifications and planning algorithms			Für eingeschränkte Veränderungen	Allgemeinere Klassen
	Certification & Justification of compliance rules				Für eingeschränkte Veränderungen	Allgemeinere Klassen
	Certification & Justification of ethical rules				Für eingeschränkte Veränderungen	Allgemeinere Klassen

7 Roadmap

Die in dieser Roadmap diskutierten Komplexitätsdimensionen und Evolutionsszenarien erscheinen auf den ersten Blick futuristisch. Betrachtet man allerdings bereits bestehende Produkt-Roadmaps z. B. der Mobilitätsindustrie, erkennt man unmittelbar, dass ein Teil dieser Zukunft dabei ist Realität zu werden. Prototypen selbstfahrender Fahrzeuge hat jeder schon gesehen. Verkehrskolonnen, die aus Effizienz- und Umweltgründen einen kooperativen Verbund von Fahrzeugen bilden, sind schon in der Diskussion der Verkehrs- und Logistikplaner. Verkehrsleitsysteme, die um Durchsatz zu erhöhen und Emission zu reduzieren Fahrzeuge und Fahrzeugverbände steuern, sind mancherorts schon in der Einführung. Smart Cities erschließen mit einem Mix aus verschiedenen Verkehrskonzepten unsere Städte, machen sie besser bewohnbar, attraktiv und verbinden Domänen wie Mobilität, Energie, Gesundheit u. a. m. Diese und ähnliche Entwicklungen sind bereits auf der Agenda von Politik und Industrie. Ihre ersten Ausbaustufen werden nicht alle Eigenschaften der Komplexität umfassen, sondern gerade dadurch möglich, dass der operative Kontext eingeschränkt wird. Die Kernfrage, wie diese Systeme safe, secure und certifiable eingeführt und betrieben werden können, bleibt uneingeschränkt gültig und muss synchron zur Einführung der Systeme technisch und regulatorisch beantwortet werden. Eine Ausbaustufenstrategie erlaubt es uns aber zu lernen und zu wachsen. Anhand von vier Beispielen werden die dominanten Charakteristika und ihre signifikanten einschränkenden Annahmen kurz skizziert und in die Komplexitätsdimensionen eingeordnet.

Beispiel 1: Autonom fahrendes Fahrzeug

Ein Fahrzeug mit SAE-Level 4/SAE-Level 5¹ erfasst auf Basis fahrzeugeigener Sensorik seinen relevanten Kontext und trifft in urbanen Fahrsituationen angemessene Manöverentscheidungen.

Die Herausforderungen liegen bei ...

- der Erstellung eines verlässlichen, vollständigen Lagebilds mit existierender Sensortechnologie
- einer belastbaren Argumentation, dass das Fahrzeug in allen bekannten und partiell bekannten Situationen angemessene Manöverentscheidungen trifft und durchführt
- der Erstellung eines Sicherheitsnachweises, dass von diesem System keine unangemessene Gefahr ausgeht, was sowohl funktionale Fehler als auch Angriffe auf das System einschließt

Komplexitätsreduzierende Einschränkungen sind:

- Fahrten in eingeschränkten Umgebungen wie Logistikhöfen, Flughäfen und ähnliches
- Fahren mit eingeschränkter Fahrdynamik, geringer Geschwindigkeit, reduzierten Manövern,
- defensiver Fahrstrategie
- Unterstützung des situativen Lagebildes durch Infrastruktur

Beispiel 2: Verkehrskolonne (Platoon)

Ein Verbund von SAE-Level 4/SAE-Level 5 autonomen LKW bildet eine kooperative Gruppe auf der Basis gemeinsamer Ziele, wie gemeinsame Route, Energieeffizienz oder Emissionsreduktion. Die Kooperation findet dabei bei der Erstellung eines gemeinsamen Lagebilds, dem kooperativen

¹ Erklärung der Stufen der Fahrautomatisierung siehe: SAE International. SAE International Releases Updated Visual Chart for Its "Levels of Driving Automation" Standard for Self-Driving Vehicles. 2018

Manöverentscheidungen und dem kooperativen Planen zur Erfüllung der gemeinsamen individuellen Mission unter Berücksichtigung der Fähigkeiten der einzelnen Fahrzeuge statt. Dazu muss jedes Fahrzeug bereit und fähig sein, individuelle Optima zugunsten gemeinsamer Ziele anzupassen.

Die Herausforderungen liegen bei ...

- einer dynamische Veränderung der Struktur der Gruppe über die Zeit
- einer kooperativen Erstellung eines vollständigen Lagebilds aus den Einzelbildern der Gruppenteilnehmer
- einer dynamischen Missionsplanung, Ziele-, Strategie- und Manöverabstimmung
- der Beherrschung aller möglichen auftretenden Fahrsituationen
- der dynamischen Koordination vieler komplexer Manöver im kooperativen Verbund
- der Sicherstellung und Überwachung der Integrität der Gruppe mit Strukturdynamik

Komplexitätsreduzierende Einschränkungen sind:

- infrastrukturbasierte Erstellung des Lagebilds
- Fahren auf definierten vorher bekannten und vermessenen Strecken (L4)
- Reduktion der möglichen Manöver, z. B. kein Spurwechsel, kein Überholen, u. a. m.
- Reduktion der Entscheidungskomplexität, indem nur das Führungsfahrzeug entscheidet
- Reduktion der Kooperationskomplexität durch homogene Fahrzeuge in einer Gruppe
- Reduktion der Strukturdynamik durch vordefinierte Zusammenstellung der Gruppe

Beispiel: Verkehrsleitsystem

Zum Zweck der Emissionsreduktion, des Verkehrsdurchsatzes oder für die vorrangige Behandlung im Notfall werden komplexe Verkehrsabschnitte so gesteuert, dass diese Ziele erreicht werden können. Dabei geht man davon aus, dass die teilnehmenden Systeme autonom anpassungsfähig sind. Gegebenenfalls können einzelne Systeme einen höheren Quality of Service (QoS) mit dem Verkehrsleitsystem aushandeln.

Die Herausforderungen liegen bei ...:

- der Erkennung und Beurteilung der aktuellen Verkehrslage
- der Vorhersage der zukünftigen Entwicklung durch externe und systeminterne Interventionen
- dem Verständnis und der Steuerung der Dynamik der Systemstruktur in einem Verkehrsabschnitt
- einem kooperativen Aushandeln von QoS für einzelne Teilnehmer, Gruppen, Kollektive
- der Beherrschung von gemischtem Verkehr mit autonomen und individuell geführten Fahrzeugen

Komplexitätsreduzierende Einschränkungen sind:

- Einschränken der Verhandlungsfähigkeit von Systemen und zentraler Verkehrssteuerung
- Verkehrsabschnitte mit exklusiver Nutzung für kooperative Teilnehmer
- Infrastruktur zur Erfassung der Verkehrslage

Beispiel: Smart City

In einer Smart City stimmen sich verschiedene Domänen wie Energie, Mobilität, Gesundheit u. a. m. ab, um wichtige Ziele der Einzeldomänen und globale Ziele der Stadt zu erreichen. Außerdem sollen Sondersituationen, wie beispielsweise Großveranstaltungen, Feinstaub, Notfälle, koordiniert behandelt werden.

Die Herausforderungen liegen bei ...

- der Beurteilung der Strukturdynamik in heterogenen Domänen und der daraus abgeleiteten Gesamtsituation
- dem gemeinsamen Planen in heterogenen Domänen, Abwägen von Prioritäten und Konflikten
- der heterogenen Kommunikation und Koordinationsmechanismen, inklusive der Auflösung von Konflikten

Komplexitätsreduzierende Einschränkungen sind:

- hierarchische Planung statt kooperativer Planung
- hierarchische Konfliktlösung statt kooperativer Konfliktlösung
- Einschränkung der Autonomie der einzelnen Domänen durch hierarchische Ressourcenzuteilung
- Einführung eines an Geldwerten orientierten Ressourcenäquivalents als Verhandlungsbasis (wie beispielsweise CO2-Preis)

Die folgende Abbildung ordnet die Beispiele in die fünf Komplexitätsdimensionen ein:

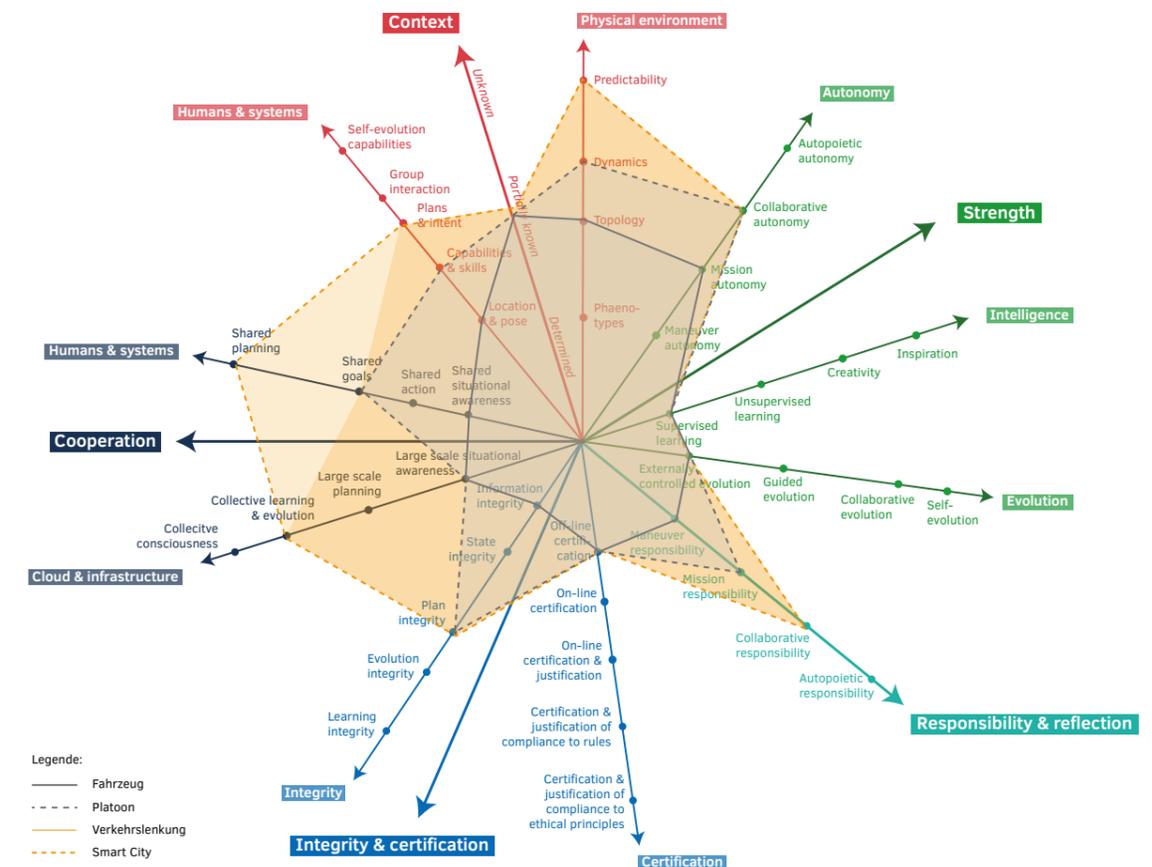


Abb. 5: Wachsende Systemkomplexität der Systemklassen am Beispiel von Mobilitätssystemen

Roadmap

Basierend auf den in den Kapitel X.2.4 vorgenommenen Identifikationen der benötigten Technologien und ihren zeitlichen Anordnungen entsteht die folgende Roadmap für die Erreichung der einzelnen Ausbaustufen in jeder der untersuchten Komplexitätsdimensionen.

Um diese Entwicklungen tatsächlich realisieren zu können, ist es notwendig, dass Forschung, Industrie und Politik zielgerichtet zusammen wirken (siehe auch „Handlungsempfehlungen“ im folgenden Kapitel).

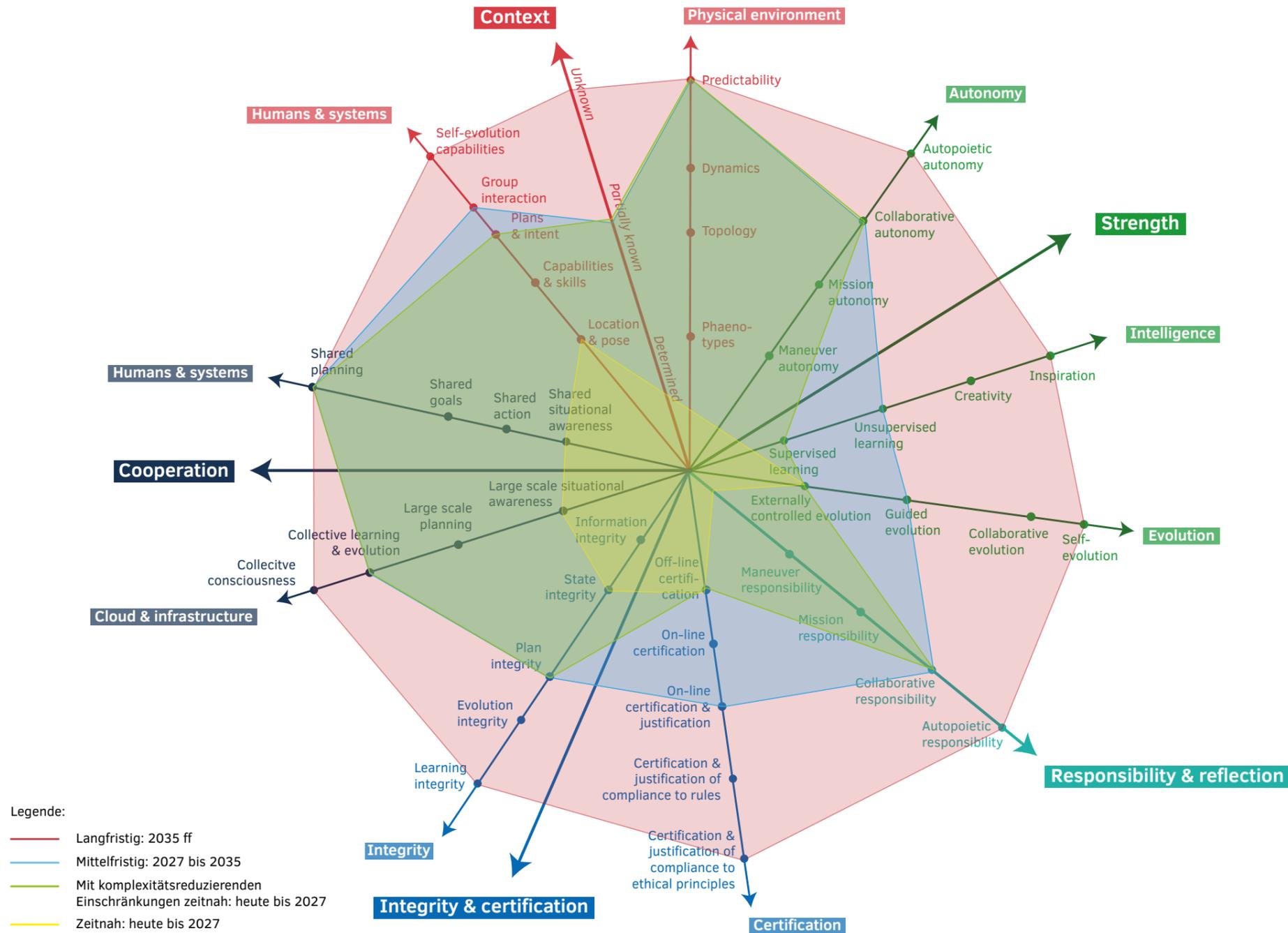


Abbildung 6: Zeitliche Vorhersage zum Erreichen der Fähigkeiten der Systemklassen

8 Handlungsempfehlungen

Zielsetzung

Absicherung des Alleinstellungsmerkmals „Quality Made in Germany“: Zukünftige in Deutschland hergestellte hochkomplexe Mensch-Maschine-Systeme gewährleisten Sicherheit, Verfügbarkeit und gesellschaftliche Akzeptanz.

A. FuE-Maßnahmen

- Aufsetzen von Förderprogrammen für sichere, zertifizierbare und verantwortliche Mensch-Maschine Systeme (Methoden, Prozesse, Metriken) einschließlich Verfahren zur virtuellen Absicherung
- Auf- und Ausbau von Forschungsinfrastrukturen, welche für Industrie und Wissenschaft frei zugänglich sind: sowohl Reallabore wie auch „Virtual Collaboration Spaces“ für repräsentative Anwendungsklassen, einschließlich begleitender empirischer Akzeptanzforschung
- Initiierung von Leuchtturmprojekten, welche die industrielle Beherrschbarkeit eines holistischen Gesamtansatzes zur Qualitätsabsicherung für verschiedene Branchen demonstrieren

B. Innovationsmaßnahmen

- Förderung von „Virtual Innovation Clustern“, welche die Großindustrie, KMU und Forschung für hochgradig relevante Anwendungsklassen zur Schaffung von dafür benötigten Innovationen unter Verwendung von „Virtual Collaboration Spaces“ zusammenbinden
- Initiierung und Förderung von Instrumenten zur Verbreitung von Best-Practices zur Erzielung dieser Qualität: Prozesse, Organisationsstrukturen, Weiterbildung

C. Regulatorische Maßnahmen

- Erarbeitung von Vorschlägen für Regularien (einschließlich Verankerung von Nachvollziehbarkeit, Zertifizierbarkeit, Verantwortung), für die Zertifizierung von hochkomplexen Mensch-Maschine-Systemen
- Einrichtung oder Benennung von unabhängigen Zertifizierungslaboren zur Vergabe von Zertifikaten für die Einhaltung der geforderten Qualitätsmaßnahmen
- Nach Vorliegen genügend umfassender Erfahrung über die Machbarkeit des Ansatzes: Etablierung internationaler Standards, einschließlich der Identifikation von System- und Risikoklassen, der Festlegung von Anforderungen an qualitätssichernde Maßnahmen pro Risikoklasse, sowie der Schaffung von dafür benötigten rechtlichen Rahmenbedingungen

D. Ausbildung

- Verankerungen der für die Herstellung hochkomplexer Mensch-Maschine-Systeme sowie deren Qualitätssicherung benötigten Kompetenzen in der universitären Ausbildung

E. Gesellschaftlicher Diskurs

- Initiierung von Diskursen zur gesellschaftlichen Relevanz und Akzeptanz solcher so qualitätsgesicherter zukünftiger Mensch-Maschine Systeme

9 Referenzen

AENEAS, ARTEMIS-IA. EPoSS (Hrsg.). Electronic Components and Systems Strategic Research Agenda (ECS SRA). Goirle (NL). 2019. <https://artemis-ia.eu/documents-1.html> (Zugriffsdatum: 02.12.2019)

BMVI, Ethik-Kommission (Hrsg.). Automatisiertes und Vernetztes Fahren. 2017. <https://www.bmvi.de/SharedDocs/DE/Publikationen/DG/bericht-der-ethik-kommission.html> (Zugriffsdatum: 02.12.2019)

European Commission (Hrsg.). ICT Standardisation Priorities for the Digital Single Market. Brüssel (B). 2016. <https://ec.europa.eu/digital-single-market/en/news/communication-ict-standardisation-priorities-digital-single-market> (Zugriffsdatum: 02.12.2019)

Guilford, Joy Paul/Hoepfner, Ralph. The analysis of intelligence. Weinheim (u.a.). Beltz. 1976

Maturana, H.R./Varela, F.J. The cognitive process. Autopoiesis and cognition: The realization of the living. Springer Science & Business Media. 1980. <https://www.springer.com/de/book/9789027710154> (Zugriffsdatum: 02.12.2019)

Plattform Lernende Systeme (Hrsg.). Sichere KI-Systeme in der Medizin - Positionspapier. Voraussichtlich 2019

Projekt: CESAR. <https://artemis-ia.eu/project/1-cesar.html> (Zugriffsdatum: 28.10.2019)

Projekt: CRYSTAL. <http://www.crystal-artemis.eu/> (Zugriffsdatum: 28.10.2019)

Projekt: ENABLE-S3. <https://www.enable-s3.eu/> (Zugriffsdatum: 28.10.2019)

Projekt VVMethoden - Verifikations- und Validierungsmethoden automatisierter Fahrzeuge Level 4 und 5. <http://www.tuvpt.de/index.php?id=vvmethoden> (Zugriffsdatum: 28.10.2019)

SAE International. SAE International Releases Updated Visual Chart for Its "Levels of Driving Automation" Standard for Self-Driving Vehicles. 2018. <https://www.sae.org/news/press-room/2018/12/sae-international-releases-updated-visual-chart-for-its-%E2%80%9Clevels-of-driving-automation%E2%80%9D-standard-for-self-driving-vehicles> (Zugriffsdatum: 02.12.2019)

SafeTRANS (Hrsg.). Eingebettete Systeme in der Automobilindustrie. 2015. <http://www.safe-trans-de.org/de/Aktivitaeten/Roadmapping.php> (Zugriffsdatum: 02.12.2019)

SafeTRANS (Hrsg.). Hochautomatisierte Systeme: Testen, Safety und Entwicklungsprozesse. 2017. <http://www.safetrans-de.org/de/Aktivitaeten/Roadmapping.php> (Zugriffsdatum: 02.12.2019)

SafeTRANS/ZVEI (Hrsg.). Nationale Roadmap Embedded Systems. 2009. <http://www.safetrans-de.org/de/Aktivitaeten/Roadmapping.php> (Zugriffsdatum: 02.12.2019)

SafeTRANS (Hrsg.). Safety, Security, and Certifiability of Future Man-Machine Systems - Positionspapier. 2019. <http://www.safetrans-de.org/de/Aktivitaeten/Roadmapping.php> (Zugriffsdatum: 02.12.2019)

The IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems. <https://standards.ieee.org/industry-connections/ec/autonomous-systems.html> (Zugriffsdatum: 29.10.2019)

