Entwurf! Diese Version ist noch nicht finalisiert.

Positionspapier

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

SafeTRANS
SAFETY IN TRANSPORTATION SYSTEMS

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: kras99/Adobe Stock, Seite 7: tawanlubfah/fotolia
Datum: November 2019

Positionspapier

Dieses Positionspapier dient der Abstimmung mit weiteren Industrieverbänden und wird in diesem Prozess finalisiert werden.

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

Inhalt

1	Motivation	6
2	Vom Einzelsystem zu heterogenen Kollektiven von Systemen: Beispiele	10
3	Komplexitätsdimensionen	14
4	Wo stehen wir heute? Wohin geht die Entwicklung?	24
5	Handlungsempfehlungen	30
6	Quellen	32
7	Teilnehmende Organisationen	34

1 Motivation

Sicherung der Qualität komplexer Systeme

Ob Mobilität[1], Energie[2], Gesundheit[3], Produktion[4]: Die Lösungen der Herausforderungen der Zukunft basieren auf komplexen, vernetzten, digitalen Systemen, welche Entscheidungs- und Führungsprozesse für Menschen auf der Basis von evolutionären, lernenden Verfahren treffen und zunehmend branchenübergreifende Lösungen erfordern. Gleichzeitig werden in der High-Tech Strategie der Bundesregierung die durch diese Komplexität resultierenden Risiken primär auf IT-Sicherheit reduziert, das als eigenes Querschnittsthema verankert ist[5]. Während zwar im Bereich Energie[6] und Mobilität[7] klar erkannt wird, dass über Security hinaus auch die funktionale Sicherheit (Safety) und die Systemverfügbarkeit zentrale Herausforderungen darstellen, fehlt eine Querschnittsaktivität, welche die allgemeine Herausforderung der Sicherung der Qualität komplexer Systeme adressiert. Eine domänenübergreifende systemische Betrachtung der Qualität komplexer Systeme dient dem übergeordneten Ziel „Quality Made in Germany“ als zentrales Attribut hier entwickelter, digitaler Lösungen für gesellschaftliche und wirtschaftliche Herausforderungen zu etablieren. Dies muss insbesondere für „Critical Applications“[8] gelten, in denen Ausfälle oder Cyber-Angriffe zu verheerenden Folgen für die Gesundheit von Menschen führen, drastische ökonomische oder ökologische Schäden produzieren und sogar - wie beim Verlust kritischer Infrastrukturen - das Funktionieren der Gesellschaft als Ganzes gefährden. Neben den Dimensionen Security, Safety und Availability sind hierbei auch vor dem Hintergrund der zunehmenden Autonomie solcher Systeme die Sicherung der Einhaltung von gesellschaftlich verankerten Normen und ethische Prinzipien zu berücksichtigen. Eine aktuelle Studie[9] von 84 hochrangigen Strategiepapieren kommt zu dem Ergebnis, dass hier fünf zentrale ethische Prinzipien durchgängig gefordert werden: “Our results reveal a global convergence emerging around five ethical principles (transparency, justice and fairness, non-maleficence, responsibility and privacy)”.

Sicherheit und Zertifizierbarkeit komplexer digitaler Systeme ist ein Kernanliegen von SafeTRANS[10]. Mit der „Nationalen Roadmap Eingebettete Systeme“ stellte SafeTRANS federführend gemeinsam mit BITKOM, ZVEI und VDMA eine systematische Übersicht über die durch Zusammenwachsen von Cyber-Physical Systems, dem Internet und dem Menschen entstehende Mensch-Maschine-Systeme her. Zehn Jahre nach dieser ersten umfassenden Analyse der Komplexität solcher Systeme legt SafeTRANS nach branchenspezifischen Roadmaps mit seiner Roadmap über „Safety, Security and Certifiability of Future Man-Machine Systems“ eine umfassende Analyse der Herausforderungen vor, die Qualität solcher Systeme zu sichern. Sie stellt einen systematischen qualitativen Ansatz zur Erfassung der Komplexität solcher Systeme vor und betrachtet die mit jedem Komplexitätszuwachs resultierenden Herausforderungen zur Sicherung der Qualität - in dem oben dargestellten umfassenden Sinne. Sie integriert damit als Spezialfälle Fragestellungen, wie sie beispielsweise behandelt werden in Verbundprojekten der Automobilindustrie zur Qualitätssicherung und Zulassung autonomer Fahrzeuge[11], in der Plattform Lernende Systeme[12] innerhalb der Risikobewertung des Einsatzes von KI-Methoden in medizinischen Anwendungen[13], in der Wirkungsforschung im Energieprogramm der Bundesregierung[14] oder im BMBF-Rahmenprogramm für die Geistes- und Sozialwissenschaften „Gesellschaft verstehen – Zukunft gestalten“[15]. Ein solcher ganzheitlicher Ansatz zur Qualitätssicherung führt zu Prozessen, in denen Risiken in der Absicherung und Akzeptanz solcher Systeme frühzeitig erkannt werden und Investionen in FuE zielgerichtet zum richtigen Zeitpunkt erfolgen können.

Dieses Positionspapier fasst Kernaussagen der SafeTRANS Roadmap „Safety, Security, and Certifiability of Future Man Machine Systems“ zusammen. Kern der Analyse ist die Einführung von fünf Komplexitätsdimensionen mit ihren spezifischen Komplexitätsstufen, welche im Abschnitt 3 anhand eines motivierenden Beispiels (siehe Kapitel 2) dargestellt werden. Kapitel 4 zeigt beispielhaft die Einordnung vor der Markteinführung stehender Systemlösungen in diesen mehrdimensionalen Komplexitätsraum. Kapitel 5 stellt diese Entwicklungsstufen als Schalenmodell mit ersten zeitlichen Einordnungen dar und zeigt exemplarisch, wie hohe Komplexitätsstufen durch Einschränkungen schon in naher Zukunft den Qualitätsansprüchen genügen können. Das Dokument schließt mit zehn Handlungsempfehlungen.

[1] Fortschrittsbericht zur Hightech-Strategie 2025. BMBF, Referat Grundsatzfragen von Innovation und Transfer (Hrsg.). 2019

[2] 7. Energieforschungsprogramm, BMWI (Hrsg.). 2018

[3] Rahmenprogramm Gesundheitsforschung der Bundesregierung. BMBF (Hrsg.) 2018

[4] Innovationen für die Produktion, Dienstleistung und Arbeit von morgen. BMBF, Referat – Forschung für Produktion, Dienstleistung und Arbeit (Hrsg.). 2014

[5] Forschung für die zivile Sicherheit 2018–2023, Rahmenprogramm der Bundesregierung. BMBF, Referat Sicherheitsforschung (Hrsg.). 2018/Selbstbestimmt und Sicher in der Digitalen Welt. BMBF (Hrsg.). 2015

[6] 7. Energieforschungsprogramm, BMWI (Hrsg.). 2018

[7] Siehe etwa: · Projekt PEGASUS. https://www.pegasusprojekt.de/ (Zugriffsdatum: 28.10.2019)
· Projekt VVMethoden – Verifikations- und Validierungsme-thoden 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, BMV (Hrsg.). 2019

[8] Roadmap on Safety, Security and Certifiability of Future Man-Machine Systems. SafeTRANS e.V. (Hrsg.). Voraussichtlich 2019

[9] Artifical Intelligence: The global landscape of ethics guidelines. Amon Jobin, Marcello Ieanca, Effy Vayena. Health Ethics & Policy Lab. ETH Zürich. 2019

[10] http://www.safetrans-de.org

[11] Siehe zum Besipiel: https://www.plattform-zukunft-mobilitaet.de/, AG 6 (Zugriffsdatum: 28.10.2019)

[12] https://www.plattform-lernende-systeme.de/ (Zugriffsdatum: 28.10.2019)

[13] Sichere KI-Systeme in der Medizin. Analyse des Anwendungsszenarios „Mit KI gegen Krebs“ mit Fokus auf IT-Sicherheit. Positionspapier der AG3. Lernende Systeme – Die Plattform für Künstliche Intelligenz. 2019

[14] 7. Energieforschungsprogramm, BMWI (Hrsg.). 2018

[15] Gesellschaft verstehen - Zukunft gestalten. BMBF-Rahmenprogramm für die Geistes- und Sozialwissenschaften (2019-2025). BMBF, Referat Sozial- und Geisteswissenschaften. 2019

ZITATE und QUELLENVERWEISE:

[1] Fortschrittsbericht zur Hightech-Strategie 2025. BMBF, Referat Grundsatzfragen von Innovation und Transfer (Hrsg.). 2019. Seite 40: „Die Mobilitätsbranche befindet sich in einem gewichtigen Umbruch. Sie wird nicht mehr nur durch die Fahrzeug-, sondern zunehmend auch durch die IT-Branche geprägt. Elektrofahrzeuge sind mit dem Stromnetz verbunden, sodass die netzdienliche Integration der Elektromobilität die Transformation des Verkehrsbereichs mit dem Umbau des Energiebereichs koppelt. […] Antworten auf die Herausforderungen müssen in der Gesamtschau gefunden werden. Mobilitätsbedarfe und Verkehrsbewegungen, Infrastrukturen, Beschäftigung, regionale Strukturen, technische Innovationen (bspw. das automatisierte und vernetzte Fahren) und neue Geschäftsmodelle (bspw. Ridesharing, Mobility as a service, netzdienliches Laden) müssen in einem vernetzten, digitalisierten und elektrifizierten Mobilitätsbereich gleichermaßen berücksichtigt werden."

[2] 7. Energieforschungsprogramm, BMWI (Hrsg.). 2018. Seite 32: „Die Digitalisierung und Vernetzung der Akteure sowie innovative, multimodale Mobilitätskonzepte, klimagerechtes Nutzerverhalten und Gestaltung im Quartier können dazu beitragen, den Energieverbrauch ganzheitlich zu optimieren und Fortschritte bei der Elektrifizierung des Verkehrs zu erleichtern." Seite 60: „Im Zuge der Energiewende befindet sich das Energiesystem in einem tiefgreifenden Wandel. Dezentrale Erzeugungsstrukturen, fluktuierende Einspeisung, Sektorkopplung, digitale Vernetzung oder auch neue Mobilitätskonzepte erfordern ein Umdenken auf vielen Ebenen. Diese Entwicklungen erfolgen innerhalb eines sehr komplexen Umfelds mit zahlreichen technischen, wirtschaftlichen, ökologischen, energiepolitischen und gesellschaftlichen Rahmenbedingungen. Um hier sinnvoll agieren zu können, benötigen Wirtschaft, Politik und Gesellschaft umfangreiches faktenbasiertes Orientierungswissen zu wahrscheinlichen Entwicklungspfaden des Energiesystems und deren potenziellen Auswirkungen."

[3] Rahmenprogramm Gesundheitsforschung der Bundesregierung. BMBF (Hrsg.) 2018. Seite 8: „Die Digitalisierung verändert Gesundheitsversorgung und Gesundheitsforschung grundlegend: Digitale Innovationen erlauben neue Formen der Kommunikation und Kooperation zwischen ärztlichem Fachpersonal und Patientinnen und Patienten, entlasten das medizinische Personal und können die Effizienz des Gesundheitssystems steigern." Seite 26: „Eine wesentliche Herausforderung ist die Entwicklung von interdisziplinären Systemlösungen und interaktiven Systemen, die medizinische Einzellösungen zusammenbinden und integrieren. Innovativen Konzepten der Mensch-Technik-Interaktion kommt dabei eine Schlüsselrolle zu".

[4] Innovationen für die Produktion, Dienstleistung und Arbeit von morgen. BMBF, Referat – Forschung für Produktion, Dienstleistung und Arbeit (Hrsg.). 2014. Seite 35: „Informations- und Kommunikationstechnologien (IKT) befördern die Entwicklung und Vermarktung von Dienstleistungen. IKT treibt einerseits die wirtschaftliche Bedeutung digital erbringbarer Leistungen und ermöglicht andererseits die Vernetzung von Beteiligten am Dienstleistungsprozess im globalen Maßstab". „Seit geraumer Zeit wirkt ein neuer Digitalisierungsschub. Stichwörter dafür sind vernetzte eingebettete Systeme (Cyber Physical Systems, CPS), Internet der Dienste und Dinge. Sie revolutionieren die Produktionstechnologie und Produktionsvorgänge und bieten wiederum die Basis für neue Dienstleistungen rund um die Produktion und um Sachgüter. Technologie in Form von Diensten und Problemlösungen in Form von Dienstleistungssystemen wachsen zusammen. Hochtechnologie im IKT-Bereich ermöglicht gleichermaßen die produktive Entwicklung von Dienstleistungen. Simulation, Visualisierung, integrierte Entwicklungsumgebungen, Modularisierung, Plattformstrategien und Service-Lifecycle-Management sind hier einige bedeutsame Stichworte".

[6] 7. Energieforschungsprogramm, BMWI (Hrsg.). 2018. Seite 50: „Allerdings müssen Schutz- und Leittechnik in zukünftigen dezentralen Versorgungsstrukturen jederzeit einen sicheren Netzzustand gewährleisten, Fehlersituationen zuverlässig erkennen und beherrschen. Dazu bedarf es der Erforschung neuartiger Verfahren und Komponenten, um die heute geltenden Anforderungen an Selektivität, Zuverlässigkeit und Schnelligkeit weiterhin zu erfüllen." Seite 50: „Dabei gilt es, die Kriterien für Sicherheit und Systemstabilität vor dem Hintergrund bisheriger und zukünftiger Veränderungen im Energieversorgungssystem zu überprüfen. Mögliche Instabilitäten sind zu erforschen und die Analyse- und Simulationswerkzeuge so anzupassen, dass sie der Komplexität des Gesamtsystems gerecht werden. Diese Werkzeuge unterstützen Planung und Betrieb des Netzes zur Sicherstellung systemdienlicher Interaktionen, Resilienzerhöhung und Systemoptimierung." Seite 63: „Die Umsetzung der Digitalisierung der Energiewende erfordert sowohl die Entwicklung von Sicherheitskonzepten als auch Konzepten für die Resilienz hochgradig vernetzter Systeme. Damit sollen im Aufbau wie Betrieb neuer Systeme Fehlersituationen ausgeschlossen oder deren Auswirkungen begrenzt werden, sodass sie beherrschbar bleiben."

[8] Roadmap on Safety, Security and Certifiability of Future Man-Machine Systems. SafeTRANS e.V. (Hrsg.). Voraussichtlich 2019

[14] 7. Energieforschungsprogramm, BMWI (Hrsg.). 2018. Seite 68: „Daher sollte die Wissenschaft anstreben, im transdisziplinären Diskurs technologische Anforderungen und soziale Bedürfnisse zu untersuchen und Technikfolgen transparent zu vermitteln. Dabei kann sie laborhaft, d. h. unter realitätsnahen Bedingungen, mögliche Lösungen erproben und auf ihre sozialen wie ökonomischen Wirkungen hin überprüfen. In interdisziplinärer Zusammenarbeit ist die Wissenschaft gefragt, um vorausschauend gesellschaftliche und institutionelle Zielkonflikte zu identifizieren. Sie sollte Leitbilder und Agenden anbieten, die eine koordinierte und zielwirksame Umsetzung der Energiewende erlauben. Und sie kann durch ihre Beratungsangebote die Organisationen der Gesellschaft und des Markts dabei unterstützen, regulatorische und strukturelle Maßnahmen, Business Cases und Marktstrategien zu entwickeln, die den Markteintritt von Innovationen wirksam unterstützen. Hierzu gehört selbst-verständlich auch Kommunikation und Transparenz der Wissenschaft. Die Forschenden verpflichten sich zu transparentem Vorgehen und aktiver Öffentlichkeitsarbeit, um über Fortschritte und Rückschläge zu berichten"

[15] Gesellschaft verstehen - Zukunft gestalten. BMBF-Rahmenprogramm für die Geistes- und Sozialwissenschaften (2019-2025). BMBF, Referat Sozial- und Geisteswissenschaften. 2019. Seite 7: „Entscheidend für eine gesellschaftliche Entwicklung zum Wohle der Menschen ist zudem, die Auswirkungen von Innovationen sowie die Konsequenzen der Ausrichtung auf stetige Innovationsfähigkeit in den Blick zu nehmen. Innovationen stoßen mitunter Veränderungen an, die weit über den Wirkungsbereich hinausgehen, für den sie ursprünglich entwickelt wurden. Gerade sogenannte Sprunginnovationen sowie soziale Erneuerungsprozesse, die viele Lebensbereiche berühren, können mittel- und langfristig weitreichende Veränderungen bestehender gesellschaftlicher Strukturen, sozialkultureller Praktiken und normativer Ordnungen herbeiführen. Was bedeutet es etwa für gesellschaftliches Miteinander, politische Systeme oder die Autonomie von Personen, wenn Künstliche Intelligenz künftig menschliche Fähigkeiten und Entscheidungen ersetzen sollte? Nicht minder weitreichend können die Wirkungen sein, wenn Strukturen, Institutionen und Ordnungen einer Gesellschaft auf das Ziel beständiger Innovationsfähigkeit ausgerichtet und dabei tradierte Werte, Normen und Identitäten infrage gestellt werden".

2 Vom Einzelsystem zu heterogenen Kollektiven von Systemen: Beispiele

Die in diesem Positionspapier diskutierten 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 Fahrzeugverbünde 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. Die dominanten Charakteristika und ihre signifikanten einschränkenden Annahmen seien hier kurz skizziert und in die Komplexitätsdimensionen eingeordnet.

Beispiel: Autonom fahrendes Fahrzeug

Beschreibung

Ein Fahrzeug, dass SAE-Level 4/SAE-Level 5[1] Autonomie realisiert. Auf Basis fahrzeugeigener Sensorik erfasst es seinen relevanten Kontext und trifft in urbanen Fahrsituationen angemessene Manöverentscheidungen.

Herausforderungen

- Erstellung eines verlässlichen, vollständigen Lagebilds mit existierender Sensortechnologie
- belastbare Argumentation, dass das Fahrzeug in allen bekannten und partiell bekannten Situationen angemessene Manöverentscheidungen trifft und durchführt
- Sicherheitsnachweis, dass von diesem System keine unangemessene Gefahr ausgeht - das schließt sowohl funktionale Fehler als auch Angriffe auf das System ein

Komplexitätsreduzierende Einschränkungen

- 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

[1] Erklärung der Stufen der Fahrautomatisierung siehe: 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: 25.11.2019)

Beispiel: Verkehrskolonne (Platoon)

Beschreibung

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

Herausforderungen

- dynamische Veränderung der Struktur der Gruppe über die Zeit
- kooperatives Erstellen eines vollständiges Lagebilds aus den Einzelbildern der Gruppenteilnehmer
- dynamische Missionsplanung, Ziele-, Strategie- und Manöverabstimmung
- Beherrschung aller möglichen auftretenden Fahrsituationen
- dynamische Koordination vieler komplexer Manöver im kooperativen Verbund
- Sicherstellen und überwachen der Integrität der Gruppe mit Strukturdynamik

Komplexitätsreduzierende Einschränkungen

- 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

Beschreibung

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.

Herausforderungen

- Erkennen und Beurteilen der aktuellen Verkehrslage
- Vorhersagen der zukünftigen Entwicklung durch externe und systeminterne Interventionen
- Dynamik der Systemstruktur in einem Verkehrsabschnitt verstehen und durch geeignete Maßnahmen steuern
- kooperatives Aushandeln von QoS für einzelne Teilnehmer, Gruppen, Kollektive
- Beherrschung von gemischtem Verkehr mit autonomen und individuell geführten Fahrzeugen

Komplexitätsreduzierende Einschränkungen

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

Beispiel: Smart City

Beschreibung

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.

Herausforderungen

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

Komplexitätsreduzierende Einschränkungen

- 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 CO_2-Preis)

Diese vier Anwendungsbeispiele veranschaulichen vier Aggregationsstufen, die in allen Anwendungen komplexer, digitaler Systeme identifiziert werden können:

1. Einzelne (Mensch-Maschine)-Systeme, wie hochautonome Fahrzeuge, Züge, Flugzeuge, Schiffe.

2. Gruppen von Mensch-Maschine-Systemen, 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 dem betreuenden Personal).

3. Homogene Kollektive von Human-Cyber-Physical Systems, welche durch die Führung des Kollektivs eine optimale Nutzung von (Anwendungsklassen spezifischer) Ressourcen ermöglichen und den Teilnehmern des Kollektivs Dienste anbieten (wie etwa Verkehrsleitplanung, Energieverbundnetze). Alle Teilnehmer des Systems gehören einer Anwendungsdomäne an.

4. Heterogene Kollektive von Human-Cyber-Physical Systems integrieren Teilnehmer verschiedener Anwendungsdomänen in ein Gesamtsystem, um durch dessen Führung eine holistische domänenübergreifende Resourcennutzung und die Einhaltung übergeordneter Vorgaben (wie etwa zur Erreichung der Klimaziele der Bundesregierung) zur ermöglichen (wie etwa Smart Cities oder Krisenmanagement-Verbünde).

3 Komplexitätsdimensionen

Wie können wir die Komplexität digitaler Mensch-Maschine-Systeme bewerten?
Wie einen Zugang zu notwendigen Maßnahmen zur Qualitätssicherung gewinnen?

Fünf Komplexitätsdimensionen

Wir stellen in diesem Abschnitt die Ergebnisse einer Analyse zahlreicher realer, komplexer Systeme vor, die uns erlaubt, einen branchenübergreifenden Begriff der Essenz dessen zu präzisieren, was die Komplexität solcher Systeme ausmacht. Wir identifizieren fünf Dimensionen - teilweise mit Untergruppen - mit jeweils dimensionsspezifischen Skalen zur Bewertung der Komplexität, die in dem nachfolgenden Schaubild dargestellt sind. 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. Gemeinsam ist allen diesen Systemausprägungen:

Relative Systemkomplexität

- die Abgrenzung zwischen dem Egosystem und einem umgebenden Kontext, der unterschiedlich komplex gestaltet sein kann und
- der Fähigkeit, diesen Kontext über eine Systemschnittstelle „wahrzunehmen“, also eine digitale oder mentale Repräsentation des aktuellen Kontextes zu erstellen.

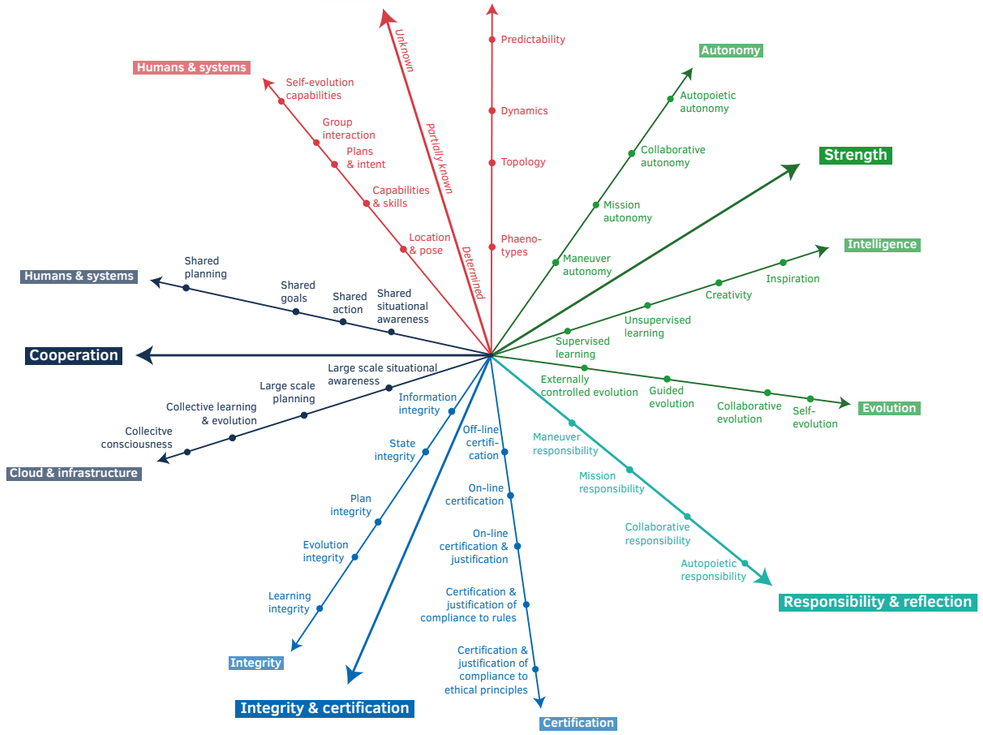


Abb. 1: Dimensionsspezifische Skalen zur Bewertung der Komplexität von Mensch-Maschine-Systemen

Die Erstellung einer digitalen oder mentalen Repräsentation erfolgt in der Regel unter Berücksichtigung der erwarteten Weiterentwicklung des wahrgenommenen Kontextes. Das System „analysiert“ den Kontext in Bezug auf die Durchsetzbarkeit der Egoziele und leitet daraus Handlungsstrategien ab um die Egoziele umzusetzen[1].

In der ersten Komplexitätsdimension 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.

Strength: Autonomy

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 unbekannten Kontext selbst erschließen und darin allein oder kooperativ handeln.

Strength: Intelligence

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

[1] Nationale Roadmap Embedded Systems. SafeTRANS, ZVEI (Hrsg.). 2009

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 unbekannten Kontext seine Ziele erreichen kann.

Strength: Evolution

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).

Ein Highway-Chauffeur kann in eingeschränkten Umfeldbedingungen (Autobahnfahrt, bestimmten Bedingungen an Witterung und Zustand der Fahrbahn) aufgrund seiner Autonomiefähigkeiten ein Fahrzeug alleine führen und ist dabei in der Lage sich an unterschiedliche Verkehrflusssituationen auf der Autobahn automatisch anzupassen. Dabei nutzt er erlernte Fähigkeiten zur Umgebungswahrnehmung und Umgebungsklassifikation, um sichere und ökologisch optimale Fahrzeugführungsstrategien zu bestimmen. SAE Level 5-Fahrzeuge können diese Führungsaufgabe vollständig autonom ohne Rückgriff auf den Menschen auch in beliebigen Verkehrssituationen übernehmen, gegebenenfalls unter Ausnutzung von Informationen, die über weitere technische Systeme wie Infrastruktur und Cloud gewonnen werden. Dazu muss ein Level 5-Fahrzeug über autopoeitische Autonomie, Kreativität und die Fähigkeit zur Selbstevolution verfügen.

In der zweiten 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.

Zum Kontext eines Highway-Chauffers zählen die umgebenden Fahrzeuge samt deren relativer Position, relativer Geschwindigkeit, relativer Beschleunigung, Markierungslinien, Verkehrszeichen, Hindernisse auf der Fahrbahn, Witterungsbedingungen, Fahrbahnzustand, Objekte und Elemente der Umgebung, welche zu fehlerhaften Umgebungswahrnehmungen führen können (z. B. in Autobahnbrücken verbautes Metall, Baustellen).

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 konfliktierend 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 opportunieren; 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 unbekannten 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.

Context: Physical environment

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.

Context: Humans and systems

Ausbaustufen Kenntnis Menschen & 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.

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 unterscheiden die folgenden Ausbaustufen von Kooperationsformen, deren Gemeinsamkeit darin besteht, dass die potenziellen Kooperationspartner das Egosystem tendentiell unterstützen.

Cooperation: Humans and systems

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.

3. Geteilte Ziele: Die Kooperationspartner stimmen sich ab, welche Ziele gemeinsam verfolgt werden.

4. Geteilte Pläne: Die Kooperationspartner stimmen ihre Pläne und Strategien zur Erreichung dieser Ziele ab.

In einem Krisenmanagement-System zur Bekämpfung eines sich schnell ausbreitenden Waldbrandes, der unmittelbar Wohngebiete bedroht, bilden Feuerwehr, Polizei, Noteinsatzzentrale für Rettungsfahrzeuge für den Zeitraum der Bekämpfung eine Kooperation, in der die jeweils lokal wahrgenommenen Lagebilder ständig über Funkkontakt ausgetauscht werden und im Krisenzentrum zu einem virtuellen Lagebild integriert werden. Unter Kenntnis der aktuellen und prognostizierten Windsituation werden dort mögliche Szenarien in der Weiterentwicklung des Brandes evaluiert. Die Analyse ergibt, dass die vor Ort vorhandenen Fähigkeiten zur Brandbekämpfung nicht ausreichen und deswegen zur Bekämpfung aus der Luft von einem Nato-Partner bereitgestellte Löschflugzeuge mit ihrem Personal integriert werden müssen. Aus dieser Analyse werden abgestimmte Pläne für zu evakuierende Gebiete und dafür verwendete Straßen verwendet, Schwerpunkte der Brandbekämpfung gesetzt und der Einsatz der Löschflugzeuge an besonders kritischen Gebieten geplant. Diese Pläne werden heruntergebrochen in Ziele für die verschiedenen Teams, die unter Kenntnis der globalen Planung koordinierte Aktionen zur Reduktion der Kritikalität der Situation umsetzen.

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 Verkehrsflussoptimierungsbeispiel 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.

Cooperation: Cloud and infrastructure

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.

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 Dilemmasituationen[2] sowie ethische Richtlinien für den Aufbau autonomer Systeme[3]). 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. Wir schlagen vor, in dieser Komplexitätsdimension die Ausbaustufen nach der zunehmenden Komplexität der Aufgaben zu ordnen, für die in diesem Sinne Verantwortung übernommen wird.

Responsibility and reflection

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 vorgenommen Handlungen zu übernehmen.

4. Autopoietische Verantwortung: Das System kann sich nicht nur einen unbekannten Kontext selbst erschließen und darin alleine oder kooperativ handeln, es besitzt darüber hinaus die Fähigkeit, für diese Handlungen Verantwortung zu übernehmen.

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 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, mit vorhandenen Ausweichstrategien übereinstimmen, die mit der Systemverschlechterung fertig werden. Die folgenden Ausbaustufen zur Gewährleistung der Systemintegrität sind nötig:

Integrity and certification

Ausbaustufen Systemintegrität

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, und sich 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).

[2] Automatisiertes und Vernetztes Fahren. BMVI. Ethik-Kommission (Hrsg.). 2017

[3] 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)

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 Security) und Verfügbarkeit solcher Systeme sowie die Respektierung gegebener etablierter gesellschaftlicher, rechtlicher und ethischer Rahmenbedingungen sind in die Qualitätsbewertung einzubeziehen. Letzteres setzt normative Reflektionsfähigkeiten voraus, die in der vorigen Komplexitätsdimension zu Verantwortlichkeit dargestellt wurden.

Auf europäischer Ebene wird dies bereits gefordert in den *ICT Standardisation Priorities for the Digital Single Market*[4]. Die VDA Leitinitiative[5] koordiniert gemeinsame FuE-Aktivitäten der Automobilindustrie zur Qualititätssicherung hochautomatisierter Fahrzeuge. In Rahmen der Plattform *Lernende Systeme* werden gegenwärtig Risikostufen für KI-basierte Systeme entwickelt, um dem jeweiligen Risiko angemessene Maßnahmen zur Qualitätssicherung vorzuschlagen[6]. 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 haben[7].

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 dieses Positionspapiers. 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.

Ausbaustufen Absicherung

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 Selbstzertifzierung 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 Selbstzertifzierung 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 Selbstzertifzierung verwendet werden durch die zuständige regulative Behörde (off-line) zertifiziert wurden.

[4] ICT Standardisation Priorities for the Digital Single Market. European Commission (Hrsg.), 2016
Electronic Components and Systems Strategic Research Agenda (ECS SRA). AENEAS, ARTEMIS-IA. EPoSS (Hrsg.). 2019.
Seite 192: „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 193: „Safety aspects have a major impact in case of public knowledge of accidents due to technical failure.“ Seite 194: „ 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.

[5] Siehe etwa: · Projekt PEGASUS. https://www.pegasusprojekt.de/ (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)
· Aktionsplan Forschung für autonomes Fahren. Ein übergreifender Forschungsrahmen von BMBF, BMWi und BMVI. BMBF, BMWi, BMVI (Hrsg.). 2019

[6] 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

[7] Ethische Grundsätze wie bereits in der Einleitung zitiert.

4 Wo stehen wir heute? Wohin geht die Entwicklung?

Wo stehen wir heute, welche Entwicklung werden zukünftige Mensch-Maschine-Systeme in dem durch die Komplexitätsdimensionen aufgespannten Entwurfsraum nehmen?

Abbildung 2 ordnet die vier Anwendungsbeispiele aus Kapitel 2 - hochautonomes Fahrzeug, Lastwagen Platoon, Verkehrsleitsystem, Smart City - in die Komplexitätsdimension ein. Sie zeigen auf, wie schon in naher Zukunft durch geschickt gewählte Einschränkungen des Kontextes Systeme hoher Komplexität beherrschbar werden können.

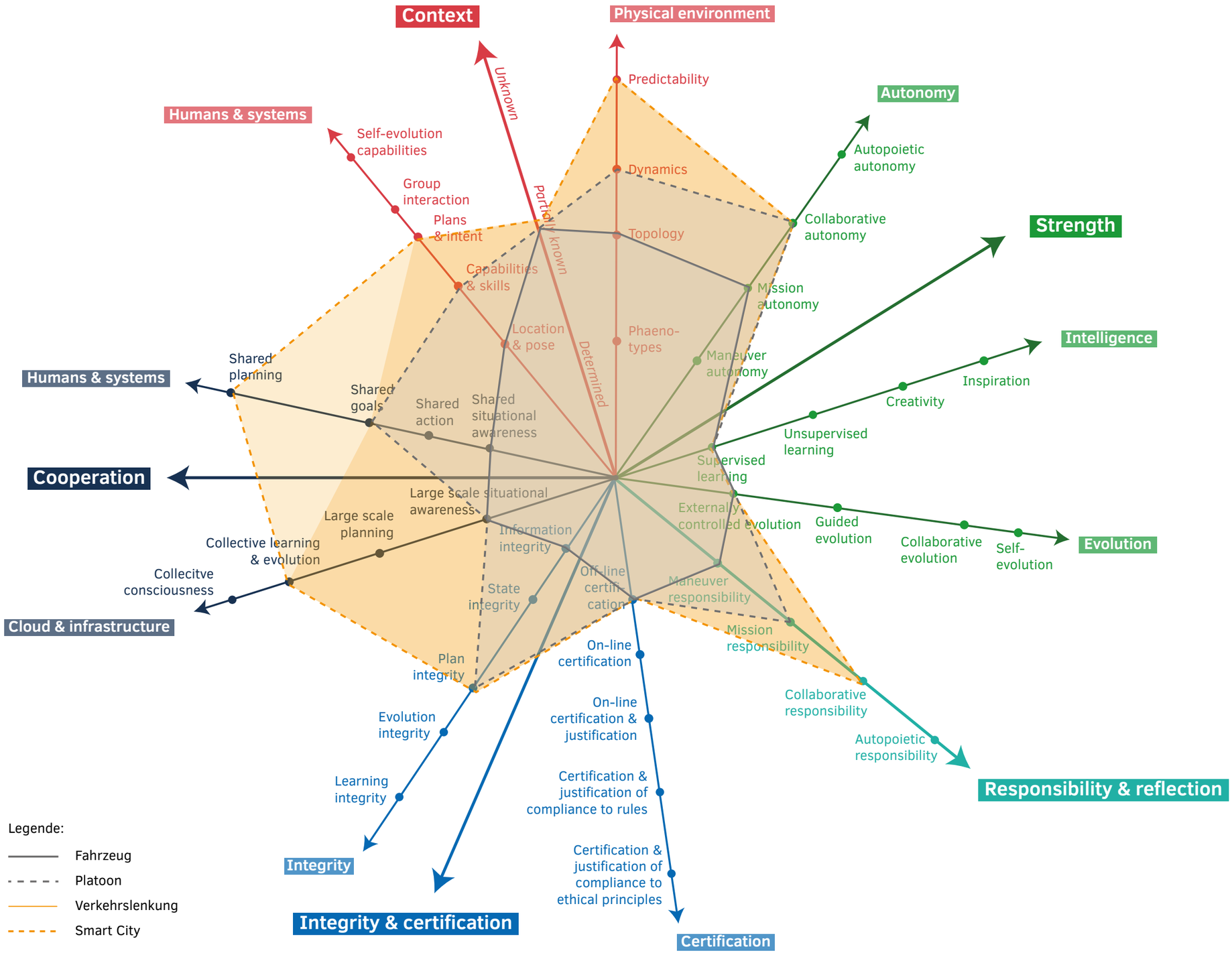


Abb. 2: Wachsende Systemkomplexität der Systemklassen am Beispiel von Mobilitätssystemen

Das erste Beispiel - hochautomatisiertes Fahren - erreicht im Zeitraum bis 2027 den hohen Überdeckungsgrad in den Dimensionen Autonomie, Erkennung und Prädiktion des Kontextes nur für eingeschränkte Kontexte (Fahren in eingeschränkten Umgebungen wie Logistikhöfen, Flughäfen und ähnliches, Fahren mit eingeschränkter Fahrdynamik, geringer Geschwindigkeit, reduzierten Manövern, defensiver Fahrstrategie) und nutzt dabei sowohl durch Infrastruktur wie durch andere Fahrzeuge über Car2X-Kommunikation vermittelte Unterstützung zur Erstellung des situativen Lagebildes. Es sichert die Integrität der dafür verwendeten Information. Die Typabsicherung erfolgt off-line, insbesondere auf der Basis von gegenwärtig entwickelten Verfahren zur virtuellen Absicherung und Szenarienkatalogen. Das System reflektiert zum Teil die Auswirkungen der Auswahl seiner Manöver auf andere Verkehrsteilnehmer.

Auch im zweiten Beispiel - eine Gruppe von Systemen, hier: ein Truck-Platoon - werden komplexitätsreduzierende Maßnahmen ausgenutzt, um einen vergleichsweise hohen Überdeckungsgrad der Komplexitätsdimensionen bis 2027 zu erreichen. Kollaborative Autonomie wird dadurch unterstützt, dass eine Reduktion auf einfache Manöver erfolgt und die Integration in einen Platoon nur dann erfolgt, wenn das neu aufgenommene Fahrzeug in seinen Fahrdynamikaspekten sich nicht signifikant unterscheidet von den anderen Fahrzeugen. Die Entscheidungskompetenz zur Auswahl von Manövern liegt nur beim Führungsfahrzeug. Eine Reduktion des Kontextes erfolgt durch Beschränkung auf vorher bekannte und vermessene Strecken. Die Infrastruktur unterstützt bei der Erstellung des Lagebildes. Von anderen Platoons sind Zielpunkte und Pläne bekannt. Auf der Basis dieser Komplexitätsreduktionen kann ein Platoon gemeinsame Ziele (shared goals) wie gemeinsame Route, Energieeffizienz oder Emissionsreduktion durch aufeinander abgestimmte Handlungen der Teilfahrzeuge vornehmen (shared action). Die Absicherung erfolgt off-line. Das System bezieht in der Umsetzung einer Mission die Auswirkungen auf die übergeordneten Ziele und andere Verkehrsteilnehmer ein.

Im dritten Beispiel - ein homogenes Kollektiv von Systemen, hier: Verkehrslenkung - wird die Erreichung übergeordneter Ziele (Emissionsreduktion, Optimierung des Verkehrsdurchsatzes, vorrangige Behandlung von Notfallfahrzeugen) durch komplexitätsreduzierende Maßnahmen unterstützt: Die teilnehmenden Fahrzeuge müssen sich durch die zentral gegebenen Anweisungen zur Erzielung der übergeordneten Ziele unterordnen, Verkehrsabschnitte werden temporär für prioritäre Teilnehmer (Rettungsfahrzeuge, ÖPNV) reserviert. Die Infrastruktur liefert aktuelle Informationen zur Verkehrslage und unterstützt damit eine vorausschauende, verantwortliche Planung der Verkehrslenkung.

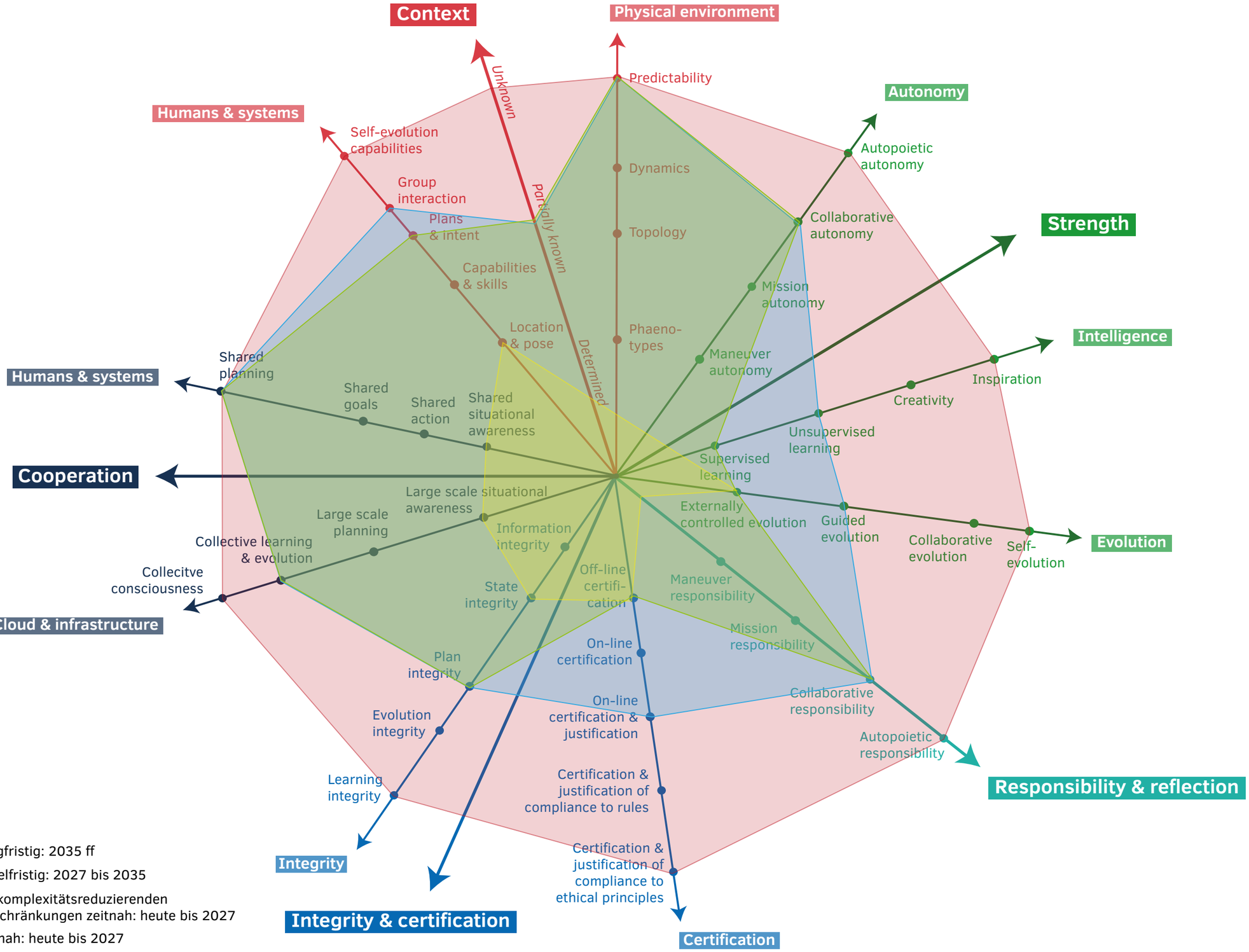
Als Beispiel von heterogenen Kollektiven von Systemen betrachten wir die Einbettung des Smart City-Beispiels in die Komplexitätsdimensionen. Hier gilt es, die möglichweise vorhandenen Konflikte zwischen Zielen der Teildomänen (Smart Energy, Smart Mobility, Smart Health, Smart Crisis Management) durch situationsabhängige Priorisierung in jeweils konfliktfreie Mengen von Teilzielen aufzulösen und dabei über die Zeit Akzeptanz und Zustimmung aller Teildomänen sicherzustellen. Von daher gilt es abzuwägen, ob hierarchische Top-down Entscheidungen, welche helfen, die Gesamtkomplexität zu reduzieren, mittelfristig durch kooperative Verhandlungslösungen ersetzt werden können.

Die Ausbaustufen in jeder Komplexitätsdimension zeigen mögliche Evolutionsstufen auf, ohne allerdings zu suggerieren, dass dieses Wachstum homogen über alle Komplexitätsdimension erfolgen wird. Während wir erwarten, dass Kooperationsfähigkeiten und die daraus resultierenden Möglichkeiten zur Beherrschung der Komplexität des Kontextes relativ schnell zunehmen, erwarten wir, dass die Systemeigenschaften Intelligenz, Evolution, Integrität und Zertifizierung sowie Verantwortung sich langsamer entwickeln. Allerdings wird es, wie in den vier Beispielen in Kapitel 2 aufgezeigt, gleichzeitig möglich sein, durch Einschränkungen der Komplexität des Kontextes durch bauliche, organisatorische oder regulative Maßnahmen hohe Grade an Absicherung bei hohen Stufen von Autonomie oder Intelligenz zu erreichen.

Die in so restringierten Kontexten gewonnenen Erfahrungen bieten sich als Lernumgebungen an, in denen schrittweise die Kontextanforderungen gelockert werden können, um so Gesamtlösungen zu erreichen, die als Zielbilder schon heute identifiziert worden sind (wie z. B. vollständig autonomes Fahren - SAE Level 5 - in beliebigen Umgebungen). Auf der Basis einer Analyse der für die einzelnen Ausbaustufen erwarteten Technologieinnovationen[1] gehen wir zeitnah (bis 2027), mittelfristig (2027 bis 2035) und langfristig (nach 2035) von vier Stufen der Überdeckung aus, wie sie Abbildung 3 darstellt. Dabei unterscheiden wir im zeitnahen Bereich solche Anwendungen, die nur durch Reduktion der Komplexität des Kontextes diese Ausbaugrade erreichen.

[1] Für eine detaillierte Darstellung der Ausbaustufen der erwarteten Technologieinnovationen siehe das ausführliche Roadmap-Dokument zu diesem Positionspapier: Roadmap on Safety, Security and Certifiability of Future Man-Machine Systems. SafeTRANS e.V. (Hrsg.). Voraussichtlich 2019

Abb. 3: Zeitliche Vorhersage zum Erreichen der Fähigkeiten der Systemklassen



Legende:

- Langfristig: 2035 ff
- Mittelfristig: 2027 bis 2035
- Mit komplexitätsreduzierenden Einschränkungen zeitnah: heute bis 2027
- Zeitnah: heute bis 2027

5 Handlungsempfehlungen

Quality Made in Germany

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 Collobaration 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

6 Quellen

Aktionsplan Forschung für autonomes Fahren. Ein übergreifender Forschungsrahmen von BMBF, BMWi und BMVI . BMBF, BMWi, BMVI (Hrsg.). 2019

Artifical Intelligence: The global landscape of ethics guidelines. Amon Jobin, Marcello Ieanca, Effy Vayena, Health Ethics & Policy Lab. ETH Zürich. 2019

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

7. Energieforschungsprogramm, BMWI (Hrsg.). 2018

Forschung für die zivile Sicherheit 2018–2023, Rahmenprogramm der Bundesregierung. BMBF, Referat Sicherheitsforschung (Hrsg.). 2018

Electronic Components and Systems Strategic Research Agenda (ECS SRA). AENEAS, ARTEMIS-IA. EPoSS (Hrsg.). 2019

Fortschrittsbericht zur Hightech-Strategie 2025. BMBF, Referat Grundsatzfragen von Innovation und Transfer (Hrsg.). 2019

Gesellschaft verstehen - Zukunft gestalten. BMBF-Rahmenprogramm für die Geistes- und Sozialwissenschaften (2019-2025). BMBF, Referat Sozial- und Geisteswissenschaften. 2019

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

Innovationen für die Produktion, Dienstleistung und Arbeit von morgen. BMBF, Referat – Forschung für Produktion, Dienstleistung und Arbeit (Hrsg.). 2014

Nationale Roadmap Embedded Systems. SafeTRANS, ZVEI (Hrsg.). 2009

Projekt PEGASUS. Förderung: BMWi. https://www.pegasusprojekt.de/de/ (Zugriffsdatum: 28.10.2019)

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

Rahmenprogramm Gesundheitsforschung der Bundesregierung. BMBF (Hrsg.). 2018

Roadmap on Safety, Security and Certifiability of Future Man-Machine Systems. SafeTRANS e.V. (Hrsg.). Voraussichtlich 2019

SAE International Releases Updated Visual Chart for Its “Levels of Driving Automation” Standard for Self-Driving Vehicles. SAE International. 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: 25.11.2019)

Selbstbestimmt und Sicher in der Digitalen Welt. BMBF (Hrsg.). 2015

Sichere KI-Systeme in der Medizin. Analyse des Anwendungsszenarios „Mit KI gegen Krebs“ mit Fokus auf IT-Sicherheit. Positionspapier der AG3. Lernende Systeme – Die Plattform für Künstliche Intelligenz. 2019

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

7 Teilnehmende Organisationen

- Carl von Ossietzky Universität Oldenburg
- Deutsches Forschungszentrum für Künstliche Intelligenz GmbH (DFKI)
- Deutsches Zentrum für Luft- und Raumfahrt (DLR)
- Fraunhofer-Institut für Experimentelles Software Engineering IESE
- FZI Forschungszentrum Informatik
- INGenX Techbnologies GmbH
- ITK Engineering GmbH
- OFFIS Institut für Informatik
- Robert Bosch GmbH
- SafeTRANS
- Siemens AG
- Siemens Mobility GmbH
- Universität Bremen
- Universität Hamm-Lippstadt

Die Erstellung dieses Positionspapiers und der dazugehörigen Roadmap „Safety, Security, and Certifiability of Future Man-Machine Systems“ erfolgte durch einen SafeTRANS-Arbeitskreis unter Beteiligung von Vertretern der oben genannten Organisationen und unter der Moderation von Prof. Dr. Werner Damm, SafeTRANS, und Peter Heidl, Robert Bosch GmbH.

www.safetrans-de.org