KI und Daten – Herausforderungen auf dem Weg zum autonomen Fahren

Workshopreihe zum Förderprogramm „Zukunftsinvestitionen Fahrzeughersteller und Zulieferindustrie“ (Konjunkturpaket 35c)

BMWi, Berlin, 04. 12. 2020

Netiquette

- Wenn es Ihr Set-Up erlaubt, schalten Sie sich bitte über einen Computer in das Meeting ein, um der Präsentation folgen und die Tools (Chat) nutzen zu können.
- **Ihr Mikrofon wurde mit dem Beitritt stummgeschaltet**. Sie können die Stummschaltung aktuell nicht selbstständig deaktivieren.
- **Fragen** stellen Sie bitte über **den Chat mit dem vorangestellten Wort „Frage“**. Während der Abschnitte „Fragen und Antworten“ liest die Moderation Fragen vor und diese werden beantwortet.

Agenda

Moderation: Matthias Marx (BMWi) & Philipp Wolf (TÜV Rheinland)

1. Begrüßung und Vorstellung	*08:00 – 08:15 Uhr*
2. Autonomy@Scale – Flexible und erweiterbare KI	*08:15 – 08:35 Uhr*
Fragen und Antworten	*08:35 – 08:45 Uhr*
3. Data-driven Development – Wie kommen wir zu der Datenbasis, die wir brauchen?	*08:45 – 09:05 Uhr*
Fragen und Antworten	*09:05 – 09:15 Uhr*
4. Podiumsdiskussion	*09:15 – 09:45 Uhr*
5. Abschluss der Veranstaltung	*09:45 – 10:00 Uhr*

Überblick 35c-Workshopreihe

Mi., 02.12.	Do., 03.12.	Fr., 04.12.
• 10:00-12:00 Uhr • **Darstellung Gesamtkonzept des Förderrahmens "35c" und Bedeutung regionaler Innovationscluster** • 13:00-15:00 Uhr • **Souveräne Datenökosysteme für die Automobilindustrie**	• 10:00-12:00 Uhr • **Innovativer Bahnverkehr - digital und nachhaltig auf der Schiene** • 13:00-15:00 Uhr • **GAIA-X als Basis für zuverlässige automatisierte und vernetzte Mobilität**	• 08:00-10:00 Uhr • **KI und Daten – Herausforderungen auf dem Weg zum autonomen Fahren** • 10:30-12:30 Uhr • **Kollaborative Nutzung von Daten in der Produktion** • 13:00-15:00 Uhr • **Production System Lifecycle Management**

Ziele

1. **Information über den geplanten Förderrahmen**
2. **Beleuchtung transformationsrelevanter Themen sowie**
3. **Vernetzung der Interessenten/innen zur Bildung von Projektverbünden**

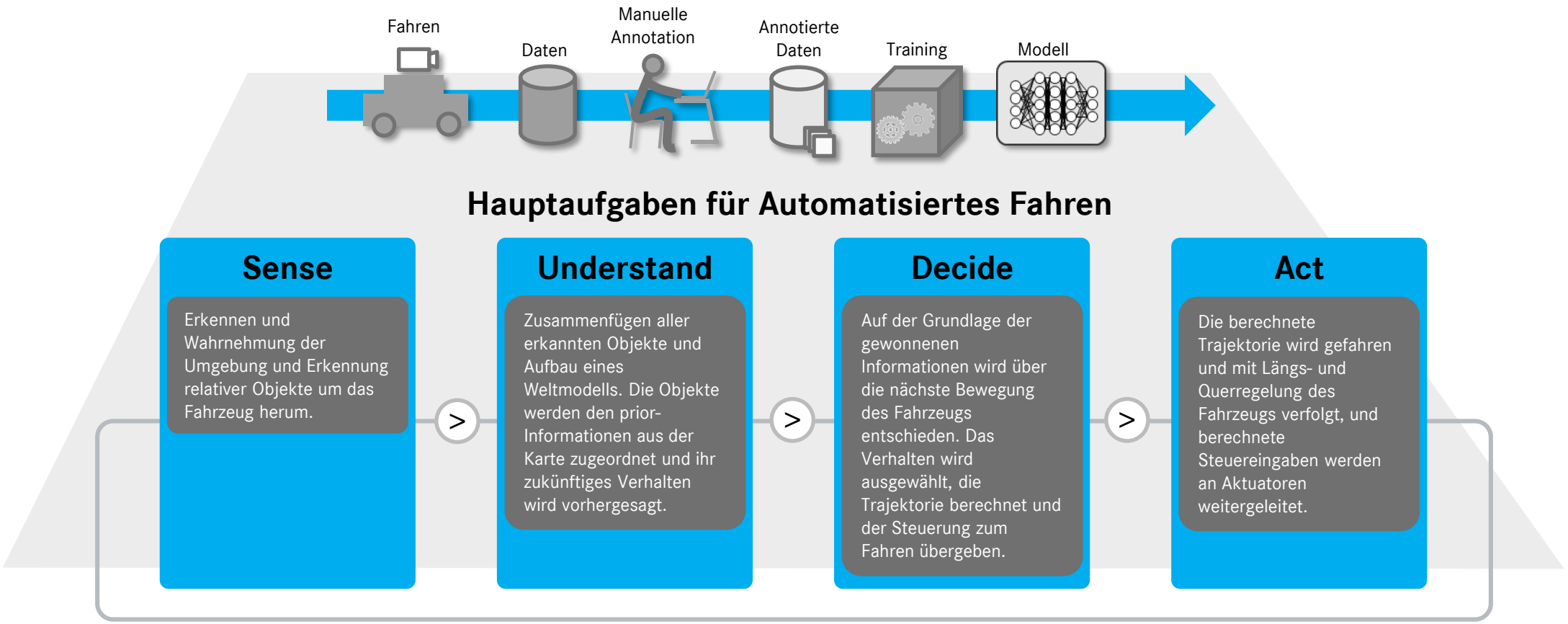
Autonomy@Scale –
Flexible und erweiterbare KI

Autonomy@Scale – Flexible und erweiterbare KI

Mohsen Sefati

Mercedes-Benz
Das Beste oder nichts.

KI und maschinelles Lernen im Kontext Automatisiertes Fahren

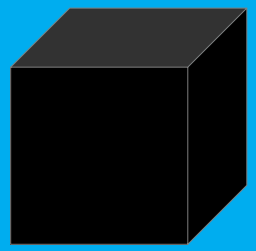


KI und Maschinelles Lernen können in allen oben genannten vier Hauptaufgaben für bessere Leistungen eingesetzt werden. Vor allem in Perzeption und Prädiktion werden hervorragende Ergebnisse erzielt.

KI und Maschinelles Lernen ist ein großartiges Werkzeug neben den immer noch herausfordernden Absicherungsfragen

KI mit Vorsicht genießen!

Back Box Natur

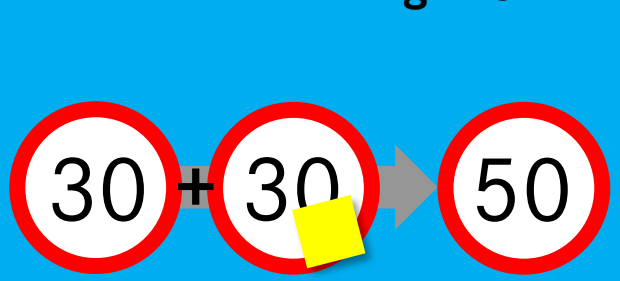


- ML-Modelle sind meist **intransparent** und ihre Berechnungen sind **nicht immer interpretierbar**.
- Das Verhalten der ML-Modelle in noch nicht erlernten Situationen ist **nicht immer vorhersehbar**.

Vergessen

- Durch das Erlernen neuer Informationen kann es passieren, dass die ML bereits erlernte Informationen abrupt vergessen. Dieser Effekt wird als **katastrophales Vergessen** bezeichnet.

Angriffe



- Die Funktionalität von MLs könnte durch absichtlich entworfenen Input angegriffen werden. Dies wird als **Adversarial Attack** bezeichnet.
- Die Funktionalität von MLs kann durch einen solchen Angriff auch vollständig gestört und gestoppt werden.

[1], [2]

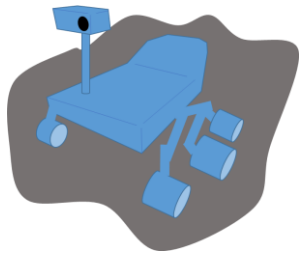
[1] Fooling aReal Car with Adversarial Traffic Signs, Morgulis et al., 2019
[2] Farbfleck könnte autonom fahrende Fahrzeuge verwirren, Behringer, CyberValley, 2019

Dies ist nicht das erste Mal, dass wir uns mit einem komplexen KI-unterstützten System befassen

Warum ist Automatisiertes Fahren so herausfordernd?

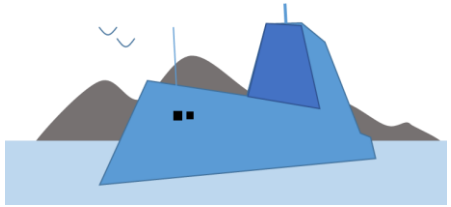
Sense | **Understand** | **Decide** | **Act**

Mars Rover Curiosity



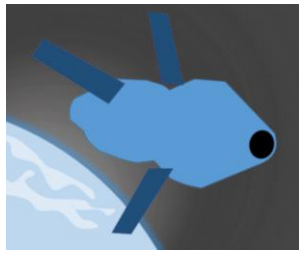
Autogroßer Rover zur Erkundung des Mars. Er wurde 2011 eingeführt und ist nach 560 Millionen km gelandet. Immer noch funktionsfähig und einsatzbereit.

USS Zumwalt



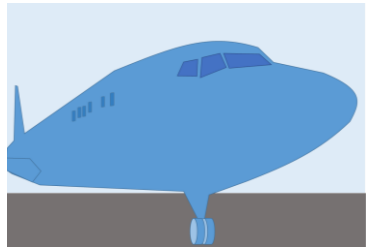
soll die Bodentruppen bei Landangriffen unterstützen.
Er wurde 2013 eingeführt und 2020 geliefert.

NASA Orion Spaceship



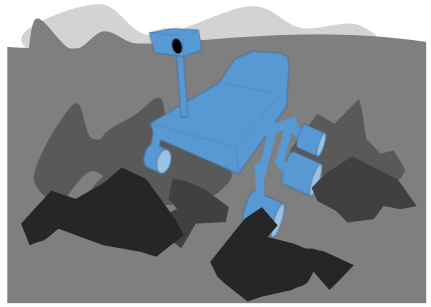
Kann eine Besatzung von sechs Personen jenseits der niedrigen Erdumlaufbahn unterstützen und bis zu 21 Tage abgekoppelt und bis zu sechs Monate angekoppelt bleiben.

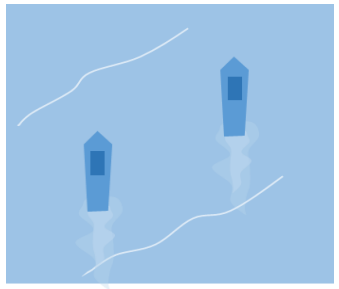
Boeing 737 Max



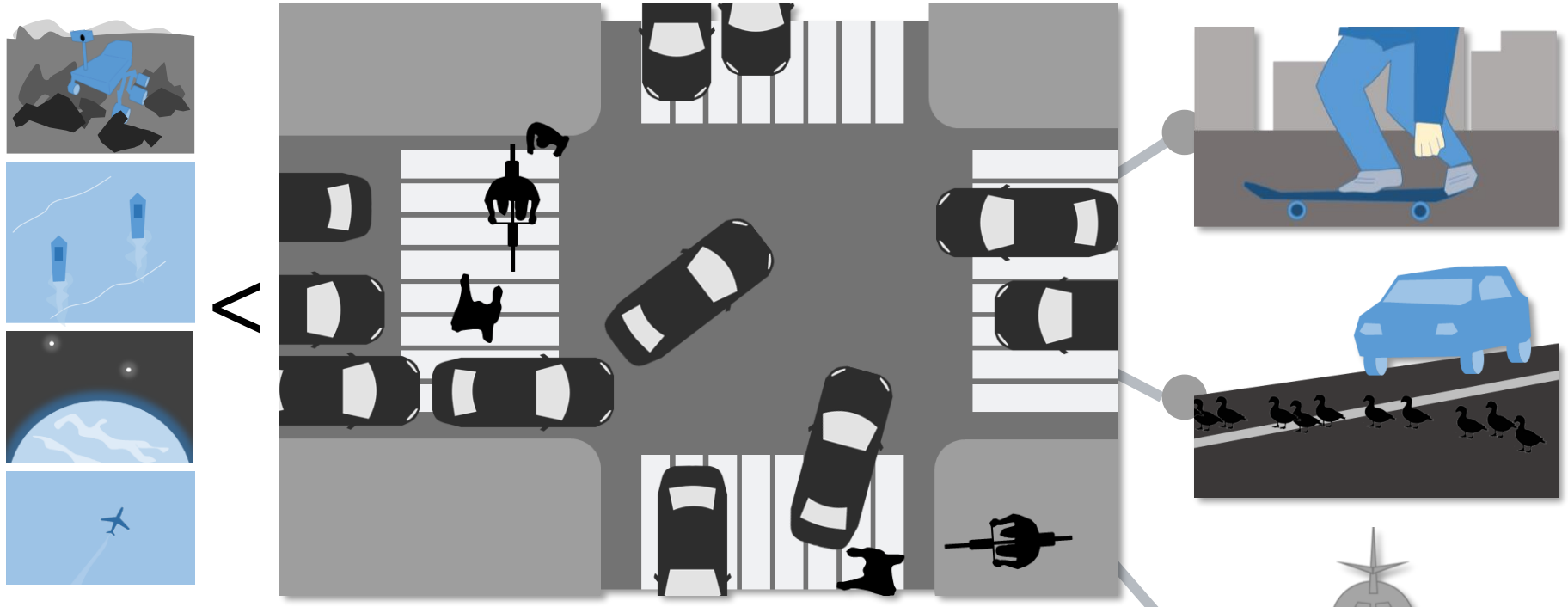
Ausgelegt für den Transport von bis zu 200 Passagieren über eine Entfernung von 6.5000 km.

Navigations-Umgebungen





Automatisiertes Fahren ist ein komplexes System in einer sehr komplexen Umgebung

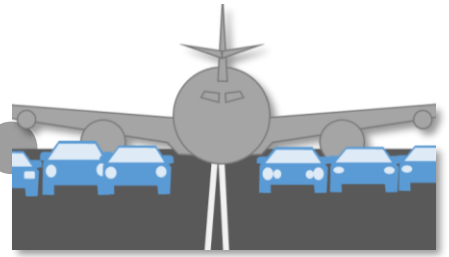


Open World Charakteristik

- Die Verkehrsteilnehmer können in jeder Form mit jeder Dynamik und jedem Verhalten auftreten.
- Es gibt unzählige Kombinationen von Szenarien, die passieren können.
- Es gibt keine Garantie, dass wir vollständige Informationen über die Umwelt entdeckt haben oder entdecken werden.

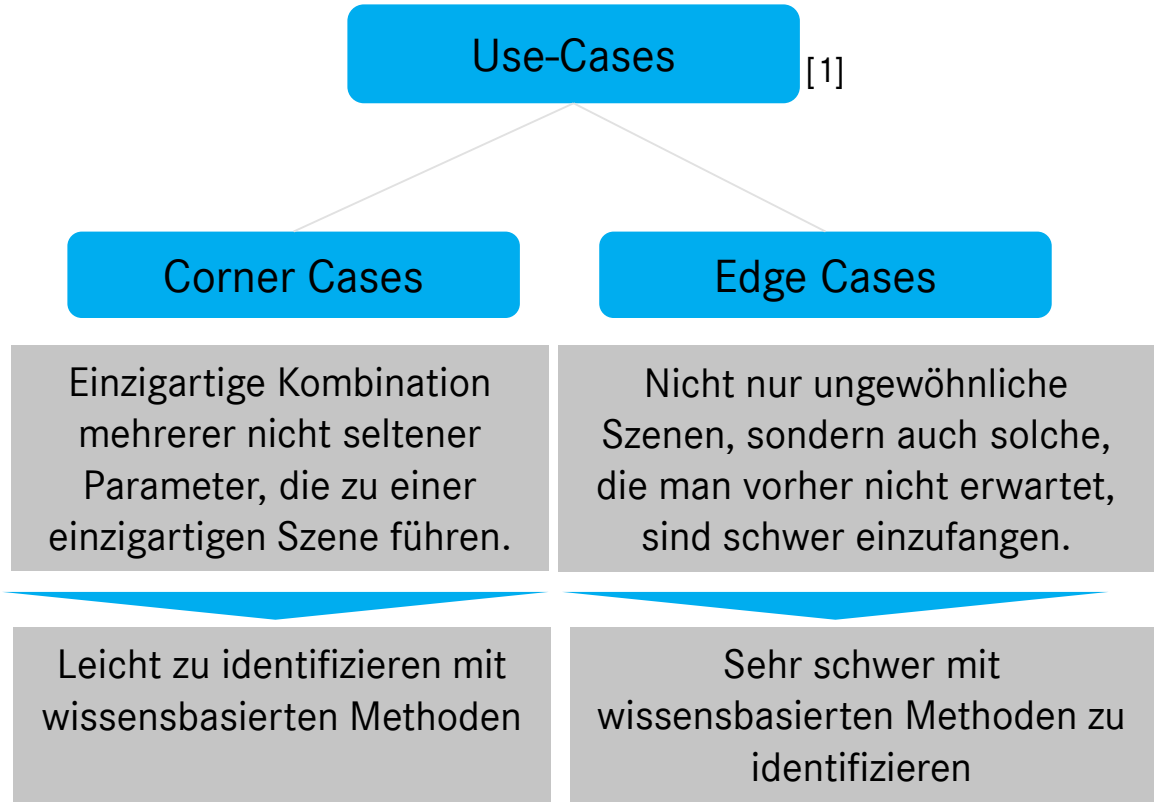
Hohe Dynamik
Ereignisse geschehen mit hoher Dynamik und abrupt, z.B. Skater kann plötzlich auf die Straße springen

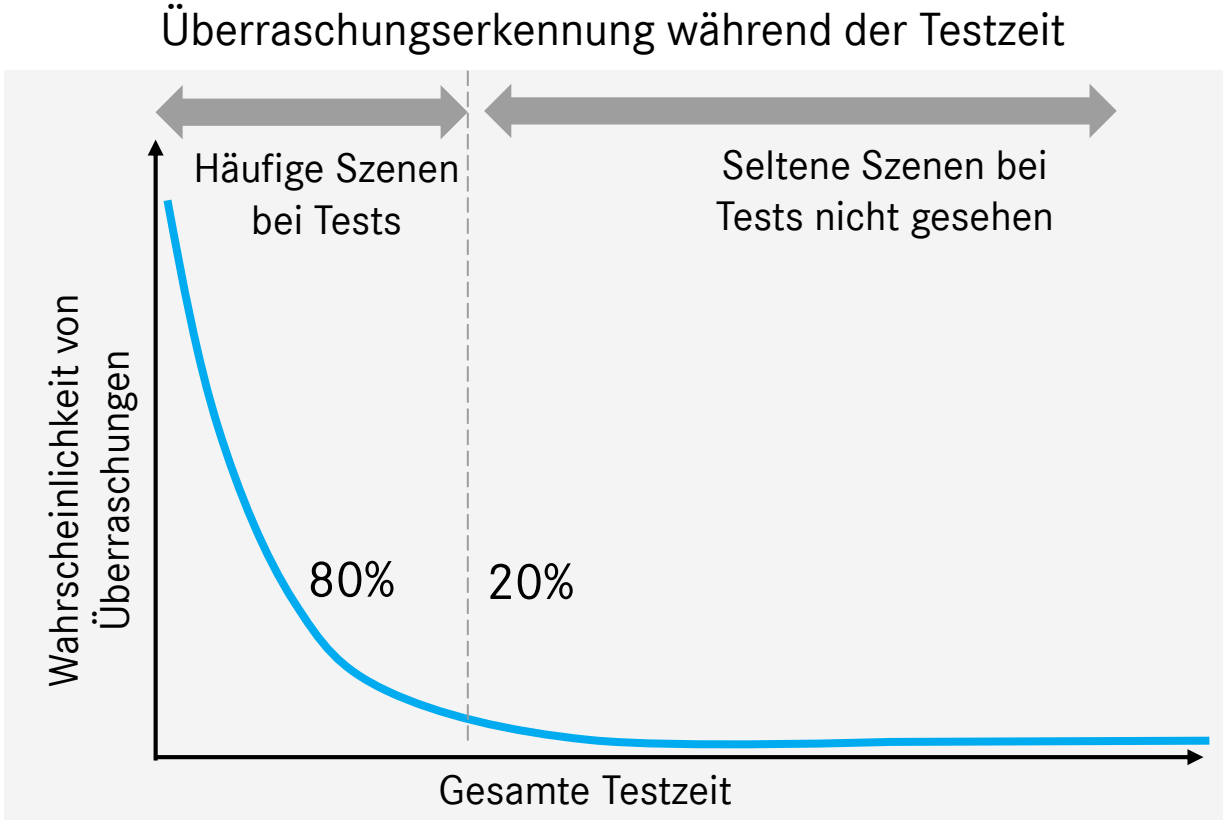
Seltene Szene
Es kann immer eine einzigartige Kombination von Ereignissen eintreten, die ungewöhnlich sind.



Ungesehene Szene
es können auch Ereignisse geschehen, die nicht vorstellbar sind, zum Beispiel das Landen auf Autobahnen

80% der Szenarien können in kürzerer Zeit gesehen werden, 20% benötigen eine lange Testzeit

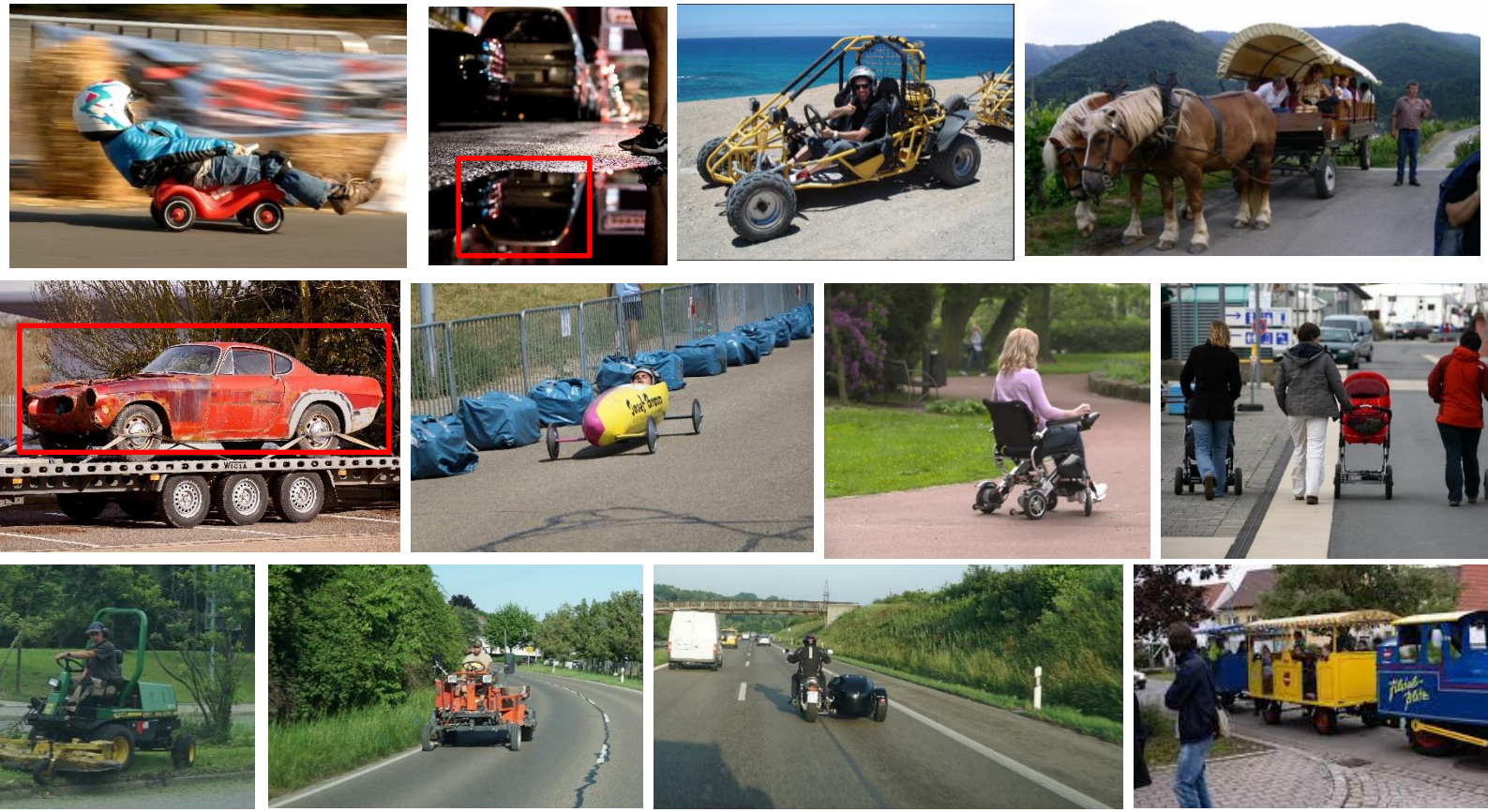




[1] Ensuring autonomous vehicle perception safety, Koopman, 2019

Selbst häufige bekannten Szenen können schwer zu erkennen sein

Was ist Auto?
vier Räder und eine Karosserie?



Wie viele Daten benötigen wir für Automatisiertes Fahren?

exemplarische Kalkulation

Das Unfallrisiko eines normalen Autos mit menschlichem Fahrer beträgt 4x10E-6 [1]

Ein automatisiertes Fahrzeug muss mindestens dreimal besser sein (Log-Skala) ~10E-9 Risiko-Wahrscheinlichkeit für automatisierte Autos [2]

Der erforderliche Datenmenge ist umgekehrt proportional, d.h. wir benötigen ~ 10E9 Key-Frames

Die Annotation der einzelnen Key-Frames dauert 20 Minuten - 1:30 Stunden

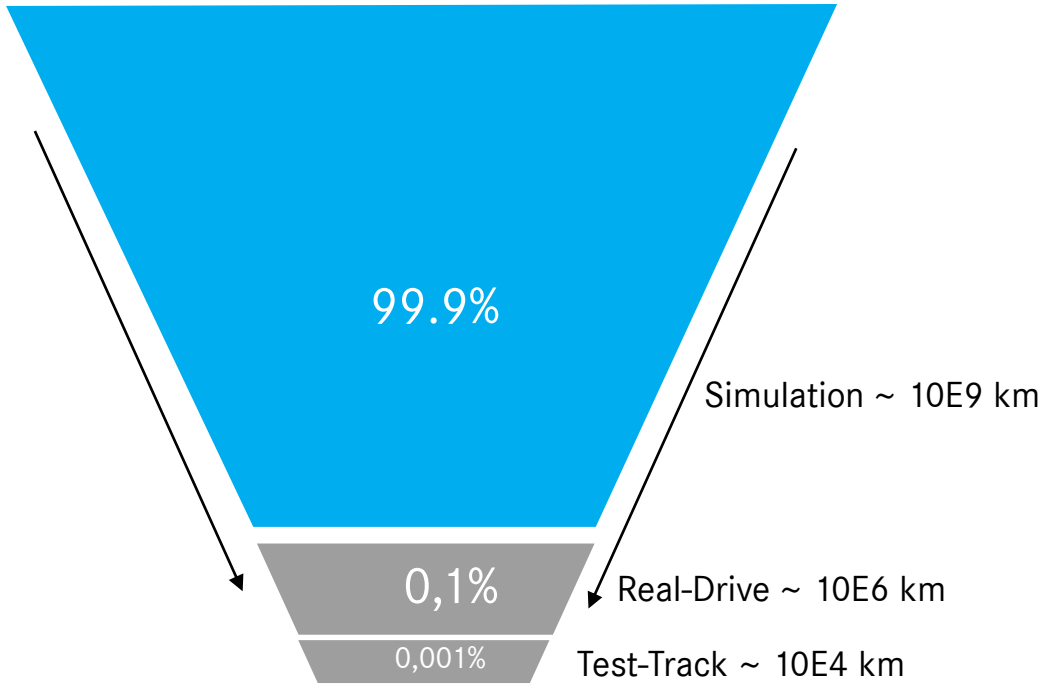
Es ist eine große Herausforderung, die erforderliche Datenmenge für Automatisiertes Fahren zu erfassen

Verringerung der Lernabhängigkeit von Daten

nicht alles aus echten Daten lernen

nur wenige Teile der Daten können real sein

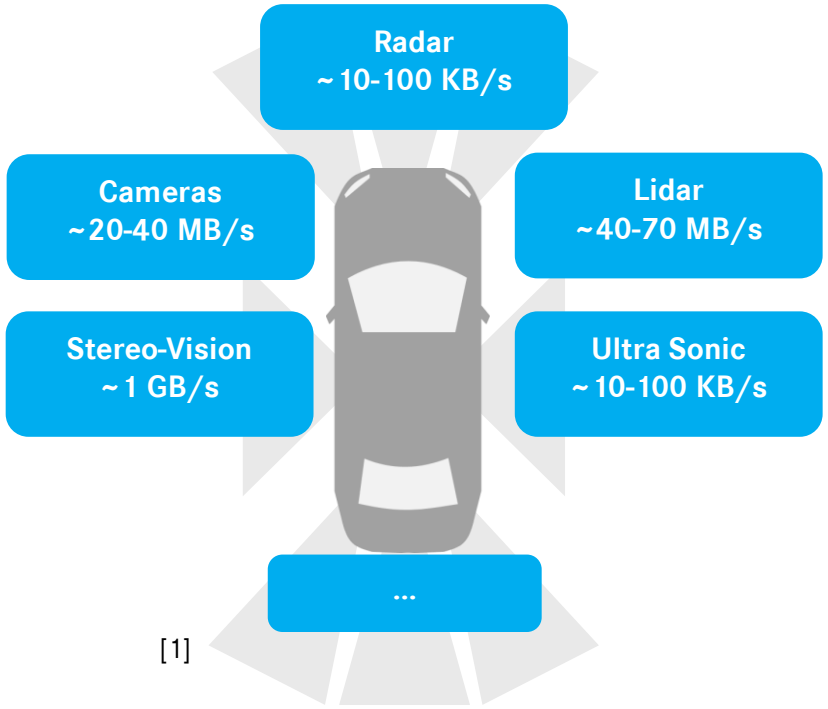
~ 6x10E9. km sind für einen vollständigen Sicherheitsnachweis erforderlich [3]



[1] Automatisches oder Autnomes Fliegen? En Statusbericht aus der Verkehrsluftfahrt, Müller, 2018
[2] On a Formal Model of Safe and Scalable Self-driving Cars, Shalev-Shwartz, Mobileye, 2017
[3] The release of autonomous vehicles, Wachenfeld et al., 2016

Big-Data mit Analytics@Scale

Typischer Sensoraufbau

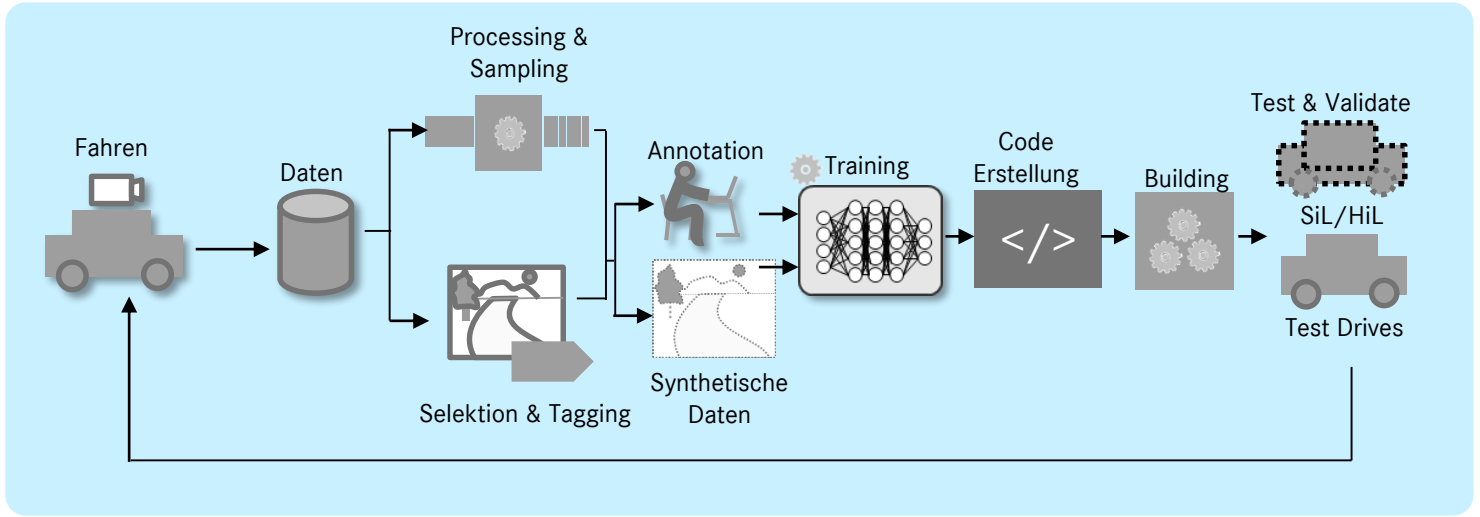


[1]

Big-Data

- ~ 100 Fahrzeuge Flotte
- ~ 5-10 Stunden Fahrzeit pro Fahrzeug
- Peta-Bytes an Daten jeden Tag, die in kurzer Zeit ausgewertet werden müssen.

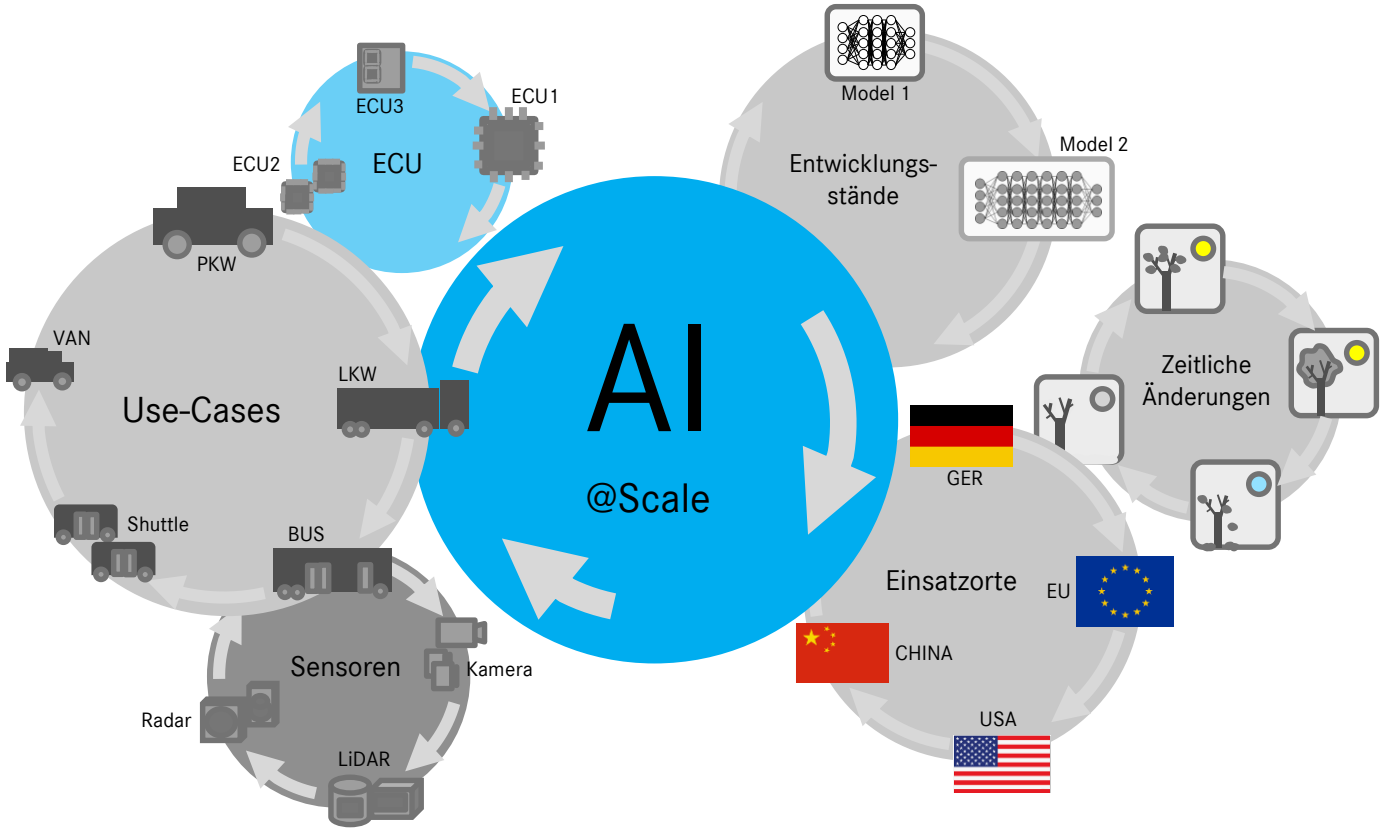
Skalierbares Data-Pipeline für Big-Data

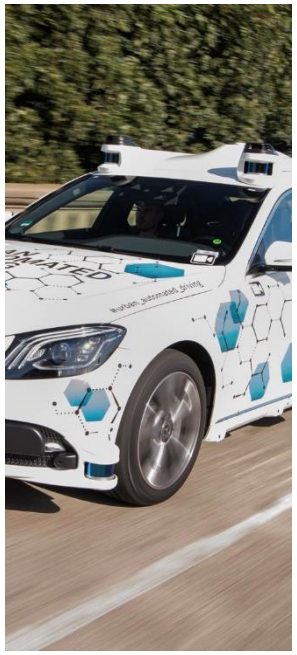


- Wenn Sie mehr Daten sehen wollen, müssen Sie mehr verarbeiten können
- Der Gewinner ist nicht derjenige, der die größte Menge an Daten hat, sondern derjenige, der eine hohe Datenqualität hat und das meiste aus den Daten herausholen kann.

[1] Automated Driving – Challenges and Opportunities, Watzenig, 2017

Noch anspruchsvoller: alles verändert sich, das Produkt, der Kunde, der Markt, der Anwendungsfall

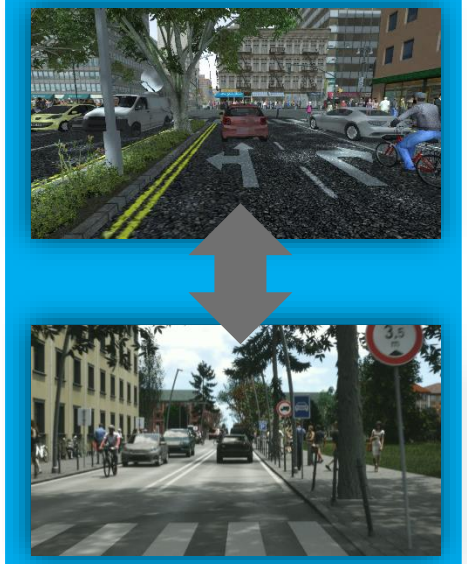




Im Kontext der Automobilindustrie befindet sich alles im ständigen Wandel. Der Markt- und Kundenbedarf, der Anwendungsfall, die Umwelt und natürlich der Entwicklungsstand.
Wir brauchen einen Ansatz für das kontinuierliche Lernen und die Nutzung des bereits gewonnenen Wissens.

Beispiele für typische Domain-Änderungen im Kontext Automatisiertes Fahren

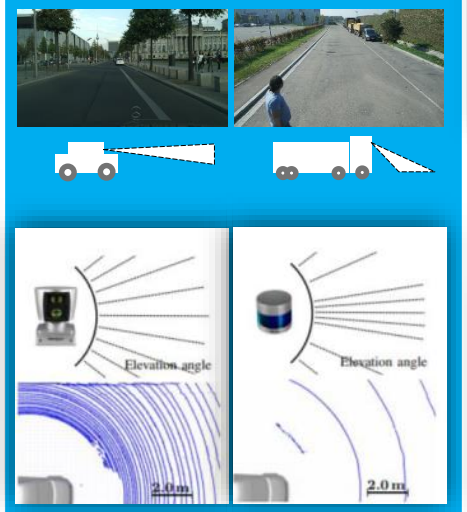
Synthetische Daten



Reale Daten

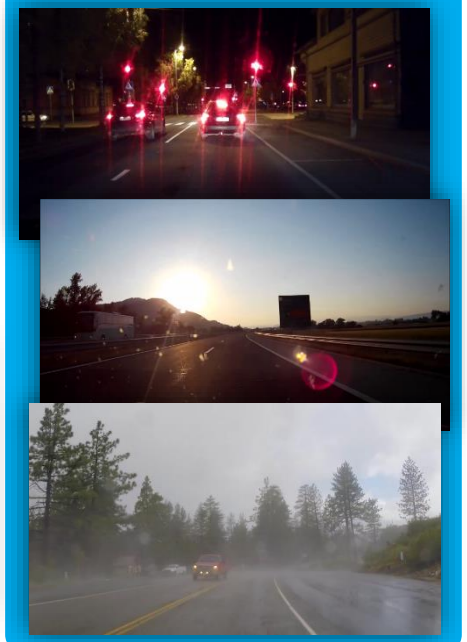
- Sensormodelle

Umgang mit Änderungen bei Sensoren



- Sensortechnologie
- Sensortyp
- Sensorpositionen
- ...

Kurzfristige Änderungen



Source: WildDash dataset

- Wetter
- Licht

Langfristische Änderungen

Berlin 2012

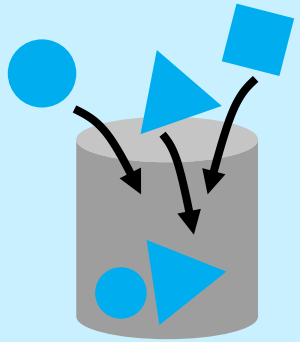
Berlin 2020

- Neue Objekte
- ...

Umgang mit örtlichen Änderungen

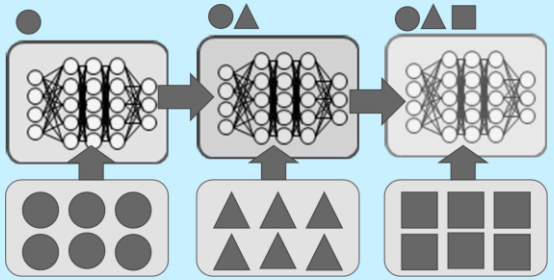
- Länder

Was sind die Enablers für Autonomy@Scale und den Umgang mit Domänenänderungen



Domain Adaptation

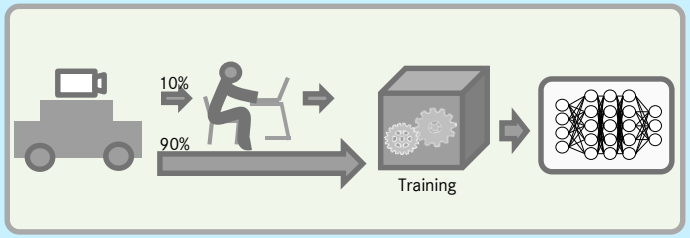
Wir müssen in der Lage sein, mit den Daten aus verschiedenen Quellen umzugehen



Transfer Learning

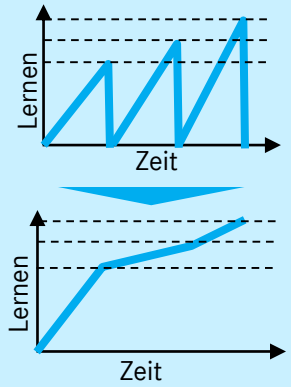
Bereits gewonnene Erkenntnisse müssen weiter genutzt werden:

- Transfer-Lernen für andere Domänen muss möglich sein



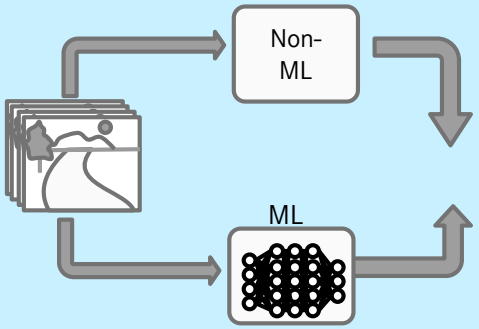
Unsupervised Learning

Weniger menschliche Intervention und Überwachung
Stärkerer Schwerpunkt auf halbüberwachten/unüberwachten Lernen



Continuous Learning

- Kontinuierliches Lernen auf der Grundlage verfügbarer und bereits abgesicherter Modelle



Kombi-Ansatz

Die Schwäche der Algorithmen des maschinellen Lernens kann mit Nicht-ML-Algorithmen kompensiert werden. Anwendung der Nicht-ML-Algorithmen dort, wo deren Leistung und Sicherheit besser ist.

Forschungsfragen

- Wie kompensieren wir die Schwächen der ML-Algorithmen?
- Die Datenerfassung ist eine langlebige Aufgabe. Datenquellen und -formate werden sich mit der Zeit ändern. Wie können wir eine kontinuierliche Nutzung garantieren?
- Ist Autonomy@scale ein echtes Bedürfnis oder eine interessante Forschungsaufgabe? Inwieweit können die aktuellen Probleme mit mehr Betreuungsunterstützung und Geld gelöst.
- Sind die KI- und ML-Algorithmen der Schlüssel zu allen Problemen?

Vielen Dank für Ihre Aufmerksamkeit

Fragen und Antworten

Bitte stellen Sie Ihre Fragen im Chat.

Data-driven Development – Wie kommen wir zu der Datenbasis, die wir brauchen?

Herausforderungen auf dem Weg zum autonomen Fahren -
Wie erreichen wir eine Datenbasis, die wir brauchen?

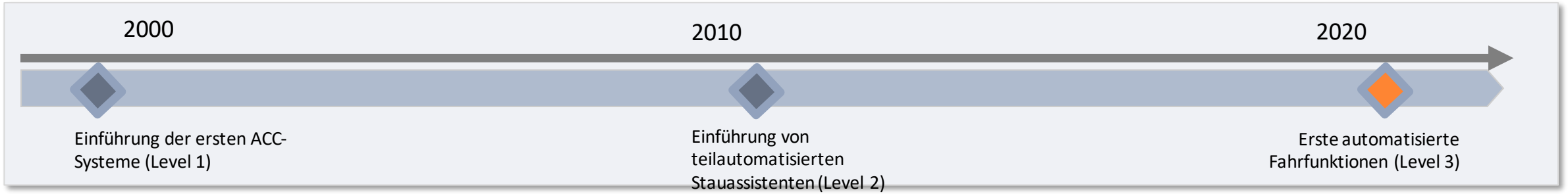
Dr. Mark Schiementz
04.12.2020

BMW GROUP THE NEXT 100 YEARS

VDA Leitinitiative autonomes und vernetztes Fahren

Rolls-Royce Motor Cars Limited

„AUSREICHENDEN DATENGRUNDLAGE“ IM WANDEL DER ZEIT.



Die Definition von „ausreichende Datengrundlage“ hat sich im Laufe der Zeit geändert.

FAS, SAE-Level 1/2:

- Fahrer trägt die Verantwortung, d.h. er muss die Fahrfunktion überwachen und ggf. eingreifen.
- „klassische“ Absicherung mit deterministischen Tests inkl. Variationen plus Realfahrversuch.

Automatisiertes Fahren, SAE-Level 3:

- Fahrfunktion ist in Verantwortung und muss innerhalb des jeweiligen Use-Cases alle Fahraufgaben meistern (u.a. mit KI-Anteilen).
- „klassische“ Absicherung reicht nicht mehr aus.
- Neue „stochastische“ Absicherungsmethoden auf Basis eines großen Wissens über den Realverkehr notwendig.

MOTIVATION.

- Der Qualitäts- und Sicherheitsnachweis einer automatisierten Fahrfunktion benötigt eine ausreichende Datengrundlage.

- Komplexe Produkte des automatisierten Fahrens können nur wettbewerbsfähig mit Daten-getriebenen Methoden entwickelt und abgesichert werden.

- Der Erfolg von Daten-getriebenen Förderprojekten hängt wesentlich von der Qualität, aber auch von der Menge der zu Verfügung stehenden (auch nicht anonymisierten Messdaten) ab.

AUTOMATISIERTES FAHREN – WOFÜR WIRD BIG DATA BENÖTIGT?

Herausforderung:
Beherrschung des offenen Kontexts der Fahraufgabe.

Lösung:
Wissen generieren, u.a. aus einer großen, repräsentativen Datenmenge.

Funktionsentwicklung:
ODD, KI-Lerndaten, etc.

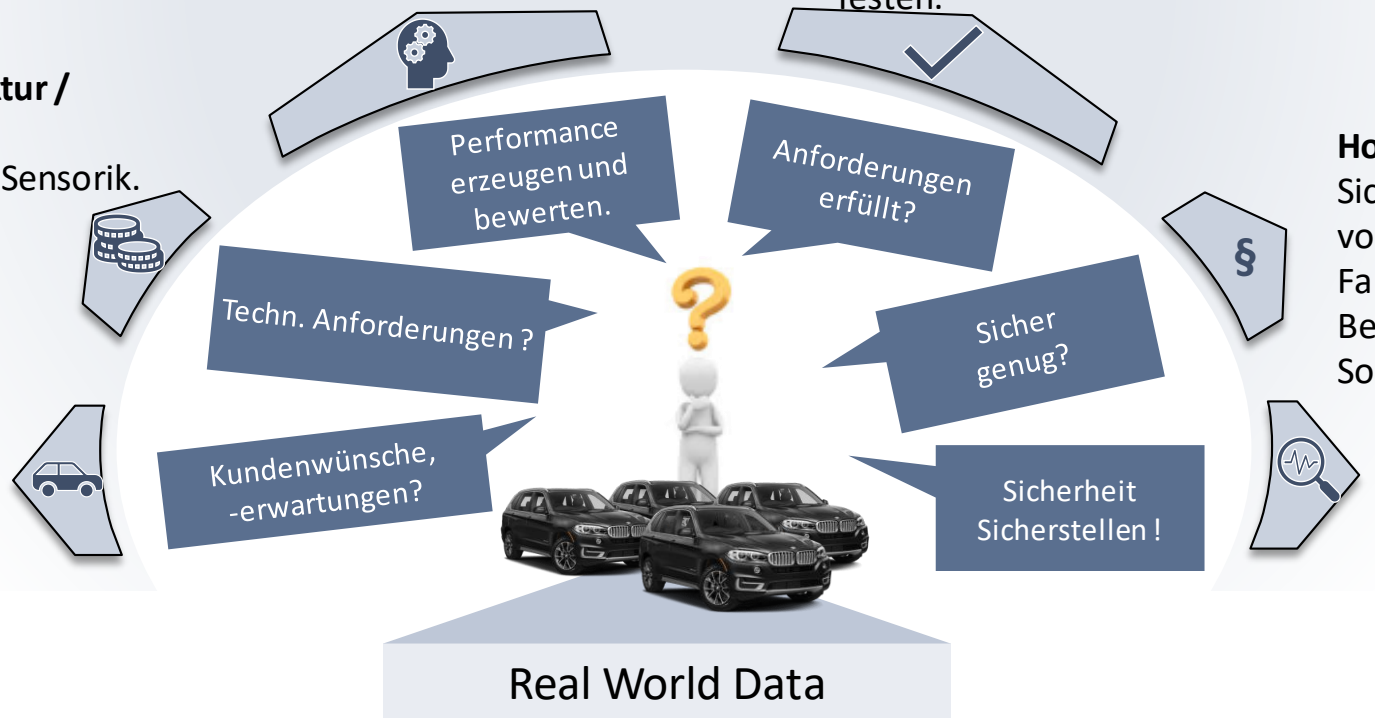
Absicherung:
Verifikation & Validierung von Funktionen, Szenarien-basiertes Testen.

Systemdesign / Architektur / Produktkosten z.B.
Speicher, Rechenpower, Sensorik.

Homologation, Typzulassung:
Sicherheitsnachweis von automatisierten Fahrfunktionen, Bestätigung eines rechtsfähigen Sollverhaltens

Kundenorientierung, Produktdesign, Produktangebot.

Qualitäts-, Sicherheitsmonitoring.



AUTOMATISIERTES FAHREN – WOFÜR WIRD BIG DATA BENÖTIGT?

Herausforderung:
Beherrschung des offenen Kontexts der Fahraufgabe.

Lösung:
Wissen generieren, u.a. aus einer großen, repräsentativen Datenmenge.

Die zur Verfügung stehende Datengrundlage entscheidet nicht nur über Produktqualität und –sicherheit, sondern v.a. auch über die **Wettbewerbsfähigkeit**.

UNTERSCHIEDLICHE ANFORDERUNGEN AN DATEN.

Unterschiedliche Use-Cases: Absicherung, Homologation, KI-Entwicklung, Verkehrsanalyse etc.

Anonymisierung akzeptabel für Anwendungszweck?

Unterschiedliches Fahrumfeld: Highway / Urban / etc

Mit / Ohne **Sensorrohdaten**

Grundsätzlicher Detaillierungsgrad / Menge der Daten pro Fzg

Mischverkehr oder abgegrenzter Bereich?

Perspektive (Daten aus Fzg, Verkehrsinfrastruktur, Drohnen, etc.)

Notwendigkeit der Repräsentativität der Gesamtheit der gesammelten Daten

Sensorset & Verbauort → Einfluss auf Nutzbarkeit der Daten

Geografische- oder Länder-spezifische Daten

Mit / Ohne Referenzsensorik

...

HERAUSFORDERUNG BEI AUFWAND UND KOSTEN.

- Die **Kosten** für die Datenerhebung, -speicherung sind in den aktuellen Förderprojekten **nicht im üblichen Kostenrahmen darstellbar** und führt zu überproportionalen hohen Kosten bei den Industriepartnern.

  → kritisch für Engagement unter den aktuellen Randbedingungen.

- Der **zeitliche Aufwand** zur Erhebung von großen Datenmengen als Grundlage für ein einzelnes Förderprojekt ist **innerhalb der Laufzeit eines Förderprojektes aktuell nicht realisierbar.**

BISHERIGES VORGEHEN IN FÖRDERPROJEKTEN.

Aktuell werden Förderprojekte hinsichtlich der Realisierbarkeit von Kosten und Inhalten **herunterskaliert**:

- Nur exemplarisches Arbeiten auf Minimal- oder künstlich erzeugten Datensätzen.
  → eingeschränkte Verifikation der entwickelten Ergebnisse / Methodik.
- Das richtige Zusammenführen und die Anreicherung der großen Datenmengen im Sinne der Projektziele stellt allein enorme zeitliche und finanzielle Herausforderung dar.
- Tools für das Prozessieren von Big Data sind kostenaufwändig und deren Einsatz kann aktuell nur eingeschränkt gefördert werden.

Prognose:

In Zukunft werden selbst die notwendigen Minimaldatensätze für die Förderprojekte immer größer werden, so dass diese innerhalb von einzelnen Förderprojekten nicht mehr erhoben werden können.

MÖGLICHE LÖSUNGEN.

→ Eine separate nicht Förderprojekt-spezifische Datenerhebung, z.B. für größere Themenkomplexe des automatisierten Fahrens.

→ Mehrfachnutzung von Datensätzen durch verschiedene Förderprojekte mit dem Ziel die Daten für möglichste viele Fragestellung / Anwendungsfälle nutzen zu können.

→ Definition eines Stands der Technik oder ggf. separate Entwicklung einer universellen Toolchain für Fragestellungen des automatisierten Fahrens.

→ Attraktivere Gestaltung des Kosten/Nutzen-Verhältnis innerhalb der Förderung notwendig.

Frage an die Workshopteilnehmer: Welche weiteren Lösungsmöglichkeiten und Bedarfe gibt es aus Ihrer Sicht?

SPANNUNGSFELD → GESETZLICHE RANDBEDINGUNGEN - DATENSCHUTZ

Zulässigkeit der Datenerhebung personenbezogener Daten (beispielhaft):

Art. 5 DSGVO

Personenbezogene Daten müssen…für festgelegte, eindeutige und legitime Zwecke erhoben werden und dürfen nicht in einer mit diesen Zwecken nicht zu vereinbarenden Weise weiterverarbeitet werden …dem Zweck angemessen und erheblich sowie auf das für die Zwecke der Verarbeitung notwendige Maß beschränkt sein („Datenminimierung“);

Art. 89 DSGVO

Die Verarbeitung zu im öffentlichen Interesse liegenden …wissenschaftlichen …Forschungszwecken oder zu statistischen Zwecken unterliegt geeigneten Garantien für die Rechte und Freiheiten der betroffenen Person gemäß dieser Verordnung. Mit diesen Garantien wird sichergestellt, dass technische *(z.B. Anonymisierung)* und organisatorische Maßnahmen bestehen, mit denen insbesondere die Achtung des Grundsatzes der Datenminimierung gewährleistet wird.

SPANNUNGSFELD → GESETZLICHE RANDBEDINGUNGEN - DATENSCHUTZ

→ **Personenbezogene Daten müssen also entsprechend deren Kritikalität nach DSGVO auch in wissenschaftlichen Projekten besonders geschützt werden.**

Welche Daten sind kritisch?

- Kritisch sind nicht anonymisierte 360°-Videos inkl. GPS-Stempel.
- Radar und Lidar-Daten gelten als unkritisch, so lange diese nicht personen-spezifisch zugeordnet werden können.

Wo liegt der Schmerz für die Förderprojekte, bzw. der Konsortialpartner?

Der vorgesehene **Bußgeldrahmen** im Fall eines Unternehmens **umfasst bis zu 4% des gesamten weltweit erzielten Jahresumsatzes** im vorangegangenen Geschäftsjahr bei gravierenden Verstößen gegen die DSGVO und 2% bei leichten Verstößen.

LÖSUNGEN VS. DSGVO.

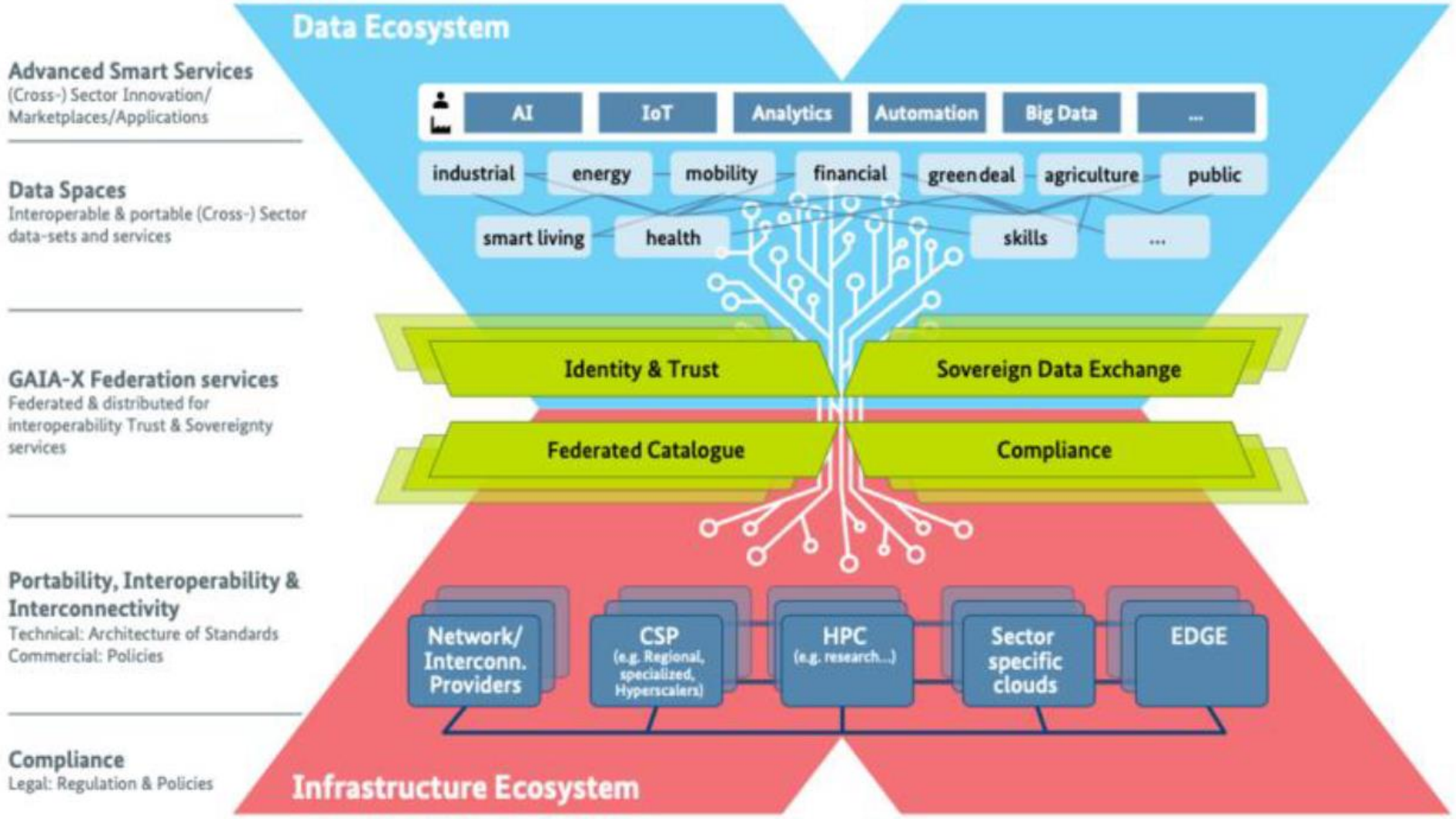
→ Eine separate nicht Förderprojekt-spezifische Datenerhebung, z.B. für größere Themenkomplexe des automatisierten Fahrens.

Spannungsfeld mit DSGVO-Formulierung „festgelegte, eindeutige und legitime Zwecke“.

→ Mehrfachnutzung von Datensätzen durch verschiedene Förderprojekte mit dem Ziel die Daten für möglichste viele Fragestellung / Anwendungsfälle nutzen zu können.

Spannungsfeld mit DSGVO-Formulierung „Datenminimierung“.

CHANCE GAIA-X.



GAIA-X

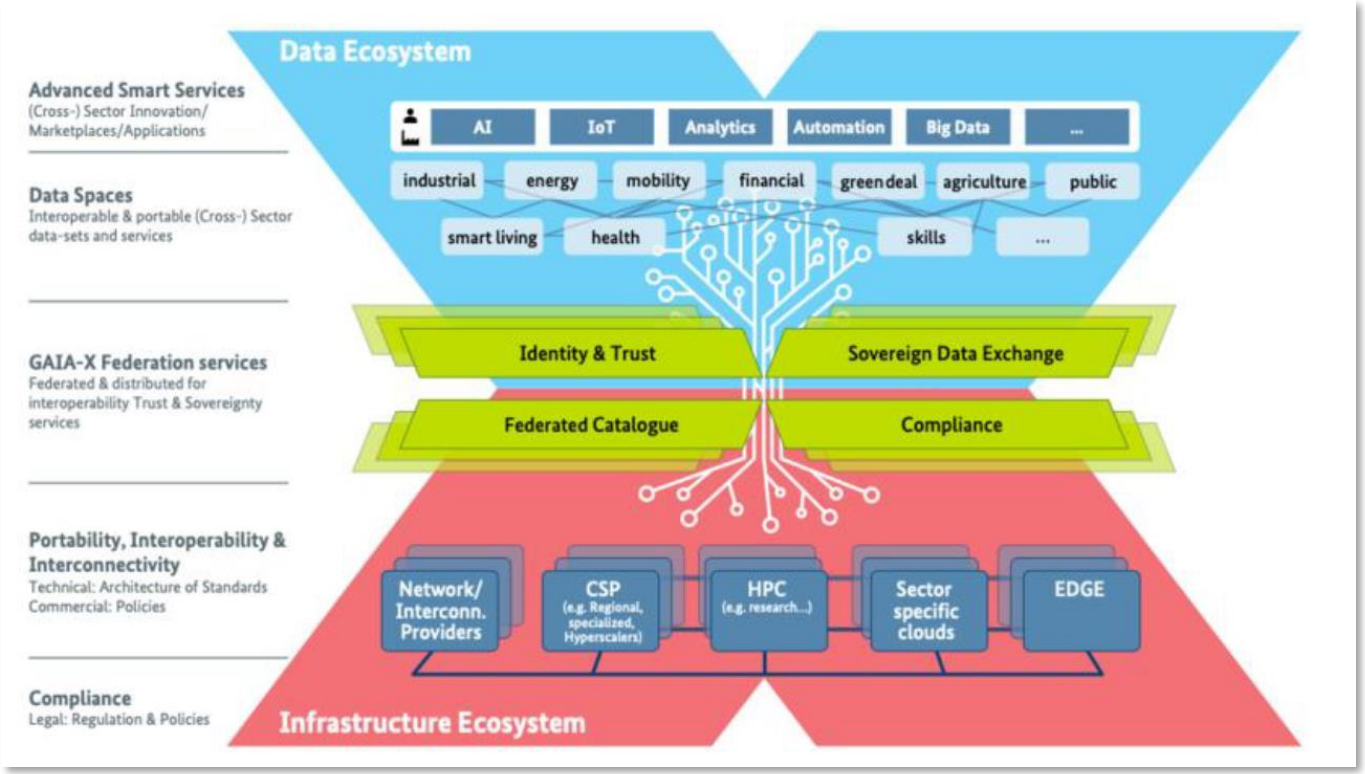
GAIA-X: IT-PLATTFORM ODER MEHR?

GAIA-X…

…hat das Potential um die genannten Problemfelder aufzulösen.

…versucht das Maximum aus vorhandenen Datensätzen herauszuholen (Plattformgedanke).

...könnte einen deutlichen Mehrwert bieten,
wenn übergreifende, große Datensätze für die Mehrfachnutzung DSGVO-konform bereit gestellt würden.



→ Quo vadis GAIA-X?

IT-Plattform als Datenvermittler oder Enabler für datengetriebene Schlüsselanwendungen?

Diskussion:
Welche Spannungsfelder können durch aufgelöst werden?

GAIA-X

WAS WIRD FÜR DATENGETRIEBENE FÖRDERPROJEKTE IN ZUKUNFT BENÖTIGT?

Daten, viele Daten...

- Datengetriebene Förderprojekte werden nur dann einen Effekt auf die zu fördernden technischen Lösungen und somit die Wettbewerbsfähigkeit haben, wenn eine ausreichend große, qualitativ angemessene Datenbasis vorhanden ist.
- Diese große Datenbasis kann aufgrund des zeitlichen Aufwands nicht innerhalb der methodischen Förderprojekte erhoben werden.
- Diese Datenbasis muss u.a. repräsentativ die Realität für die jeweiligen Use-Cases abbilden.
- Die Kosten für die Datenerhebung müssen über den aktuell üblichen Rahmen förderfähig sein.
- Diese übergreifende Datenbasis bietet neue Chancen für Arbeiten von KMUs und Wissenschaftspartnern.

WAS WIRD FÜR DATENGETRIEBENE FÖRDERPROJEKTE IN ZUKUNFT BENÖTIGT?

DSGVO konforme Lösungen und Klärungen zur Rechtssicherheit:

- Um die effiziente Mehrfachnutzung der Daten (also für unterschiedliche Zwecke) im Sinne der Rechtssicherheit zu ermöglichen, ist eine Bewertung und Klärung bzgl. der DSGVO notwendig.
- Eine Mehrfachnutzung von Datensätzen bedeutet auch, dass diese Datensätze mehreren und ggf. höheren Anforderungen genügen müssen. Dieses ist nicht im Sinne der Datenminimierung in der DSGVO → Bewertung und Klärung notwendig.
- Diese DSGVO-Klärung ist durch die Industrie bisher alleine nicht möglich gewesen.
- Von Drohnen aufgenommene Verkehrsdaten stellen für manche Use-Cases (z. B. Szenarien-basiertes Wissen über Realverkehr) eine DSGVO-konforme Ergänzung oder Alternative dar.

WAS WIRD FÜR DATENGETRIEBENE FÖRDERPROJEKTE IN ZUKUNFT BENÖTIGT?

Datenbank(en) und Data-Toolchain auf Stand der Technik:

- Es muss geklärt werden wo Förderprojekt-übergreifende Daten abgelegt und dauerhaft gespeichert werden können.
  - Wer übernimmt die Kosten hierfür?
  - Ohne externe Kostenübernahme oder -beteiligung wird z.B. der Datenbankbetreiber ein eigenes Datengeschäftsmodell zur Finanzierung entwickeln. Was bedeutet das für die Akzeptanz der Daten-liefernden Partner?

- Nutzbarkeit der Daten, Toolchain für Datenverarbeitung, -services notwendig.
  - IT-technisch leicht durch Förderprojekte nutzbar → Kompatibilität und generische Schnittstellen.
  - Entwicklung übergreifender Tools für Datenerhebung, -verarbeitung und -services, damit sich z.B. methodische Projekte auf die Inhalte fokussieren können.

VORSCHLAG ZUR DISKUSSION.

Ein staatlich gefördertes oder finanziertes Projekt erhebt und aktualisiert regelmäßig einen DSGVO konformen Datensatz (Realdaten und Simulationsdaten), der unter einer geeigneten Lizenz zur freien Nutzung bereitgestellt wird.

Welche Themenkomplexe / Aspekte müssen hierzu und darüber hinaus berücksichtigt werden?

Fragen und Antworten

Bitte stellen Sie Ihre Fragen im Chat.

Podiumsdiskussion

Zusammenfassung und nächste Schritte

1. Wesentliche Ergebnisse
2. Weiterer Zeitplan
3. Weitere Vernetzung

Kontakt

Internet: www.kopa35c.de
E-Mail: kopa35c@bmwi.bund.de

Referat IVA3 Digitalisierung, Industrie 4.0
Bundesministerium für Wirtschaft und Energie

Hannoversche Straße 28 - 30,
10115 Berlin
Internet: http://www.bmwi.de