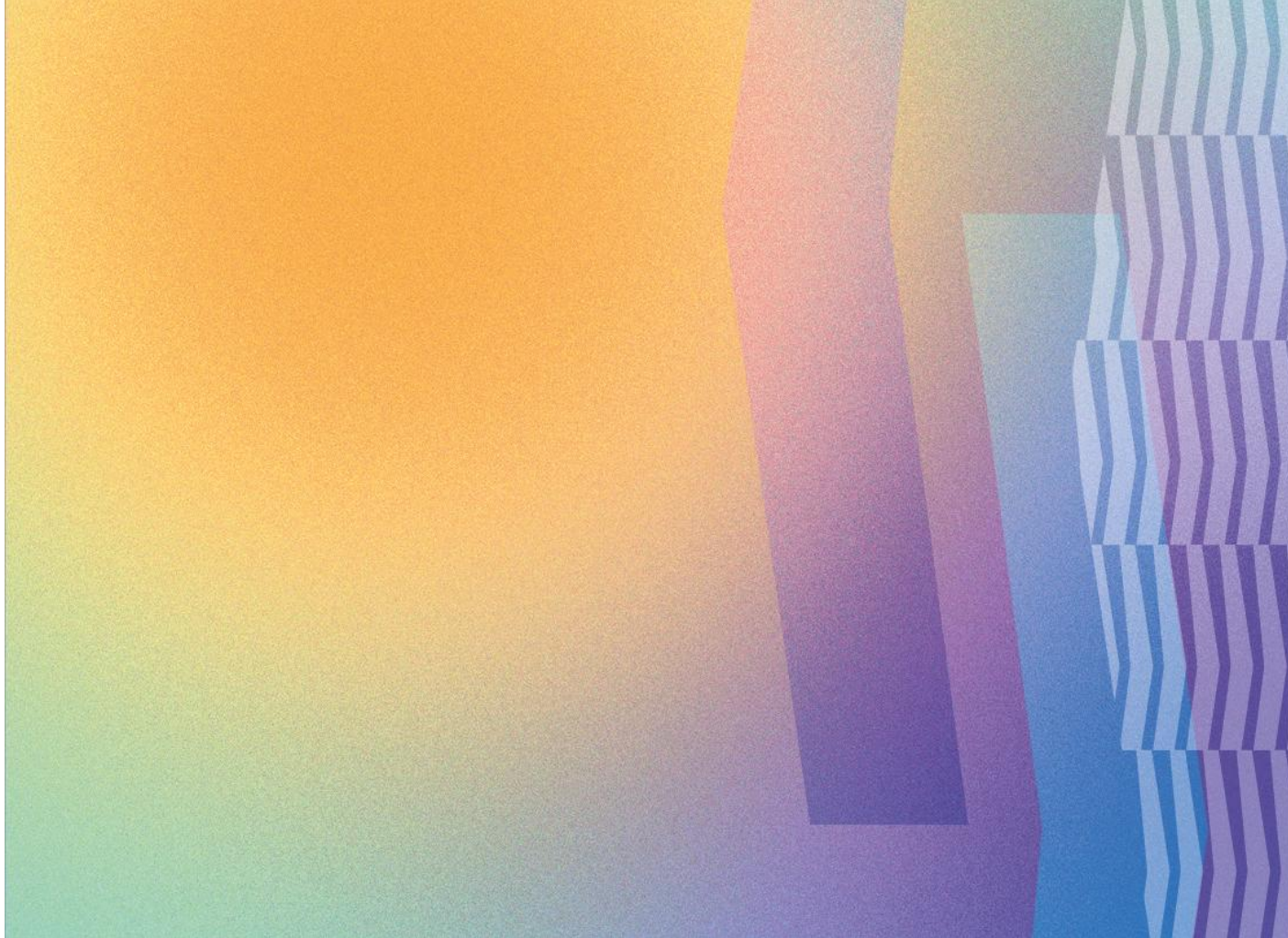


Eine Studie der Begleitforschung zum Technologieprogramm Edge Datenwirtschaft im Auftrag des Bundesministeriums für Forschung, Technologie und Raumfahrt (BMFTR)



BEWERTUNG VON EDGE-CLOUD-SYSTEMEN

Vorgehen, Instrumente und Metriken

Impressum

Autoren

Dr. Benjamin Alt, ArtiMinds Robotics GmbH, *Projekt EASY*

Sebastian Bast, Institut für Softwaresysteme, Umwelt-Campus Birkenfeld, Hochschule Trier, *Projekt EASY*

Matthias Babel, Fraunhofer-Institut für Angewandte Informationstechnik FIT, *Projekt DEER*

Julian Buitmann, Hochschule Osnabrück, *Projekt EDNA*

Dr. Lars Creutz, Institut für Softwaresysteme, Umwelt-Campus Birkenfeld, Hochschule Trier, *Projekt EASY*

Prof. Dr.-Ing. Guido Dartmann, Institut für Softwaresysteme, Umwelt-Campus Birkenfeld, Hochschule Trier, *Projekt EASY*

Marvin Ehaus, Fraunhofer-Institut für Angewandte Informationstechnik FIT, *Projekt DEER*

Dr. Thomas Engel, Siemens AG, Foundational Technology, *Projekt GEMIMEG-II*

Fabian Gast, Institut für Produktionsmanagement, Technologie und Werkzeugmaschinen (PTW) | TU Darmstadt, *Projekt ESCOM*

Nina Großegesse, Salzburg Research Forschungsgesellschaft m.b.H., *Projekt EASY*

Achim Guldner, Institut für Softwaresysteme, Umwelt-Campus Birkenfeld, Hochschule Trier, *Projekt EASY*

Nils Jahnke, Fraunhofer-Institut für Software- und Systemtechnik ISST, *Begleitforschung Edge Datenwirtschaft*

David Jilg, Deutsches Forschungszentrum für Künstliche Intelligenz (DFKI) GmbH – Erfahrungsbasierte Lernende Systeme (EBLS), *Projekt EASY*

Dr. Marc-Fabian Körner, Fraunhofer-Institut für Angewandte Informationstechnik FIT, *Projekt DEER*

Justus Leskow, Hochschule Osnabrück, *Projekt EDNA*

Johannes Mundorf, Deutsches Forschungszentrum für Künstliche Intelligenz (DFKI) GmbH – Innovative Fabrikssysteme (IFS), *Projekt EASY*

Prof. Dr. Stefan Naumann, Institut für Softwaresysteme, Umwelt-Campus Birkenfeld, Hochschule Trier, *Projekt EASY*

Dr.-Ing. Alexandra Pehlken, OFFIS e.V. - Institut für Informatik, *Projekt EDNA*

Marvin Schacht, Institut für Softwaresysteme, Umwelt-Campus Birkenfeld, Hochschule Trier, *Projekt EASY*

Leo Schick, Fraunhofer-Institut für Angewandte Informationstechnik FIT, *Projekt DEER*

Dr. Tobias Schlagenhauf, Robert Bosch GmbH, Bosch Research - AI in Production, *Projekt EASY*

Alexander Schultheis, Deutsches Forschungszentrum für Künstliche Intelligenz (DFKI) GmbH –
Erfahrungsbasierte Lernende Systeme (EBLS), *Projekt EASY*

Dr. Karsten Schörner, Siemens AG, Foundational Technology, *Projekt ECO:DIGIT*

Sebastian Straub, Institut für Innovation und Technik (iit), *Begleitforschung Edge
Datenwirtschaft*

Prof. Dr. Jens Strüker, Fraunhofer-Institut für Angewandte Informationstechnik FIT, *Projekt
DEER*

Ralph Traphöner, Empolis Information Management GmbH, *Projekt EASY*

Moritz-André Weiher, Deutsches Forschungszentrum für Künstliche Intelligenz (DFKI) GmbH –
Smart Enterprise Engineering (SEE), *Projekt SECAI*

Redaktion

Nils Jahnke

Begleitforschung Edge Datenwirtschaft
Fraunhofer-Institut für Software- und Systemtechnik ISST

Herausgeber

Peter Gabriel

Dr. Nicole Wittenbrink

Begleitforschung Edge Datenwirtschaft
Institut für Innovation und Technik (iit)
in der VDI / VDE Innovation + Technik GmbH
Berlin

Datum

Juni 2025

Layout

PRpetuum GmbH

Diese Orientierungshilfe wurde im Auftrag des Bundesministeriums für Forschung, Technologie und Raumfahrt (BMFTR) im Rahmen der Begleitforschung zum Technologieprogramm „Edge Datenwirtschaft“ erstellt.

Management Summary

Edge-Cloud-Systeme (ECS) sind die Grundlage neuartiger intelligenter Dienste mit wirtschaftlichem und gesellschaftlichem Mehrwert in vielfältigen Wirtschaftszweigen. ECS adressieren zentrale Herausforderungen moderner Softwaresysteme wie die Einhaltung von Latenzzeiten, die Gewährleistung von Datensicherheit und Datenschutz und den effizienten Umgang mit Ressourcen durch die Nutzung verschiedener Software- und Hardwarekomponenten entlang des Kontinuums von der Edge zur Cloud. Je nach Anwendungsfall sind verschiedene ECS-Konfigurationen möglich.

Der echte Mehrwert einer bestimmten ECS-Konfiguration gegenüber ihren Alternativen kann jedoch nur gezeigt werden, wenn ihre Vorteile bewiesen werden können. Zentrales Instrument hierzu ist die Durchführung der Bewertung eines ECS. Die besonderen Charakteristika von ECS erschweren jedoch die Planung, Durchführung und Auswertung der ECS-Bewertung. Zu den Bewertungshemmnissen zählen insbesondere die verteilte Struktur von ECS, die große Vielfalt und Heterogenität eingesetzter Hardware- und Softwarekomponenten sowie die dynamische Natur von ECS – beispielsweise können einzelne Systembausteine über die Lebensdauer ausgetauscht oder aktualisiert werden. Entsprechend fällt es Anwendenden häufig schwer, kontextrelevante Bewertungskriterien zu bestimmen und geeignete Metriken auszuwählen, um diese in einer konkreten Ausprägung eines ECS zu analysieren.

Vor diesem Hintergrund haben sich Early-Adopter von Edge-Computing aus den Projekten des Technologieprogramms „Edge Datenwirtschaft“ Bundesministeriums für Forschung, Technologie und Raumfahrt (BMFTR) sowie zwei Projekte aus angrenzenden BMFTR-Förderprogrammen zu einer Task-Force zusammengeschlossen. Ziel der Task-Force war es, individuelle Erkenntnisse und Lösungsansätze der Early-Adopter bei der Bewertung von ECS aus ihren Praxisprojekten zu diskutieren und zu konsolidieren. Konkret waren an der Task-Force die Projekte EASY, EDNA, ESCOM, DEER, und SECAI (Edge Datenwirtschaft) sowie die Projekte ECO:DIGIT (GreenTech Innovationswettbewerb) und GEMIMEG-II (Smart Service Welt II) beteiligt. Diese Orientierungshilfe macht die Ergebnisse der Task-Force für eine breite Öffentlichkeit verfügbar.

Die Orientierungshilfe richtet sich vorrangig an Verantwortliche in den Bereichen der Systemarchitektur und Systemintegration sowie in Entwicklungsbereichen, die eine ganzheitliche Evaluation ihres ECS durchführen oder Hinweise auf die Bewertung einzelner Charakteristika von ECS erhalten möchten. Sie erhalten einen umfassenden Einblick in die relevanten Konzepte zur Bewertung von ECS und praktikable Handlungshilfen, um eine Evaluation von ECS in der Praxis durchzuführen. Zudem können sich Führungskräfte, Projektmanagement-Teams und weitere Akteure mit Interesse an der Bewertung von ECS einen allgemeinen Überblick über die Thematik verschaffen. Schließlich adressiert die Orientierungshilfe auch Politik und Forschung, indem sie Handlungsempfehlungen zur Schaffung der notwendigen Rahmenbedingungen für eine nachhaltige Befähigung von Organisationen zur ECS-Bewertung formuliert.

Konkret präsentiert diese Orientierungshilfe zunächst einen über die Teilnehmenden der Task-Force hinweg harmonisierten Überblick über die Kernkonzepte einer Bewertung von ECS. Dazu zählen insbesondere die Vorstellung eines fünfstufigen Vorgehensmodells aus dem Bereich der Softwaresysteme für die Bewertung von ECS und eine konsolidierte Übersicht über relevante Bewertungskriterien und zugehörige Metriken aus den Bereichen Organisation, Technologie und

Umwelt. Um Hands-on-Einblicke in die Bewertung einzelner Kriterien wie der Ressourcennutzung, der Netzwerkperformanz oder der Umweltwirkungen von ECS in der Praxis zu geben, präsentieren die Early-Adopter zudem ausgewählte Vorgehensweisen und Ergebnisse aus ihren Forschungsprojekten, die über den aktuellen Stand der Technik hinausgehen.

Basierend auf den Erkenntnissen aus den Forschungs- und Entwicklungsarbeiten werden gemeinsam identifizierte wirtschaftliche, technische und organisatorische Herausforderungen bei der Bewertung von ECS aufgezeigt. Zu den größten Herausforderungen gehören die Anwendungsfallabhängigkeit der Bewertungen, die hohen Kosten bei der Bewertung bestimmter Kriterien und die geringe Kompetenz im Bereich der Bewertung von ECS in der Praxis.

Abschließend präsentiert diese Orientierungshilfe gemeinsam erarbeitete Handlungsempfehlungen in zwei Handlungsfeldern. Das erste Handlungsfeld bietet praxisnahe Empfehlungen für jene Akteure, die sich konkret mit der Bewertung von ECS befassen (z. B. Anwendende, Plattformbetreibende, Systemintegrierende und Herstellende von Komponenten). Es zeigt auf, wie sich eine unternehmensweite Bewertung etablieren lässt und welche Kriterien (u. a. Resilienz, Datenschutz, regulatorische Vorgaben) dabei berücksichtigt werden sollen. Das zweite Handlungsfeld beschäftigt sich mit der Verbesserung der übergreifenden Rahmenbedingungen, um Organisationen langfristig zur Bewertung von ECS zu befähigen. Hierbei werden insbesondere Forschungseinrichtungen und politische Entscheidungsträger adressiert. Die Empfehlungen umfassen den Kompetenzaufbau, die Einführung von Transparenzstandards, semantische Standardisierungen, die Steigerung der Erklärbarkeit von (KI-) Systemen und die Schaffung niedrigschwelliger Einstiegsmöglichkeiten für ECS.

Inhalt

1	Einleitung	6
2	Grundlagen der Bewertung von Edge-Cloud-Systemen: Definitionen, Vorgehen und Bewertungskriterien.....	8
2.1	Definition und Terminologie von Edge-Cloud-Systemen.....	8
2.2	Exemplarisches Vorgehensmodell zur Bewertung von Edge-Cloud-Systemen.....	9
2.3	Relevante Kriterien zur Bewertung von Edge-Cloud-Systemen.....	11
3	Methoden, Instrumente und Metriken zur Bewertung von Edge-Cloud-Systemen – Einblicke in die Praxis.....	15
3.1	Bewertung von Computing-Plattformen	16
3.2	Analyse und Optimierung der Datenübertragung.....	17
3.3	Bewertung von Energieverbrauch und CO ₂ -Emissionen der Gesamtsysteme	22
3.4	Bewertung von Methoden des Maschinellen Lernens – Föderiertes Lernen.....	25
3.5	Bewertung von KI-Planung.....	29
3.6	Bewertung von Methoden des Maschinellen Lernens – Fallbasiertes Schließen	31
3.7	Ansätze zur Sicherstellung von Privatsphäre, Datensouveränität und Vertrauen.....	33
3.8	Bewertung der Kosten.....	42
3.9	Ansätze zur rechtlichen Bewertung	44
4	Aktuelle Herausforderungen bei der Bewertung von Edge-Cloud-Systemen	47
5	Handlungsempfehlungen.....	53
5.1	Handlungsempfehlungen für die praktische Durchführung einer Bewertung von Edge-Cloud-Systemen	55
5.2	Handlungsempfehlungen zur nachhaltigen Befähigung von Organisationen zur Umsetzung von Bewertungen der Edge-Cloud-Systeme	58
6	Referenzen.....	63

1 Einleitung

Edge-Computing ermöglicht die Durchführung von Rechen- und Speicheroperationen nahe dem Ort der Leistungserbringung und Datenerzeugung. Es ermöglicht die Umsetzung neuer intelligenter Dienste mit besonderen Anforderungen an Latenzzeiten, Datenschutz, Datensicherheit oder die Anwendungsverfügbarkeit. Durch eine geringere Nutzung zentralisierter Systeme wie der Cloud und der damit verbundenen Datensparsamkeit stellt Edge-Computing zudem einen wichtigen Baustein zur Reduktion des ökologischen Fußabdrucks von IT-Infrastrukturen dar. Eine vollständige Abkehr von der Cloud ist dabei in den seltensten Fällen sinnvoll. Im Gegenteil: In Edge-Cloud-Systemen (ECS) ergänzen sich Edge und Cloud durch die Ausführung komplementärer Aufgaben. An der Edge erfolgt die Aufbereitung und Verarbeitung von Rohdaten, beispielsweise mittels Methoden der Künstlichen Intelligenz (KI). Die Cloud übernimmt hingegen Aufgaben wie die Systemorchestrierung, die langfristige Verwahrung von Analysedaten und die Bereitstellung aggregierter Analyseergebnisse. Ein Beispiel hierfür ist die bildbasierte Qualitätskontrolle: Bild-daten von Objekten werden an der Edge mittels Methoden der KI ausgewertet, die notwendigen Aktionen abgeleitet und beispielsweise an eine Maschinensteuerung weitergegeben. In der Cloud werden die KI-Modelle trainiert und an die Edge ausgerollt, relevante Trainingsdaten gesichert und aggregierte Qualitätsdaten für relevante Stellen langfristig zugänglich gemacht.

ECS generieren nur dann einen echten Mehrwert, wenn sie in ihrem jeweiligen konkreten Anwendungskontext nachweislich wirksam und effizient sind und belastbare Vorteile gegenüber anderen Lösungen bieten. Zur Erzeugung dieser Evidenz ist eine zielgerichtete Bewertung von ECS erforderlich. Eine Bewertung von ECS ist jedoch mit besonderen Herausforderungen verbunden: ECS resultieren aus vielfältigen Hardware- und Softwarekomponenten, die in verschiedensten Konstellationen integriert sind. Entsprechend ergeben sich diverse Topologien von ECS. Zudem können sich diese Komponenten mit fortlaufender Zeit verändern, etwa, indem Hardware ausgetauscht oder Softwaredienste aktualisiert werden. Durch diese Heterogenität und Dynamik fällt es einerseits schwer, die für den jeweiligen Anwendungskontext relevanten Bewertungskriterien aus den Bereichen Technologie, Organisation und Umwelt zu identifizieren. Andererseits müssen zu diesen Bewertungskriterien die geeigneten Ansätze und Metriken gewählt werden, um sie für eine spezifische Auslegung eines ECS zu beurteilen. Genau an dieser Stelle fehlt es derzeit an Hilfestellungen, die eine systematische Evaluation von ECS für die Praxis vereinfachen.

Um diese Herausforderung gemeinsam anzugehen, haben sich die Förderprojekte des Technologieprogramms Edge Datenwirtschaft EASY, EDNA, ESCOM, DEER, und SECAI sowie die Projekte ECO:DIGIT (GreenTech Innovationswettbewerb) und GEMIMEG-II (Smart Service Welt II) zu einer Task-Force zusammengeschlossen. In diesen Projekten, die vom Bundesministerium für Forschung, Technologie und Raumfahrt (BMFTR) gefördert werden, arbeiten Expertinnen und Experten aus verschiedenen Branchen an Innovationen entlang des Edge-Cloud-Kontinuums. Ziel der Task-Force war es, die Erkenntnisse der Early-Adopter-Projekte bei der Bewertung von ECS zu diskutieren und zu konsolidieren. Das vorliegende Dokument präsentiert die Ergebnisse der Task-Force und bietet somit eine erste Orientierungshilfe im Themenbereich der Bewertung von ECS.

Die Orientierungshilfe gibt zunächst einen über die Projekte hinweg konsolidierten Überblick über die Grundlagen der Bewertung von ECS. Hierzu gehören die relevante Terminologie und Definitionen, ein Vorgehensmodell zur Durchführung der Bewertung von ECS und eine ganzheitliche

Übersicht über mögliche Bewertungskriterien (Abschnitt 2). Anschließend werden ausgewählte Bewertungskriterien im Detail betrachtet. Dazu stellen die Expertinnen und Experten anhand von individuellen Erfahrungen und Beispielen aus ihren Projekten vor, wie einzelne Bewertungskriterien in der Praxis analysiert und operationalisiert werden können (Abschnitt 3). Abschnitt 4 fasst die vorhabensübergreifend relevanten Herausforderungen bei der Bewertung von ECS zusammen. Abschließend präsentiert diese Orientierungshilfe gemeinsam erarbeitete Handlungsempfehlungen an Systemintegratoren, Hardware-, Software- und Plattformbetreibende sowie Forschung und Politik zur Unterstützung der Bewertung von ECS (Abschnitt 5).

Diese Orientierungshilfe richtet sich insbesondere an Systemarchitektinnen und -architekten und Entwicklungsabteilungen in Unternehmen, die sich mit der Entwicklung und Bewertung von ECS befassen. Darüber hinaus ist diese Orientierungshilfe für Führungskräfte, Projektmanagement-Teams und weitere Akteure relevant, die sich einen allgemeinen Überblick über die Bewertung von ECS verschaffen möchten. Zuletzt adressiert diese Studie alle im Rahmen der Handlungsempfehlungen genannten Akteure, darunter auch die Politik und Forschende, die mit dieser Orientierungshilfe Anregungen zur Weiterentwicklung des aktuellen Status quo der Bewertung von ECS erhalten.

Die Orientierungshilfe wurde im Rahmen der Begleitforschung zum Technologieprogramm „Edge Datenwirtschaft“ des Bundesministeriums für Forschung, Technologie und Raumfahrt (BMFTR) erstellt.

2 Grundlagen der Bewertung von Edge-Cloud-Systemen: Definitionen, Vorgehen und Bewertungskriterien

Um eine Bewertung von Edge-Cloud-Systemen (ECS) durchführen zu können, bedarf es zunächst eines gemeinsamen Verständnisses über die Bestandteile eines ECS als Bewertungsgegenstand. Weiterhin muss geplant werden, wie die Bewertung konkret durchzuführen ist und welche Eigenschaften anhand welcher Bewertungskriterien beurteilt werden sollen. Dieser Abschnitt beschreibt die notwendigen Grundlagen und gliedert sich dementsprechend in drei Teile: eine Darstellung der relevanten Bestandteile von ECS, die Vorstellung eines Vorgehensmodells für die Durchführung der ECS-Bewertung (Green Software Measurement Modell, GSMM) sowie eine strukturierte Übersicht über potenziell relevante Bewertungskriterien.

2.1 Definition und Terminologie von Edge-Cloud-Systemen

ECS sind hybride Rechnerarchitekturen, die auf verschiedene Elemente des Edge-Cloud-Kontinuums (ECC) zurückgreifen, um die Ressourcennutzung zu optimieren, die Latenz zu verringern und die Leistung von verschiedenen Anwendungen zu verbessern. Das ECC setzt sich aus verschiedenen Ressourcen zusammen, die Daten entlang der Datenwertschöpfungskette verarbeiten (Strnadl, 2021). Diese Ressourcen reichen von IoT-Geräten bis zur Cloud-Infrastruktur. Je nach Referenz unterscheiden sich der Detailgrad und die Abstufungen in der Beschreibung des ECC. Bei der Klassifizierung der Elemente des ECC stellen unter anderem die Verantwortung des Infrastrukturbetriebs oder die zugrunde liegende Hardware wichtige Unterscheidungskriterien dar.

Abbildung 1 fasst ausgewählte Einteilungen des ECC zusammen. Die Linux Foundation (LF) (2020) führt eine Gliederung des ECC anhand des Ortes der Datenverarbeitung und der zugrunde liegenden Hardware durch. Geräte der „Constrained-Device-Edge“ basieren auf typischen Microcontrollern mit eingebetteter Software. In der „Smart-Device-Edge“ kommen x86- oder ARM-Architekturen zum Einsatz. Ab dem Bereich „On-Premise-Data-Center-Edge“ werden Serverarchitekturen verwendet. Willner und Gowtham (2020) klassifizieren das ECC detaillierter nach der Hierarchiestufe und gliedern die User-Edge stärker auf. Die Einordnung von Strnadl (2021) weist die größte Detaillierung auf. Insbesondere im Bereich der „Smart-Device-Edge“, die in räumlicher Nähe zu den Anwendenden betrieben wird, werden weitere Typen von Edge-Geräten unterschieden.

Entlang des ECC ermöglichen Edge-Geräte die lokale Datenverarbeitung und -speicherung sowie die lokale Ausführung von Diensten. Die Cloud stellt in dieser Architektur prinzipiell die gleichen Funktionen mit höheren Kapazitäten, aber auch größeren Latenzen, bereit. Edge-Knoten an der „User-Edge“ können von einzelnen Mitgliedern des ECC, wie beispielsweise Industrieunternehmen, betrieben werden. Die zentrale Cloud-Plattform wird üblicherweise durch eine separate Einheit, beispielsweise durch einen Hyperscaler, zur Verfügung gestellt. Daten und Berechnungen können im ECC flexibel zwischen den Edge-Geräten und der Cloud verschoben werden. Dadurch kann die Übertragung von Daten in die Cloud minimiert werden, da diese größtenteils von Edge-Rechenressourcen übernommen wird. Andersherum werden rechenintensive oder weniger zeitkritische Anfragen auf der Cloud-Infrastruktur bearbeitet. Somit lassen sich die durch die

Datenübertragungen anfallenden Kosten und Verzögerungen reduzieren und sowohl echtzeitkritische als auch rechenintensive Anfragen umsetzen, ohne alle für die Datenverarbeitung notwendigen Ressourcen an jeder Maschine vorzuhalten. Die Cloud übernimmt zudem die Orchestrierung des ECC. Sie bestimmt, wo Prozesse innerhalb des Kontinuums durchgeführt werden. Durch die Architektur des ECC können auch Datenschutzaspekte adressiert werden, da Unternehmen bestimmte Daten lokal verarbeiten und nicht in unternehmensübergreifende oder externe Cloud-Lösungen übertragen müssen (Schultheis, Alt et al., 2024).

Location nach LF	User Edge							Service Provider		Cloud Provider
Linux Foundation	Constrained Device Edge	Smart Device Edge				On-Premise Data Center Edge		Access Edge	Regional Edge	Centralized Data Centers
Willner & Gowtham	Product Edge	Deep Edge	Gateway Edge		Network Edge	Private Edge		Collocation Edge		Cloud Computing
Strnadl	Device Edge	Control Edge	Thin Gateways	Thick Gateways	Industrial/Telco Edge	Micro Data Center	On-Premise Data Center	Far-Edge Cloud	Metro-Level Edge	Cloud

Abbildung 1: Gegenüberstellung der verschiedenen Detaillierungen des Edge-Cloud-Kontinuums

2.2 Exemplarisches Vorgehensmodell zur Bewertung von Edge-Cloud-Systemen

Zur Bewertung von ECS existiert derzeit noch kein etabliertes kontextspezifisches Vorgehensmodell. Dieser Abschnitt präsentiert mit dem „Green Software Measurement Modell“ (GSMM, Guldner et al. (2024)) ein Referenzmodell zur Bewertung von Softwaresystemen. Ursprünglich im Kontext der Bewertung von Energie- und Ressourcenverbrauch durch ein internationales Forschungsteam entwickelt, kann dieses, wie im weiteren Verlauf gezeigt, auf den Kontext von ECS appliziert werden und einen strukturierten Ablauf der Bewertung gewährleisten.

Die Bewertung und Evaluation von ECS ist generell und insbesondere im Produktionsumfeld eine komplexe Aufgabenstellung. Zu Beginn ist daher die Frage zu stellen, was die Ziele der Bewertung sind. Typische Fragestellungen umfassen dabei die Sicherheit und Privatheit eines ECS, seine Effektivität, Ressourcen- und Dateneffizienz und seine strategische Verbesserung. Zudem muss zwischen einer Neu- und einer Weiterentwicklung eines ECS sowie der (gegebenenfalls externen) Bewertung eines bestehenden ECS unterschieden werden. Dabei ist zu beachten, dass die Vielzahl von ECS-Architekturen direkte Vergleiche zwischen verschiedenen ECS erschwert (Guldner et al., 2024). Daher ist eine Bewertung besonders für den Ist-Zustand hilfreich, um Vergleiche zu anderen Versionen eines ECS zu ermöglichen sowie die Optimierung während der Entwicklung zu unterstützen.

Abbildung 2 verdeutlicht exemplarisch das Vorgehen, wie Bewertungskriterien der Energie- und Ressourceneffizienz von Softwaresystemen identifiziert und erhoben werden können. Typische Schritte sind dabei (A) die Identifikation der Ziele der Messung (der Bewertungskriterien) sowie der zu messenden Bestandteile des Systems, (B) die Definition der aufzunehmenden Maße und Metriken (Kennzahlen), (C) die Festlegung des Messablaufs, (D) der Aufbau der Messumgebung und (E) die Evaluation der Messung. Dabei ist zu beachten, dass eine solche Bewertung zunächst immer eine Momentaufnahme ist, die gegebenenfalls für sich genommen wenig aussagt (bspw.

kWh pro Prozess, Datenmenge oder auch Verschlüsselungsoptionen), aber im Verlauf der (Weiter-)Entwicklung einen Pfad zu ressourceneffizienteren ECS aufzeigen kann.

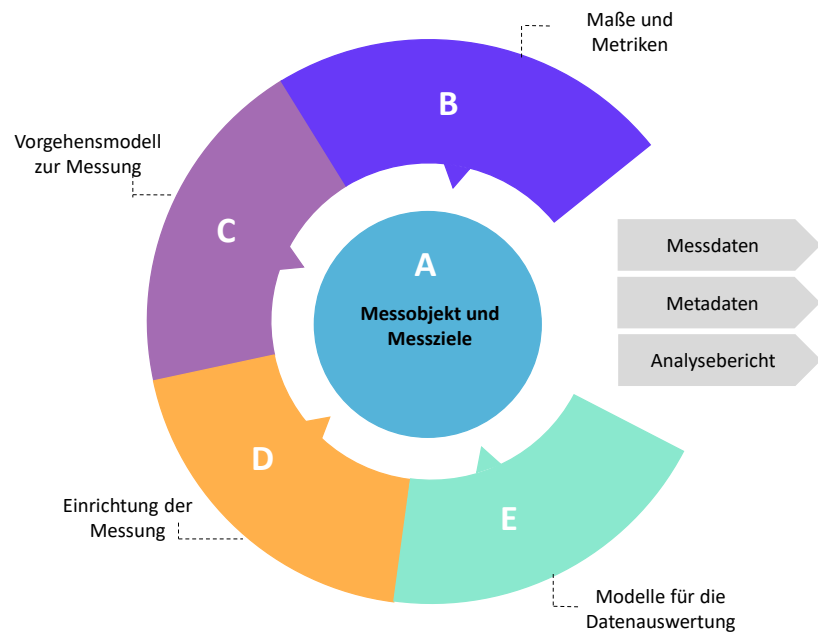


Abbildung 2: Green Software Measurement Modell (GSMM) nach Guldner et al. (2024)

Entsprechend ist ein kontinuierliches Monitoring notwendig, um Unternehmensvorteile trotz der zunächst aufwändigeren Einrichtung der Mess- und Analysekomponenten für das ECS erzielen zu können. Gleichzeitig führen dieses Monitoring und die Diskussion der Ergebnisse in Fachforen, Branchenvereinigungen und auf Konferenzen im Austausch zwischen Wirtschaft und Wissenschaft zu einer Weiterentwicklung der ECS und ihrer Bewertungsmethoden insgesamt. Mittelfristiges Ziel ist die automatisierte und in die Komponenten integrierte Bewertung mittels standardisierter Methoden. Dabei ist immer zu berücksichtigen, dass eine Betrachtung über den gesamten Lebenszyklus eines ECS komplex ist: Die Rolle und die Ressourcenbedarfe der industriellen Assets können sich verändern, die eingesetzte Hardware über die Zeit weiterentwickeln, die Kosten der Datenübertragung in die Cloud verringern oder erhöhen und die Cloud-Verteilung angepasst werden. Die Bewertung muss daher auf das jeweilige ECS zugeschnitten sein und wird in der Praxis naturgemäß Unvollständigkeiten aufweisen.

Anwendungsbeispiel: Bewertung eines ECS zur Fehlererkennung im 3D-Druck

Für den konkreten Anwendungsfall zur Erfassung der Energie- und Dateneffizienz eines Systems zur Fehlererkennung im 3D-Druck kann das GSMM als Orientierung dienen. Es zeigt exemplarisch, wie Bewertungskriterien identifiziert und erhoben werden können.

Um die Dateneffizienz eines bildbasierten Fehlererkennungssystems im 3D-Druck mithilfe des GSMM zu bestimmen, wird zunächst das zu messende Objekt definiert. Hierbei kann es sich um den gesamten Erkennungsprozess oder nur um spezifische Module, wie die Bildvorverarbeitung oder das Modelltraining, handeln. Das Hauptziel besteht in der Optimierung der Dateneffizienz. Dabei soll die für eine zuverlässige Fehlererkennung erforderliche Datenmenge minimiert werden, ohne die Erkennungsgenauigkeit zu beeinträchtigen. Im nächsten Schritt werden die relevanten Metriken festgelegt. Für ein bildbasiertes Fehlererkennungssystem sind hier die Daten-

menge pro Bild, die Rechenleistung, der Speicherverbrauch, die Verarbeitungszeit sowie die Erkennungsrate (Präzision, Recall, F1-Score) und der Energieverbrauch pro Verarbeitungsschritt relevant. Die Metriken dienen dazu, den Einfluss der Datenmenge auf die Leistung und die Energieeffizienz des Systems sichtbar zu machen.

Anschließend wird das Messverfahren ausgewählt, angepasst oder, sofern kein passendes Verfahren existiert, neu entwickelt. Es ist sinnvoll, verschiedene Szenarien zu untersuchen, die einen Einfluss auf das Messziel und damit mögliche Optimierungsansätze haben. Beispiele hierfür sind der Einfluss unterschiedlicher Bildauflösungen auf die Erkennungsgenauigkeit und den Energieverbrauch. Ein weiterer Vergleich könnte zwischen der Trainingsphase des Systems und dessen Anwendung in Echtzeit erfolgen, um Unterschiede im Datenverbrauch und der Energieeffizienz zu identifizieren. Auch die Effizienz der Verarbeitung mehrerer Bilder gleichzeitig gegenüber der sequenziellen Einzelbildverarbeitung könnte Gegenstand der Untersuchung sein. Im Anschluss wird der Messaufbau definiert. Dies umfasst die Auswahl der benötigten Hardware wie Kameras, 3D-Drucker und Bildverarbeitungssoftware sowie der erforderlichen Messsoftware. Externe Strommessgeräte können den Energieverbrauch während des Betriebs überwachen, während interne Monitoring-Tools die Hardwareauslastung aufzeichnen. Nach Abschluss der Messungen erfolgt die Auswertung der gesammelten Daten. Besonders wichtig ist dabei die Analyse der Korrelation zwischen der Datenmenge und der Erkennungsgenauigkeit sowie des Energieverbrauchs. Statistische Auswertungen helfen, den optimalen Punkt zwischen Datenreduktion und Effizienz zu bestimmen.

Zum Abschluss werden die Ergebnisse in einem Bericht zusammengefasst, der aufzeigt, welche Optimierungsmöglichkeiten zur Steigerung der Dateneffizienz bestehen. Dies könnte beispielsweise die Reduktion der Bildgröße sein, die keine signifikanten Einbußen bei der Fehlererkennung verursacht. Zusätzlich könnten Empfehlungen zur Optimierung der Hard- und Software gegeben werden, um die Effizienz des 3D-Druck-Prozesses insgesamt zu steigern.

2.3 Relevante Kriterien zur Bewertung von Edge-Cloud-Systemen

Ein zentraler Aspekt der Bewertung von ECS ist die Festlegung der während der Bewertung betrachteten Kriterien (im GSMM: Messziele) und der zugehörigen Metriken, mit denen diese Kriterien bewertet werden sollen. Dieser Abschnitt bietet eine im Rahmen der Task-Force erstellte konsolidierte Übersicht über mögliche Bewertungskriterien und deren Metriken als Orientierungshilfe in der Bewertungsplanung. Zunächst werden relevante Bewertungskriterien strukturiert erläutert. In einer abschließenden Tabelle werden diesen Kriterien relevante Metriken zugeordnet.

Die Bewertungskriterien für ECS lassen sich entlang der Kategorien „Technologie“, „Organisation“ und „Umwelt“ klassifizieren. Während im wissenschaftlichen Diskurs zunächst insbesondere technische Kriterien im Fokus standen (Aslanpour et al., 2020), werden zunehmend auch organisatorische und Umweltkriterien zu einem wichtigen Betrachtungsgegenstand in Forschung und Praxis. Abbildung 3 visualisiert mögliche Kriterien zur Bewertung von ECS in den Kategorien Technologie, Organisation und Umwelt.

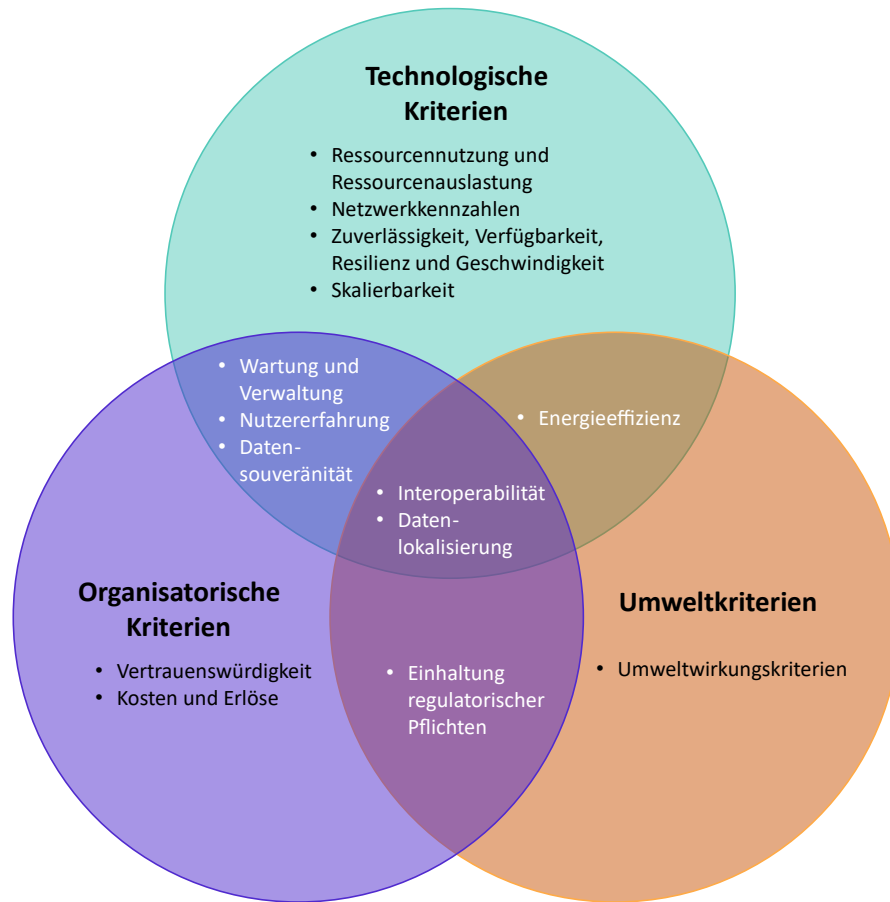


Abbildung 3: Bewertungskriterien von Edge-Cloud-Systemen

Die technologische Dimension umfasst Kriterien, welche die Eigenschaften des technischen Systems selbst charakterisieren. Für diese Kriterien existieren üblicherweise konkrete (objektive) Kennzahlen und Metriken, die sich durch bestimmte Messmethoden quantifizieren oder zumindest schätzen lassen. Zu den technologischen Kriterien zählen die Nutzung und Auslastung der bereitgestellten physischen und virtuellen Ressourcen, zu denen Elemente wie Grafikprozessoren und Speicher gehören. Auch das Netzwerk eines ECS kann anhand eigener Kennzahlen bewertet werden. Hierbei werden beispielsweise Metriken wie die benötigte Bandbreite oder Latenzzeiten berücksichtigt. Darüber hinaus ist die Zuverlässigkeit und Verfügbarkeit eines ECS ein wesentliches Kriterium, das insbesondere im Bereich kritischer Systeme zu bewerten ist. Die Skalierbarkeit eines ECS beschreibt, inwiefern das System auf schwankende Anforderungen hinsichtlich der Nutzlast reagieren kann. Die Resilienz des operativen ECS ist ebenfalls ein wichtiges technisches Kriterium.

Organisatorische Kriterien beschreiben insbesondere die menschlichen, kulturellen und strukturellen Aspekte, die bei der Einführung von Edge- und Cloud-Technologien eine Rolle spielen. Ebenso werden hier Kostenüberlegungen zugeordnet, da diese häufig mit der internen Entscheidungsfindung, Budgetierung und Ressourcenallokation innerhalb einer Organisation verbunden sind. Die Bewertung organisatorischer Kriterien erfolgt oftmals auf Basis theoretischer Konstrukte in Form von Befragungen, Beobachtungen oder Fallstudien und anhand subjektiver Metriken. So kann beispielsweise die Einschätzung der Vertrauenswürdigkeit eines Systems zwischen verschiedenen Organisationen und Anwendungskontexten divergieren, da diese von individuellen Erfahrungen der Bewertenden abhängt.

Die Umweltdimension umfasst Aspekte, die sich aus dem politischen und gesellschaftlichen Umfeld ergeben, in dem eine Organisation tätig ist. Hierzu gehören Anforderungen des Marktes ebenso wie der zutreffende regulatorische Rahmen. Zusätzlich zu den Umweltkriterien, die sich auf das System im Betrieb beziehen, existieren darüberhinausgehende Umweltwirkungskriterien. Diese beschreiben die über das gesamte System hinweg aggregierten Umweltwirkungen über den gesamten Lebenszyklus, der neben der eigentlichen Nutzungsphase auch die Herstellungsphase und die Entsorgung umfasst. Bei der Bewertung von Umweltkriterien wird in der Praxis meist geprüft, ob exogene, z.B. durch die Gesetzgebung vorgegebene Anforderungen erfüllt werden.

Viele der Bewertungskriterien lassen sich nicht eindeutig einer Kategorie zuordnen, sondern sind an der Schnittstelle von Technologie, Organisation und Umwelt angesiedelt. Dazu gehören Bewertungskriterien wie der Aufwand für die Pflege und Verwaltung des ECS, die Benutzererfahrung oder die Datenhoheit, die sowohl technische als auch organisatorische Mechanismen umfassen. Die Einhaltung regulatorischer Pflichten wird durch den Gesetzgeber vorgegeben und ist daher als Umweltkriterium zu sehen. Die konkrete Umsetzung dieser Vorgaben wird jedoch oftmals durch die organisatorischen Rahmenbedingungen beeinflusst. Die Energieeffizienz von Softwaresystemen ist eine Eigenschaft, die zunehmend von Regulatorik und Gesellschaft gefordert wird und sich mit technischen Mechanismen messen oder berechnen lässt. Die Bewertungskriterien Interoperabilität und Datenlokalisierung stehen im Schnittpunkt aller drei Bewertungsdimensionen. Das Kriterium Interoperabilität lässt sich dabei, gemäß dem European Interoperability Framework in die Bereiche rechtliche Interoperabilität, organisatorische Interoperabilität, technische Interoperabilität und semantische Interoperabilität unterteilen (European Commission, 2017). Zur Operationalisierung der abgebildeten Bewertungskriterien sind konkret messbare beziehungsweise bewertbare Kennzahlen oder Metriken zu verwenden. Tabelle 1 zeigt eine Übersicht über relevante Bewertungskriterien und zugehörigen Metriken für ECS. Die Tabelle verweist zudem auf die relevanten Absätze in Abschnitt 3, in denen die Expertinnen und Experten der Task-Force Einblicke in praktische Ansätze zur Bemessung der jeweiligen Kriterien geben. Weiterhin werden ausgewählte Referenzen aus der Wissenschafts- und Fachliteratur gelistet, die eine ausführlichere Erläuterung der Metriken liefern.

Bewertungskriterium	Metriken	Absatz	Weitere Referenzen (Auswahl)
Ressourcennutzung und -auslastung	CPU-/ GPU- und Speicherverbrauch Ressourcennutzung (Bspw. Container oder VM) Ressourcenauslastung	3.1	Das et al. (2018) Gill et al. (2018) Aslanpour et al. (2020)
Netzwerkkenzahlen	Bandbreite Durchsatz (Datenrate) Latenz Jitter Packet-Loss-Ratio Netzwerküberlastung	3.2	Rouf et al. (2024) Zhu et al. (2018) Aslanpour et al. (2020) Luber und Donner (2018) Herlich et al. (2017) Slattery (2019)
Zuverlässigkeit, Verfügbarkeit, Resilienz und Geschwindigkeit	Fehlertoleranz Verfügbarkeit Antwortzeit Anzahl beschädigter Aufgaben Zuverlässigkeit = mittlere Betriebsdauer zwischen Ausfällen + mittlere Wiederherstellungszeit	3.2 3.6	Wang et al. (2024) Gill et al. (2018) Maciel et al. (2022) Aslanpour et al. (2020) Gill et al. (2018)
Skalierbarkeit	Elastizität Effizienz Lastenverteilung	3.5	Coutinho et al. (2015) Hu et al. (2024)
Wartung und Verwaltung	Installationsaufwand Wartungsaufwand Benötigte Kompetenzen		Lenhard et al. (2013) Souza et al. (2024) ---
Nutzererfahrung	Nutzerzufriedenheit Net-Promoter-Score		Li et al. (2022) Kühnapfel (2021)
Datensouveränität	Nutzung Privacy-Enhancing-Technologies Föderiertes Lernen Zero-Knowledge-Proofs	3.4 3.7	Cha et al. (2019) McMahan et al. (2016) Sheybani et al. (2025)
Vertrauenswürdigkeit	Verwendung von Standards (ja/nein) Zertifizierung (ja/nein) Attestierung durch Drittparteien (ja/nein)	3.7	Karagiannis et al. (2025) Omar et al. (2012) Mo et al. (2020)
Kosten und Erlöse	Gesamtbetriebskosten Kapitalrendite Break-Even Analyse	3.8	Horsch (2023); Mondello (2022)
Interoperabilität	Kompatibilitätstests (ja/nein) Zertifizierung (ja/nein)		European Commission (2017)
Datenlokalisierung	Ort der Datenspeicherung und Verarbeitung		Baudoin (2018)
Einhaltung regulatorischer Pflichten	Compliance zu DSGVO / Data Act / AI Act / ... (ja/nein)	3.9	Straub und Bogenstahl (2024)
Energieeffizienz	Energieverbrauch Emissionen (bspw. CO ₂ -Äquivalente [CO ₂ e])	3.3	Ahvar et al. (2022) AIOTI (2024)
Umweltwirkung	Kumulierter Energiebedarf Treibhauspotenzial Abiotischer Ressourcenverbrauch Wasserverbrauch Elektro- und Elektronik-Altgeräte (Waste Electrical and Electronic Equipment [WEEE])	3.1	Gröger et al. (2024)

Tabelle 1: Relevante Bewertungskriterien und zugehörige Metriken in Edge-Cloud-Systemen identifiziert in der Task-Force

3 Methoden, Instrumente und Metriken zur Bewertung von Edge-Cloud-Systemen – Einblicke in die Praxis

Dieser Abschnitt stellt dar, wie ausgewählte Bestandteile und Bewertungskriterien von ECS in der Praxis anhand von spezifischen Metriken evaluiert werden können. Die hier vorgestellten Ansätze präsentieren den aktuellen Stand der Technik und darüberhinausgehende Forschungs- und Entwicklungsarbeiten, die den in der Einleitung beschriebenen Förderprojekten im Bereich „Entwicklung digitaler Technologien“ des BMFTR¹ entstammen. Abbildung 4 fasst die in dieser Publikation behandelten Themenbereiche und Bewertungsgrößen von ECS zusammen und listet die Projekte, die sich mit den jeweiligen Themen befassen.

Themen und Kategorien der Bewertung	Computing-Plattformen (3.1)	Datenübertragung (3.2)	Energieverbrauch und CO ₂ -Emissionen (3.3)	Künstliche Intelligenz <ul style="list-style-type: none"> Föderiertes Lernen (3.4) KI-Planung (3.5) Fallbasiertes Schließen (3.6) 	Privatsphäre, Datensouveränität und Vertrauen (3.7)	Kostenfaktoren (3.8)	Rechtliche Aspekte (3.9)
Bewertungsgrößen	Gesamtenergieverbrauch, Ressourcenauslastung, Umweltwirkungsmetriken	Latenz, Packet-Loss-Ratio, Durchsatz, Bandbreite, Netzüberlastung, Laufzeitvarianz	Energieverbrauch in Produktion und Logistik, Produktbezogener CO ₂ -Fußabdruck, Netzwerkauslastung, Speicherplatzbedarf, Skalierbarkeit	Genauigkeit, Zeit, Energieverbrauch, Planungskosten, Lösungsqualität	Qualitative Betrachtung der Kriterien	Gesamtkosten	Rechtliche Einschätzung der Compliance
Projekte	ECO:DIGIT	EASY, ESCOM, GEMIMEG-II	EDNA	EASY, SECAI	DEER, EASY, ESCOM	Begleitforschung Edge Datenwirtschaft	Begleitforschung Edge Datenwirtschaft

Abbildung 4: Themenbereiche der Bewertung von ECS und zugehörige Bewertungsgrößen in dieser Publikation

Abschnitt 3.1 stellt vor, wie mithilfe eines Prüfstandes die Ressourcen- und Energieverbräuche von Softwarediensten in Abhängigkeit von der zugrundeliegenden Computing-Plattform analysiert und optimiert werden können, um Softwarelösungen in Bezug auf ihre ökologische Effizienz zu bewerten. Zur Bewertung der Qualität der Datenübertragung im Netzwerk zwischen einzelnen Knoten eines ECS können sowohl Messverfahren als auch Verfahren der Künstlichen Intelligenz (KI) zur Schätzung von Netzwerkparametern eingesetzt werden. Abschnitt 3.2 beschreibt ein beispielhaftes Verfahren zur Abschätzung von Netzwerkparametern und erläutert die daraus folgenden Verteilungsentscheidungen von Softwarediensten im Edge-Cloud-Kontinuum (ECC). Die Reduktion von Ressourcenverbräuchen bei der Nutzung von Softwarediensten und den darauf basierenden Anwendungen stellt einen der größten Motivationsfaktoren zum Einsatz von Edge-Computing dar. Abschnitt 3.3 erläutert, wie CO₂-Äquivalente (CO₂e-Emissionen) im Kontext industrieller Produktions- und Logistikanwendungen erfasst werden können. Zudem wird gezeigt, wie Edge-Computing in diesen Bereichen eine präzise CO₂-Berechnung ermöglicht und die Umweltwirkung der Anwendungen verringert. ECS stellen die Plattform für Anwendungen im Bereich

¹ Mehr Informationen zum Förderbereich und zu den Projekten erhalten Sie unter <https://www.digitale-technologien.de/DT/Navigation/DE/Home/home.html> oder direkt über die Verlinkungen in der Einleitung

der KI dar. Auch die KI-Verfahren selbst sollten im Zuge der ganzheitlichen Systembewertung evaluiert werden. Diese Publikation fokussiert sich auf die Bewertung von Verfahren des Föderierten Lernens (FL) (Abschnitt 3.4), der KI-Planung (Abschnitt 3.5) sowie des Fallbasierten Schließens (CBR) (Abschnitt 3.6). Ein weiterer Motivationsfaktor zum Einsatz von ECS stellt die Verbesserung von Privatsphäre und Datensouveränität der Datengebenden sowie eine Erhöhung des Vertrauens in die bereitgestellten Daten dar. Abschnitt 3.7 führt dazu eine qualitative Betrachtung von Ansätzen durch, die eine Verbesserung dieser Kriterien erzielen können. Abschnitt 3.8 beschreibt relevante Kostenkriterien zur monetären Bewertung von ECS. Abschließend wird eine Einschätzung hinsichtlich der zu bewertenden rechtlichen Anforderungen bei der Umsetzung von ECS gegeben (Abschnitt 3.9).

3.1 Bewertung von Computing-Plattformen

Anwendungsbereich	Verschiedenartige Software in ECS
Bewertungskriterien	Energieeffizienz, Ressourcennutzung und -auslastung, Umweltwirkung
Kennzahlen und Metriken	Gesamt-Energieverbrauch (kWh), Ressourcen-Auslastung (%), verschiedene Metriken zur Umweltwirkung (bspw. Global Warming Potential GWP in kg CO ₂ e, Waste Electrical and Electronic Equipment WEEE in kg)
<ul style="list-style-type: none"> ▪ Die Bewertung einer Software hinsichtlich ihres Energie- und Ressourcenverbrauchs stellt Transparenz und Objektivität zu den mit dem Einsatz der Software verbundenen Umweltwirkungen her. Dabei sind die über die reine Nutzungsphase hinaus auch mit der eingesetzten Hardware verbundenen Umweltwirkungen relevant und zu berücksichtigen. ▪ Im Projekt ECO:DIGIT wird ein Prüfstand geschaffen, der eine leicht durchzuführende Bewertung ermöglichen soll. Dazu sind außer der Bereitstellung der zu untersuchenden Software, den Lastszenarien und einer konkreten Infrastrukturddefinition keine weiteren Angaben notwendig. ▪ Durch den vielfachen Einsatz des Prüfstands zur Bewertung unterschiedlichster Software kann im Laufe der Zeit ein Trainingsdatensatz entstehen, der es KI zukünftig erlaubt, sowohl Code-Optimierungen als auch optimierte Hardware-Setups vorzuschlagen. 	

Mit der Ausführung von Software geht immer auch eine Nutzung von (Computing-) Systemressourcen sowie ein Verbrauch von Energie und anderen Ressourcen wie Wasser (zum Kühlen) einher. Dies gilt für verschiedenste Software, wenn auch in unterschiedlichem Ausmaß: Eine Smartphone-App verbraucht beispielsweise bei der Ausführung gespeicherte Energie aus der Smartphone-Batterie, während großskalige Cloud-Anwendungen bei der Ausführung im Rechenzentrum Strom und Wasser verbrauchen. Die genannten Beispiele beziehen sich auf den Ressourcenverbrauch (Energie, Wasser) während der Nutzungsphase. Hinzu kommen weitere Umweltwirkungen aus den anderen Lebensphasen der beteiligten Hardware, insbesondere aus der Herstellungs- und der Entsorgungsphase, die es zu berücksichtigen gilt.

Das Ziel des Projektes ECO:DIGIT ist es, einen Prüfstand zu entwickeln, der die Bewertung unterschiedlicher Software hinsichtlich ihres Ressourcen- und Energieverbrauchs transparent und auf einfache Weise verfügbar macht. Dabei soll ein manueller Messvorgang (von Energie- und Ressourcennutzung der Software) durch ein automatisiertes Berechnen der gewünschten Größen ersetzt werden, was den Bewertungsvorgang wesentlich vereinfacht. Dem liegt eine entsprechende Modellierung der Vorgänge zugrunde.

Die Bewertung erfolgt dabei anhand einer vom Öko-Institut erarbeiteten Methodik (s. Gröger et al. (2024)), die die Ökobilanzierung einer digitalen Lieferkette beschreibt (s. Abbildung 5). Die digitale Lieferkette setzt sich aus allen bei der Auslieferung der digitalen Dienstleistung beteiligten Hardware-Plattformen zusammen. Zur Erstellung der Ökobilanz müssen alle Glieder der digitalen Lieferkette bewertet und geeignet zugeordnet werden.

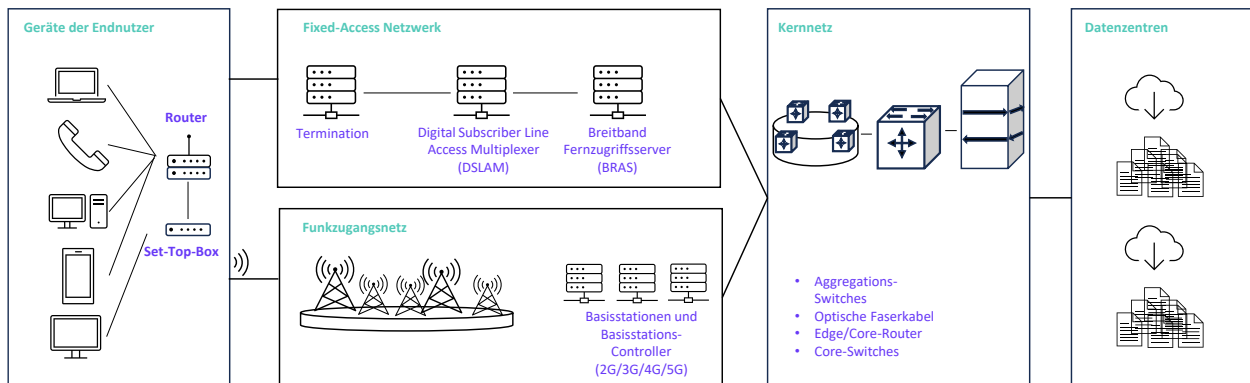


Abbildung 5: Die digitale Lieferkette nach Gröger et al. (2024)

Neben Cloud-Anwendungen rücken dabei in dem Projekt auch Szenarien aus dem industriellen Edge-Computing Bereich in den Fokus. Ziel ist es, mit dem Prüfstand eine möglichst genaue und zuverlässige Bewertung des i) Energieverbrauchs und ii) der mit den eingesetzten Hardware-Geräten verbundenen Umweltwirkungen für u. a. Edge-Computing-Anwendungen zu ermöglichen.

Damit soll es in Zukunft möglich sein, für vorhandene Softwarelösungen mit einfachen Mitteln eine Bewertung durchzuführen und somit eine objektive Aussage zu den mit dem Einsatz der Software einhergehenden Umweltwirkungen zu erhalten. Darauf aufbauend können auch Entwickelnde und Systemarchitektinnen und -architekten davon profitieren, da sie mit einer in den Entwicklungsprozess integrierten Bewertung ihre Implementierung bereits im Entstehungsprozess kontinuierlich hinsichtlich ihres Energiebedarfs überwachen und weiter optimieren. Darüber hinaus ist es denkbar, dass im Beschaffungsprozess von Softwareprodukten neben anderen Vergleichskriterien wie Preis, Funktionalität etc. die ökologische Bewertung auf Basis des Prüfstands als weitere Bewertungsdimension hinzukommt. Schließlich wird der Prüfstand auch als Simulationsplattform dienen, um unterschiedlichste Konfigurationen in Form eines digitalen Zwillings, d. h., ohne die Notwendigkeit, diese real aufbauen zu müssen, bewerten und optimieren zu können (Gröger et al., 2024).

3.2 Analyse und Optimierung der Datenübertragung

Anwendungsbereich	Datenübertragung und Netzwerkkonfigurationen in Edge-Cloud-Systemen
Bewertungskriterien	Netzwerkkennzahlen, Zuverlässigkeit, Geschwindigkeit
Kennzahlen und Metriken	Latenz, Packet-Loss-Ratio, Durchsatz, Bandbreite, Netzüberlastung (Congestion), Laufzeitvarianz (Jitter)
<ul style="list-style-type: none"> ▪ In ECS spielt die Bewertung der Datenübertragung oft eine entscheidende Rolle sowohl für die Performanz als auch für die Entscheidung, welche Anwendungen und Dienste wo ausgeführt werden können. ▪ Die Messung von Kennzahlen zur Beurteilung der Datenübertragung ist zwar grundsätzlich möglich, häufig jedoch mit zusätzlichem Aufwand und Einbußen in der Performanz der Datenübertragung verbunden. 	

- Durch den zunehmenden Einsatz von Technologien aus den Bereichen der KI und Heuristik ergeben sich auch bei der Bewertung der Datenübertragung neue Möglichkeiten zur Schätzung von Metriken, die es ermöglichen, Softwareverteilungen in ECS zu bewerten.

Eine effiziente und performante Datenübertragung ist der Schlüssel für leistungsfähige ECS. Typischerweise nehmen Edge-Systeme Daten aus der Umgebung auf. Dies geschieht sowohl direkt als intelligentes lokales Monitoring- oder Sensorsystem über eingebaute Controller oder Sensoren als auch über sogenannte Constrained-Devices, also sehr einfache, aber spezifische Geräte mit sehr eingeschränkter Rechen- und Kommunikationsleistung, die mit einem Edge-System verbunden sind. Ein ECS verfügt in den unterschiedlichen Ebenen daher über Hardwarekomponenten, die in Bezug auf Rechenleistung und Kommunikation mit benachbarten Systemen unterschiedliche Anforderungsprofile erfüllen und entsprechende Leistungsparameter haben. Das Netzwerk zwischen diesen Hardwarekomponenten ist nicht nur das notwendige funktionale Bindeglied der diversen ECS-Komponenten. Ein leistungsfähiges und auch adaptives Netzwerk kann über seine Funktionalität, Performanz und Resilienz die Funktion eines ECS massiv beeinflussen und mitbestimmen. In ECS werden Sensornetzwerke (beispielsweise Zigbee oder Bluetooth-Low-Energy) genutzt, um Daten von verteilten Sensoren zu erfassen und zu Edge-Gateways zu transportieren. Lokale Netzwerke wie WLAN oder LAN verbinden nahegelegene Geräte und Edge-Knoten für lokale Kommunikation. Mobilfunknetze (4G/5G) bieten flächendeckende, drahtlose Hochgeschwindigkeitsverbindungen, die insbesondere für mobile oder entfernte Edge-Geräte von Vorteil sind. Je nach Anwendung muss ein passender Netzwerktyp gewählt werden. Abbildung 6 visualisiert die Problemstellung der Auswahl eines geeigneten Netzwerktyps.

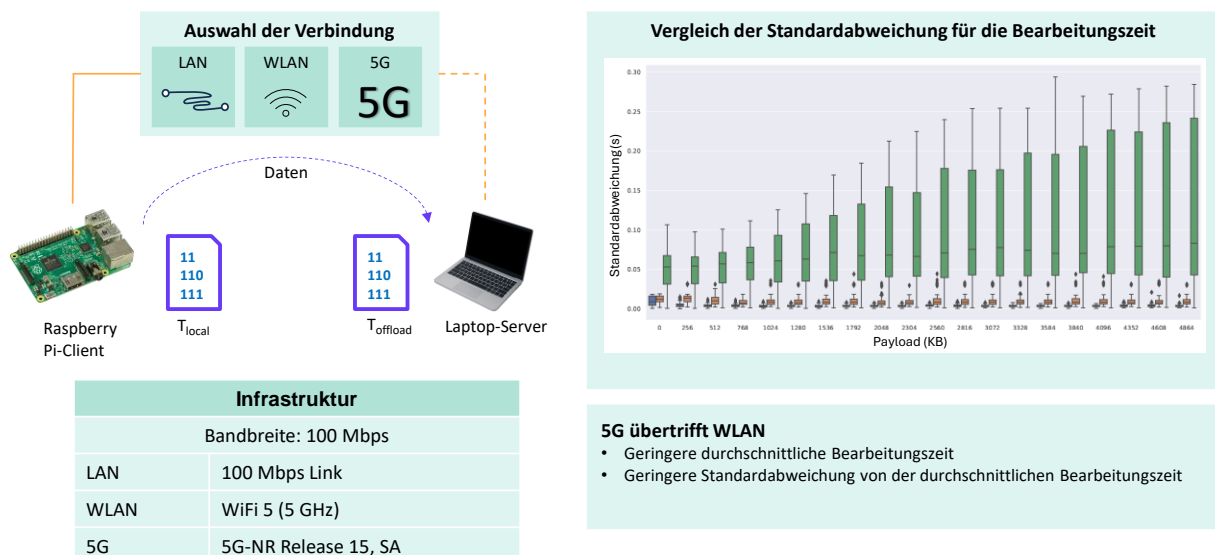


Abbildung 6: Vergleich der Ende-zu-Ende-Latenz für unterschiedliche Netzwerke und übertragene Datenmengen (GEMIMEG-II, 2024)

Funktionale Netzwerkparameter

Die Funktionsfähigkeit eines Netzwerks kann über diverse Parameter charakterisiert werden. Die maximale Übertragungsleistung eines Netzwerks wird über die nominale Bandbreite beschrieben und typischerweise in Gigabit pro Sekunde (Gb/s) angegeben. Die Latenz des Netzwerks beschreibt die Zeitdauer für die Übertragung eines Datenpakets bzw. Datensatzes von einem Netz-

werkteilnehmer zu einem anderen. Die Zeitdauer wird typischerweise in Milli- (ms) oder Mikrosekunden (μ s) angegeben. Die Latenz hängt stark von der Netzwerkauslastung ab, so dass die Angabe von z. B. minimaler, mittlerer und auch maximaler Latenz für einen Datensatz für einen Netzwerktest oder eine Applikation sinnvoll sein kann. Unterschiedliche Anwendungen haben z.T. sehr unterschiedliche Anforderungen an die zulässige Latenz. So können diese von Echtzeitanforderungen bis hin zu Minuten oder Stunden reichen. Applikationen in der Produktionstechnik – insbesondere Maschinenregelungen – zeichnen sich häufig durch Echtzeitanforderungen aus. Die Norm ISO/IEC 2382 definiert Echtzeitverarbeitung als Verarbeitung von Daten unter Einhaltung von Zeit-Anforderungen eines externen Prozesses (ISO, 2015). Der Kernaspekt von Echtzeit ist somit ein deterministisches Zeitverhalten beim Reagieren auf externe (oder auch interne) Ereignisse (Hüning, 2019). Typische Latenzen für interaktive Systeme mit Anwendenden oder in der Regelungstechnik reichen von knapp unterhalb 200 ms bis hin zu wenigen Millisekunden, je nach Kritikalität.

Im Projekt GEMIMEG-II wurde die Leistungsfähigkeit von unterschiedlichen Netzwerken für die Anwendung für Mess-, Steuer- und Regelungsaufgaben untersucht (Engel, 2023; s. Abbildung 6). Für die aktuellen verfügbaren Netzwerke für eine Fabrikautomation standen die Funktechniken 5G (SA-Release 15) für Mobilfunk und WiFi 6 für WLAN zur Verfügung. Die durchschnittliche Ende-zu-Ende-Latenz in WiFi 6 ist mit 6,5 ms deutlich kleiner als die Ende-zu-Ende-Latenz in 5G. Allerdings ist die maximale Latenz unter typischen Anwendungsbedingungen mit 117 ms in 5G deutlich besser als 330 ms in WiFi 6. Die Datenraten für den Download sind sehr ähnlich. Die Datenrate für den Upload ist in WiFi 6 deutlich größer und flexibler zu definieren als in 5G (GEMIMEG-II, 2024).

Zudem können unter anderem folgende weitere funktionale Netzwerkparameter erfasst und bewertet werden:

- **Zuverlässigkeit (Packet-Loss-Ratio):** Die Zuverlässigkeit ist die Paketzustellungsrate zwischen zwei Endpunkten bei einer bestimmten maximalen Übertragungsdauer (Herlich et al., 2017).
- **Durchsatz:** Bei der Datenübertragung ist der Netzdurchsatz die Menge an Daten, die in einem bestimmten Zeitraum tatsächlich und erfolgreich von einem Ort zum anderen (in der Regel zwischen zwei Netzwerkknoten) übertragen wird. Der Netzdurchsatz wird in der Regel in Bits pro Sekunde (bps), wie in Megabit pro Sekunde (Mbps) oder Gigabit pro Sekunde (Gbps) gemessen (Aslanpour et al., 2020). Im Gegensatz dazu gibt die nominale Bandbreite die Kapazität des Netzwerks an, also die maximale Datenmenge, die das Netzwerk übertragen kann.
- **Netzüberlastung (Congestion):** Netzüberlastung in Datennetzen ist die verminderte Dienstqualität, die auftritt, wenn ein Netzwerkknoten oder eine Verbindung mehr Daten transportieren soll als verarbeitet werden können. Eine Möglichkeit, die Netzüberlastung zu beurteilen, ist die Betrachtung von Paketverlusten.
- **Laufzeitvarianz (Jitter):** Die Laufzeitvarianz bezeichnet die Abweichung der Latenzzeit bei der Übertragung von Datenpaketen (IONOS Redaktion, 2022). Insbesondere in Industrieranwendungen mit Echtzeitanforderungen ist essenziell, dass die Laufzeitvarianz innerhalb vorgegebener Toleranzen liegt. Anwendungen wie Sensordatenfusion erfordern bei hoher Laufzeitvarianz Datenpuffer, was die Latenz deutlich erhöht.

Schätzung von Netzwerkparametern mittels KI

Die im vorangegangenen Abschnitt beschriebenen Netzwerkparameter zur Charakterisierung der Funktionalität und Performanz von Kommunikationsnetzwerken, sind zwar grundsätzlich messbar, jedoch ist eine Messung in vielen Fällen mit erheblichem Aufwand verbunden. Gegebenenfalls wird das Netzwerk durch eine Messung auch zusätzlich belastet, sodass Performanzeinbußen entstehen. Im Kontext von ECS ist für die Entscheidung zur Verteilung der Services entlang des ECC außerdem eine kontinuierliche Erhebung der Parameter entscheidend, um Dauer, Aufwand und Zuverlässigkeit der damit einhergehenden Datenübertragung bestimmen zu können. Aus diesem Grund ist es sinnvoll, die Performanzmetriken mittels KI-Methoden zu schätzen und damit die Messungen im laufenden Betrieb zu ersetzen. Ein solches KI-Modell zur Schätzung von Leistungsparametern muss zunächst anhand von Daten aus Messungen trainiert werden, die im Vorfeld aufgezeichnet werden müssen. Danach kann die Inferenz des Modells jedoch ohne weitere Messungen verwendet werden. Dies erspart folglich den Aufwand für kontinuierliche Messungen während des Betriebs.

Da es sich bei Netzwerken im ECC um dynamische Topologien² handelt, haben sich sogenannte Graph-Neural-Networks (GNNs) (Scarselli et al., 2009) zur Modellierung des Problems und zur Schätzung der Leistungsparameter als besonders geeignet erwiesen. GNNs wurden spezifisch für die Verarbeitung und Modellierung von Daten, die in einer Graphstruktur vorliegen, entwickelt. Sie sind invariant gegenüber Permutationen von Knoten und Kanten der Graphen und bieten damit die Möglichkeit, auch auf Graphen mit anderer Struktur und Größe zu generalisieren, d.h., akkurate Vorhersagen für bislang ungesehene Graphen treffen zu können. Entsprechend können sich diese auch Topologieänderungen anpassen. Da sich Netzwerktopologien in Form von Graphen darstellen lassen und sich außerdem in vielen Fällen dynamisch verändern, eignen sich GNNs besonders für die Anwendung im Kontext der Datenübertragung in Kommunikationsnetzen (Suárez-Varela et al., 2023).

Zur Schätzung der Netzwerkparameter werden dem GNN die vorliegende Netzwerktopologie im ECC sowie weitere Parameter zum Netzwerkverkehr (z. B. Paketgrößen, Abstände zwischen den Paketen) und die gemessenen Zielwerte der jeweiligen Performanz-Metriken zum Training zur Verfügung gestellt, um anschließend mithilfe des Inferenz-Modells die entsprechenden Metriken schätzen zu können (s. Abbildung 7).

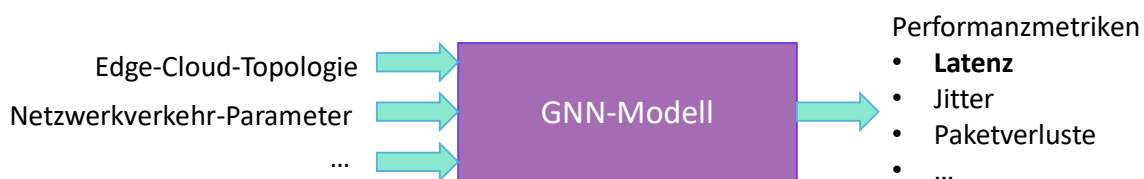


Abbildung 7: Netzwerktopologie und Performanzmetriken

Zur optimalen Ausführung von Aufgaben in ECS kann es insbesondere für die Gesamtdauer eines Prozesses sinnvoll sein, Datenverarbeitungsschritte auf eine andere, typischerweise nachfolgende und in Bezug auf CPU, GPU, RAM etc. leistungsfähigere Rechenressource zu verlagern. Zur Ermittlung der besten Konfiguration müssen die zeitlichen Aufwände für die eigentliche Datenverarbeitung auf der jeweiligen Ressource und auch die Aufwände für die Übertragung der

² Eine dynamische Topologie impliziert, dass sich die Netzwerkstruktur dynamisch ändert, etwa, indem einzelne Geräte ein Netzwerk von intelligenten Geräten flexibel betreten oder verlassen können.

Eingangsdaten und der Ergebnisdaten im Netzwerk betrachtet werden. Dabei hängen die Übertragungszeiten stark von der aktuellen Netzwerkauslastung zum Übertragungszeitpunkt ab. Diese sind zum Zeitpunkt der Verteilungsentscheidung a priori nicht bekannt und können nur geschätzt werden. Im Nachhinein (a posteriori) kann dann die Güte der Schätzung bestimmt werden und so die Qualität bzw. Richtigkeit der Verteilungsentscheidung bewertet werden. Im Projekt GEMIMEG-II wurde die Verteilung von Datenverarbeitungsaufgaben unter Berücksichtigung der Netzwerkauslastung untersucht. Zur Optimierung der Gesamt-Bearbeitungszeit wurden Methoden des Reinforcement-Learning (RL) und des Deep-Reinforcement-Learning (DRL) angewendet und verglichen. Abbildung 8 zeigt die Ergebnisse für die Gewinne an reduzierter Durchlaufzeit (Offloading-Gain). Je nach Netzwerklast und Komplexität der Datenverarbeitungsaufgabe konnten Gewinne um den Faktor 3 bis 9 erzielt werden. Die Ergebnisse von RL und DRL liegen nah aneinander. DRL ist dennoch die geeignetere Methode in dynamischen Umgebungen, um das gemessene Systemverhalten in zuverlässige Vorhersagen zu verwandeln, da es mit weniger Eingangsdaten auskommt und zudem im Lernprozess fehlende Eingangsdaten durch Interpolieren in gewissen Grenzen auch approximieren kann.

Client: Raspberry Pi 3B, Server: FUJITSU-CELSIUS H770, CPU Ratio: 10, Dynamic Case, BW 50 Mbps, 5G

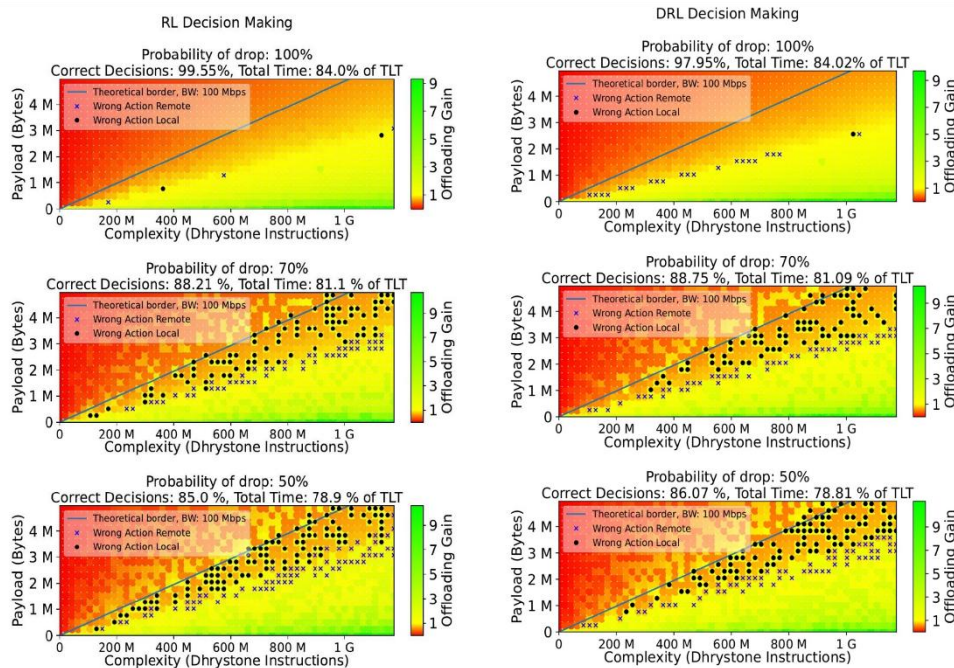


Abbildung 8: Optimierung der Bearbeitungszeit durch Verteilung von Datenverarbeitungsaufgaben in Netzwerken in Abhängigkeit von der Netzwerkbelastung. Optimierung mit Reinforcement-Learning (RL) und Deep-Reinforcement-Learning (DRL) (GEMIMEG-II, 2024).

Werden zusätzlich zur Auslagerung einzelner Datenverarbeitungsaufgaben voneinander abhängige Teilaufgaben betrachtet, die im ECC platziert werden können, sind weitere Überlegungen für eine Verteilungsentscheidung notwendig. Die wichtigste Entscheidung ist dabei die Optimierungsfunktion, nach der die Verteilungen im Netzwerk bewertet werden. Diese kann sich je nach Anwendungsfall stark unterscheiden und Echtzeit-, Latenz- und Ressourceneffizianz anforderungen abbilden. So können neben der schnelleren Übertragung und Bearbeitung von Aufgaben auch Kosten durch die gezielte Abschaltung von Ressourcen eingespart werden (Brockhaus, 2024).

Um vor der Ausführung von Aufgaben und der Datenübertragung im Netzwerk bereits Verteilungsentscheidungen zu treffen, ist eine Modellierung der Rechnerinfrastruktur und des Netzwerks nötig. Zur Bestimmung der Aufwände und Zeiten für die eigentliche Datenverarbeitung wird im Projekt ESCOM das bestehende Netzwerk über gerichtete azyklische Graphen abgebildet, wobei die Knoten Edge-Computing-Ressourcen und Switches und die Kanten die Netzwerkverbindungen darstellen (Brockhaus et al., 2024). Zudem können zu verteilende Softwareaufgaben ebenfalls als Graphen dargestellt werden, bei denen einzelne Teilaufgaben als Knoten und die benötigten Datenübertragungen zwischen diesen Teilaufgaben als Kanten modelliert werden. In Verbindung mit einer für den Anwendungsfall angepassten Optimierungsfunktion können daraufhin zusätzlich zu einer Schätzung der Netzwerkauslastung die Verteilungen der Teilaufgaben im Kontinuum bestimmt werden (Task-Scheduling). Hierzu haben sich Metaheuristiken zur Lösungsfindung des NP-vollständigen Optimierungsproblems als geeignet herausgestellt. Es zeigt sich also, dass die Platzierung von Software-Tasks einen großen Einfluss auf die Datenübertragungen im Netzwerk hat und neben der generellen Betrachtung von Netzwerkparametern eine entscheidende Rolle bei der Optimierung im ECC spielt.

3.3 Bewertung von Energieverbrauch und CO₂-Emissionen der Gesamtsysteme

Anwendungsbereich	Modellierung von Verbrauchs- und Emissionsdaten in Produktion und Logistik
Bewertungskriterien	Energieeffizienz, Netzwerkkennzahlen, Ressourcennutzung und -auslastung, Skalierbarkeit, Verfügbarkeit
Kennzahlen und Metriken	Energieverbrauch (Produktion und Logistik), produktbezogener CO ₂ e-Fußabdruck, Netzwerkauslastung (Datenübertragung), Latenz, Datenverfügbarkeit (Hochfrequente Daten), Speicherplatzbedarf, Schnittstellenauslastung
<ul style="list-style-type: none"> ▪ Die Analyse der Umweltwirkung von industriellen Produkten und Prozessen spielt eine immer wichtiger werdende Rolle in der Gesamtsystembewertung. ▪ Der Einsatz von Edge-Computing zur Berechnung von CO₂-Äquivalentemissionen (CO₂e) in der Produktion und Logistik verbessert die Genauigkeit der Emissionsberechnung durch die dezentrale Verarbeitung hochaufgelöster Sensordaten. ▪ Durch die Aggregation und Vorverarbeitung der Daten vor Ort wird die Netzwerkbelastung reduziert, während gleichzeitig eine effizientere Analyse und Optimierung industrieller Prozesse gefördert wird. Weitere Eigenschaften wie Kosten oder IT-Security sollten zukünftig mit in die Überlegungen einfließen. 	

In diesem Abschnitt werden Methoden, Instrumente und Kennzahlen aufgezeigt, um die Eigenschaften von Edge-Computing im Kontext von ECS und insbesondere in Bezug auf den Energieverbrauch und die CO₂-Emissionen in verschiedenen Anwendungsfällen zu bewerten. Der Fokus liegt darauf, wie die lokale Datenverarbeitung durch Edge-Computing in betrieblichen Prozessen, beispielsweise in den Bereichen Produktion und Logistik, den Energieverbrauch senken und die Nachhaltigkeit verbessern kann.

Methode zur Berechnung der CO₂e-Emissionen

Das Projekt EDNA beschäftigt sich mit der nachhaltig automatisierten Fertigungswirtschaft eines Produzenten von LKW-Auflegern (Trailern). Ein zentraler Aspekt ist die Betrachtung des Energieverbrauchs sowie die Berechnung der CO₂-Äquivalentemissionen (CO₂e) zur Bewertung der Umweltwirkung von Prozessen und Technologien. Die Kennzahl CO₂e umfasst nicht nur CO₂,

sondern auch andere Treibhausgase wie Methan (CH₄) und Distickstoffmonoxid (N₂O) und gewichtet diese Treibhausgase nach ihrem globalen Erwärmungspotenzial relativ zu CO₂.

Die Berechnung dieser Emissionen basiert maßgeblich auf der ISO-Norm-Familie 14000 und den daraus hervorgegangenen Berechnungsmethoden. Innerhalb des EDNA-Projektes werden die CO₂e-Emissionen für zwei verschiedene Use-Cases berechnet: die während der Produktion von LKW-Trailern entstehenden Emissionen („Cradle-to-Gate“) und die im Betrieb („Cradle-to-Grave“), also bei Logistikoperationen, entstehenden Emissionen durch LKW-Trailer.

Die Emissionsberechnung für die Produktion basiert auf der Norm DIN EN ISO 14067, während die Emissionsberechnung für die Logistik auf Norm DIN EN ISO 14083 beruht. Für die automatisierte Berechnung der CO₂e-Emissionen werden entsprechend der Normen für den Anwendungsfall der Produktion die Eingangsfaktoren Energieverbräuche (z. B. elektrische Energie, Druckluft, Erdgas), Verbräuche sonstiger Betriebsstoffe (z. B. Schutzgas, Schweißdraht) und ihre verbrauchsspezifischen Herstellungsenergiebedarfe (z. B. kg CO₂e / kg Material) benötigt. Für den Anwendungsfall der Logistik bedarf es Informationen über die Distanz, Masse und die Emissionsintensität. Die entsprechenden Datenquellen sind zum einen direkt an den Produktionsmaschinen aufgenommene Metriken und zum anderen im Trailer ermittelte Sensorwerte. Diese Daten werden mittels des firmeninternen TCP/IP-Netzwerks (Modbus TCP & OPC UA) d. h. des Produktionsnetzwerkes auf der OT-Ebene oder einer Telemetrieinheit innerhalb des Trailers übertragen und an einem zentralen Ort ausgewertet. Je höher die Sensorwerte aufgelöst sind, desto genauer kann die Berechnung der Verbräuche und damit folglich der CO₂e-Emissionen angesehen werden.

Bei der Übertragung von Daten aus den operativen Umgebungen existieren jedoch einige Limitationen. Hierzu gehören die Auslastung des Produktionsnetzwerkes, da dieses durch die begrenzte Datenrate die hochfrequenten Daten nicht mit der notwendigen Geschwindigkeit übertragen kann, die durch die Übertragung von Telemetriedaten entstehenden Mobilfunkkosten sowie die Mobilfunkverfügbarkeit und -geschwindigkeit. Für dieses Problem bietet Edge-Computing eine mögliche technische Lösung: Daten werden nicht sofort an ein zentrales Rechenzentrum gesendet, sondern vor Ort, also dezentral („on-the-Edge“), auf Mini-Computern verarbeitet. Diese Methode reduziert den Datenverkehr und damit die benötigte Bandbreite und entlastet sowohl die Netzwerkinfrastruktur als auch die zentralen Systeme, d. h. die Cloud. Das entsprechende Vorgehen kann in vier Phasen aufgeteilt werden:

- **Lokale Datenaggregation:** Sensordaten werden in Echtzeit erfasst und auf lokalen Edge-Geräten aggregiert.
- **Vorverarbeitung:** Die Datenströme werden lokal vorverarbeitet, gefiltert und aggregiert, um nur relevante Informationen weiterzuleiten.
- **Berechnung:** Die notwendigen Berechnungen zur Bestimmung der CO₂e-Emissionen werden ebenfalls lokal durchgeführt.
- **Übertragung der Ergebnisse:** Nur die berechneten Ergebnisse werden mit reduzierter Frequenz gesendet.

Ein Vergleich des Einsatzes dieser Methodik im Anwendungsfall Logistik mit der herkömmlichen Cloud-basierten Ermittlung der CO₂e-Emissionen ist in Abbildung 9 dargestellt und wird folgend erläutert. Alternativ zur im Logistik-Kontext relevanten Kenngröße „Distanz“ ließe sich eine vergleichbare Darstellung auch für den Stromverbrauch oder den Druckluftverbrauch im Anwendungsfall Produktion erstellen, um den beschriebenen Sachverhalt zu veranschaulichen.

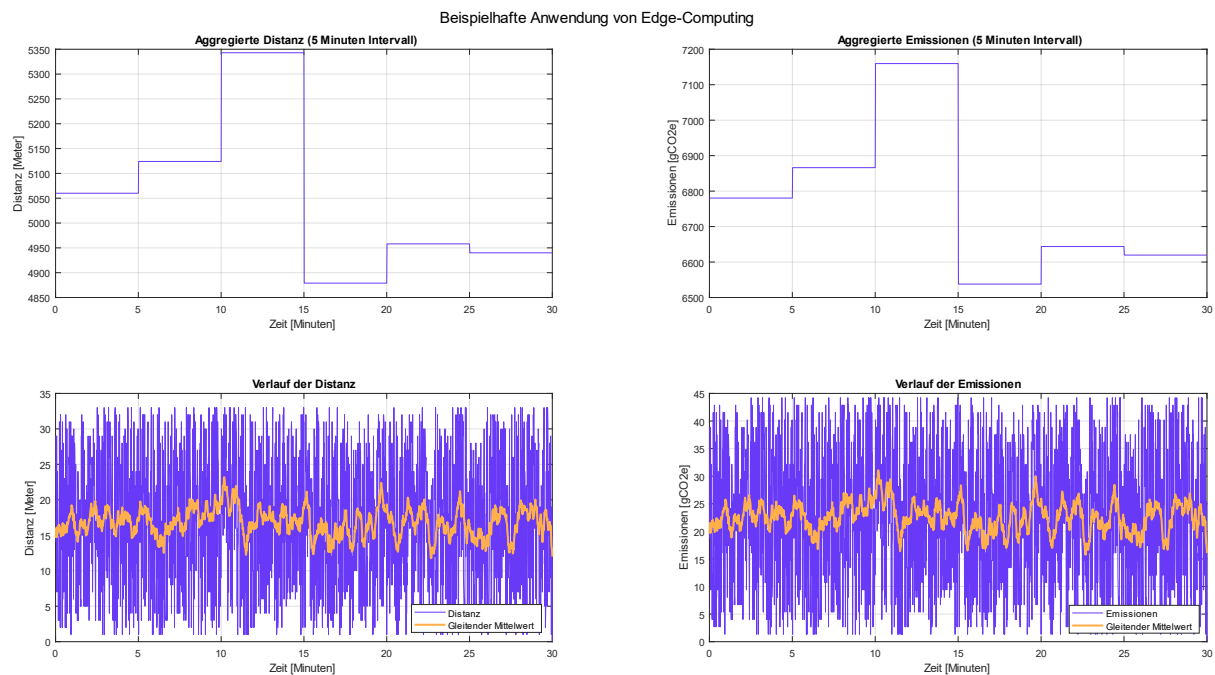


Abbildung 9: Granularere Sensordaten eröffnen tiefere Prozesseinblicke

Das übliche Übertragungsintervall einer Telemetrieinheit liegt im Minuten- bis Stundenbereich. In Abbildung 9 sind entsprechende Beispieldaten dargestellt. Die obere Zeile simuliert eine Telemetrieinheit und zeigt die jeweils für fünfminütige Intervalle aggregierten Werte der zurückgelegten Distanz und der aus der Masse, Distanz und Emissionsintensität berechneten CO_{2e}-Emissionen. Analog dazu sind in der unteren Zeile die im Sekundentakt aufgezeichneten Werte der zurückgelegten Distanzen und CO_{2e}-Emissionen samt ihrem gleitenden Mittelwert dargestellt. Durch die reine Aggregation der Distanzen gehen detaillierte Informationen über den Bewegungsverlauf zwischen den Datenpunkten verloren. Dadurch können beispielsweise starke Beschleunigungs- oder Bremsvorgänge in den Emissionsberechnungen nicht berücksichtigt werden, was zu ungenauen Ergebnissen führt und den Streckenverlauf zu einer Art „Blackbox“ macht. Im ungünstigsten Fall würde dies bedeuten, dass die komplette Distanz in den ersten Minuten bei sehr hoher Geschwindigkeit zurückgelegt worden wäre, wobei in den restlichen Minuten des Intervalls keine Distanz mehr zurückgelegt wird. Dies wäre folglich aus den aggregierten Daten nicht ersichtlich. Gleiches gilt in der Produktion: Wenn nur aggregierte Daten betrachtet werden, könnten kurzfristige Spitzenverbräuche oder Schwankungen unentdeckt bleiben.

Eine Verknüpfung der Prozess- und Auftragsdaten der Anlagensteuerung mit den Sensordaten durch das Edge-Gerät ermöglicht, produktions- und auftragsbezogene Energieverbräuche zu ermitteln. Dadurch können sowohl Verbrauchsspitzen als auch Anomalien wie z. B. Druckluftleckagen identifiziert und analysiert werden. Indem hochaufgelöste Sensordaten direkt „on-the-Edge“ verarbeitet werden, können detaillierte Informationen in die Berechnungen einfließen, die bei einer rein aggregierten Datenübertragung verloren gehen würden. Dadurch wird nicht nur die Netzwerkbelastung reduziert, da lediglich vorverarbeitete Ergebnisse übertragen werden müssen, sondern auch eine präzisere Analyse und Optimierung des Energieverbrauchs in Produktionsprozessen, analog zum vorherigen Anwendungsfall der Logistik, ermöglicht.

Zusammenfassung

Edge-Computing bietet eine vielversprechende Lösung, um die Herausforderungen hochfrequenter Datenerfassung und -übertragung bei der präzisen CO₂e-Emissionsberechnung gezielt und anwendungsspezifisch zu adressieren – wie anhand der Anwendungsfälle aus Produktion und Logistik verdeutlicht wird. Durch die lokale, dezentrale Verarbeitung hochaufgelöster Daten direkt an der Quelle lassen sich komplexe Zusammenhänge erfassen – etwa zwischen Distanz und Masse im Anwendungsfall Logistik sowie zwischen Energieverbrauch, Maschinenzuständen und Produktionsausfällen in der Fertigung. Dieser Ansatz reduziert die Auslastung der Kommunikationsinfrastruktur erheblich, während dennoch hochaufgelöste und präzise Daten für Berechnungen und weitere Analysen, beispielsweise für Optimierungsmaßnahmen in der Produktion, genutzt werden können. Dies trägt nicht nur zur Genauigkeit und Effizienz der CO₂e-Berechnungen bei, sondern fördert auch die Nachhaltigkeit und Leistungsfähigkeit industrieller Prozesse.

3.4 Bewertung von Methoden des Maschinellen Lernens – Föderiertes Lernen

Anwendungsbereich	Modelltraining, Daten- und Informationsaustausch
Bewertungskriterien	Datensouveränität, Dateneffizienz, Modellqualität
Kennzahlen und Metriken	Zu erreichende Genauigkeit für neu in Betrieb genommene Clients
<ul style="list-style-type: none"> ▪ Neu in Betrieb genommene Produkte, Maschinen, Prozesse oder Anlagen können direkt mit einem funktionsfähigen Modell ausgestattet werden, welches Informationen aller anderen Clients enthält. Ein Austausch von Rohdaten zwischen den Clients ist dafür nicht notwendig. ▪ Dadurch wird nicht nur die Vertraulichkeit sensibler Daten gewahrt, sondern auch die Einhaltung von Datenschutzvorschriften erleichtert. Zudem reduziert sich der Bandbreitenbedarf, da nur Modellaktualisierungen und nicht große Datenmengen übertragen werden. ▪ Durch die kontinuierliche Verbesserung der Modelle auf Basis dezentraler Datenquellen können außerdem robustere und generalisierbare KI-Modelle entstehen, die besser auf verschiedene Betriebsbedingungen reagieren. Schließlich ermöglicht Federated-Learning eine potenziell effizientere Nutzung von Rechenressourcen, da die Trainingsprozesse auf den jeweiligen Endgeräten (Edge-Devices) der Maschinen stattfinden und dynamisch entschieden werden kann, zu welchem Zeitpunkt unter welchen Bedingungen die einzelnen Edge-Devices beim Training mit aktiv werden. 	

Föderiertes Lernen (engl. Federated-Learning) (FL) (McMahan et al., 2016) ist ein Verfahren zum verteilten Lernen von maschinellen Lernmodellen. Die Besonderheit ist, dass beim FL keine Rohdaten geteilt werden. Dies ist besonders relevant für Anwendungen, bei denen die Eingangsdaten kritisch sind und daher nicht ohne weiteres geteilt werden können oder dürfen. Beispiele hierfür sind Daten von Smartphones oder Daten von Maschinen im industriellen Kontext, die einen Personenbezug oder Geschäftsgeheimnisse aufweisen. Bei diesen Anwendungen besteht dennoch ein Bedarf für Applikationen, bei welchen Modelle des maschinellen Lernens im Hintergrund laufen, um bspw. das nächste Wort bei der Benutzung der Smartphone-Tastatur vorherzusagen oder einen sich ankündigenden Maschinenstillstand anhand der Maschinendaten zu prognostizieren.

Klassischerweise benötigen diese Applikationen (Roh-) Daten zum Training. Weiterhin werden die trainierten Modelle immer besser, je mehr Daten von unterschiedlichen Anwendungen (sog. Clients) zur Verfügung stehen. Wird ein Modell mit Daten von mehreren Clients trainiert und dann wieder an die einzelnen Clients zurückgegeben, so können alle Clients voneinander lernen. Eine

Maschine in der Produktion könnte damit befähigt werden, Fehlerzustände frühzeitig zu erkennen, obwohl diese noch nie bei dieser Maschine aufgetreten sind. In den Daten der anderen Maschinen waren diese aber vorhanden und stehen nun allen Maschinen zur Verfügung. Mit diesem Prozess würde das vorhandene Wissen einzelner Akteure klassischerweise über mehrere Clients geteilt. Daten werden gesammelt, ein sogenanntes globales Modell wird mit allen Daten trainiert und das globale Modell an alle Clients verteilt.

FL bietet die Möglichkeit, globale Modelle zu trainieren, ohne die (Roh-) Daten zwischen verschiedenen Clients zu teilen. Die Funktionsweise von FL ist schematisch in Abbildung 10 dargestellt.

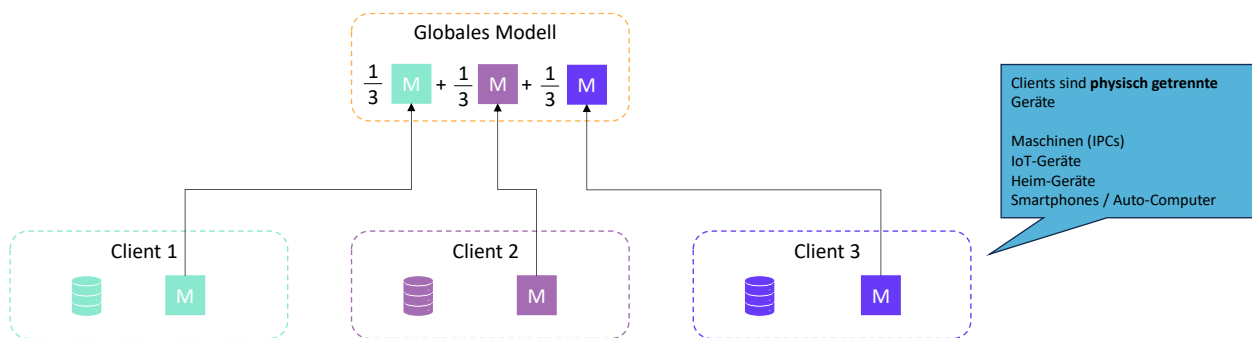


Abbildung 10: Federated-Learning für industrielle Anwendungen

Der Prozess ist dabei nahezu identisch zum beschriebenen klassischen Ansatz, mit der Ausnahme, dass keine Rohdaten, sondern lokale Modelle geteilt werden. Wir gehen davon aus, dass jeder Client ein sogenanntes lokales Modell besitzt (Schritt 1), dass die gewünschte Aufgabe mit den vorliegenden lokalen Daten lösen soll. Dieses Modell soll nun verbessert werden. Dazu wird das Modell an einen zentralen Server gesandt (Schritt 2). Die Modelle aller Clients werden zu einem gemeinsamen globalen Modell kombiniert. Dieses globale Modell enthält nun die Modelle und damit das Wissen aller Clientmodelle und wird im Schritt 3 an alle Clients verteilt. Dieses Vorgehen wird rundenweise durchgeführt, um eine kontinuierliche Verbesserung des globalen Modells zu ermöglichen.

Somit kann Wissen der Clients verteilt werden, ohne dabei Rohdaten zu versenden. Dies hat unmittelbare Vorteile hinsichtlich der Datensouveränität und meist auch hinsichtlich der Netzwerkauslastung. Es ergeben sich weitere wichtige, nicht unmittelbar ersichtliche Vorteile durch FL, speziell in industriellen Anwendungen: Hier werden Maschinen oder Kundenprodukte zu einem bestimmten Zeitpunkt initial in Betrieb genommen. Zu diesem Zeitpunkt liegen noch keine lokalen Daten vor und das Training eines Client-spezifischen Modells ist nicht möglich. Durch FL steht nun aber ein globales Modell zur Verfügung, das bereits das Wissen der sich in Betrieb befindlichen Maschinen enthält und mit der neuen Maschine geteilt werden kann. Es ist damit möglich, Clients mit funktionsfähigen Modellen zu initialisieren. Sobald ein Client an einem FL-System teilnimmt, teilt dieser kontinuierlich das lokal angepasste globale Modell aus der vorherigen Runde mit allen weiteren Clients und stellt somit ebenfalls das lokale Wissen des Clients zur Verfügung.

Eine zentrale Herausforderung sind heterogene Daten. Dies ist besonders für industrielle Applikationen relevant, da hier größtenteils heterogene Anwendungen vorliegen. Beispiele sind unterschiedliche Maschinen, unterschiedliche Kunden und unterschiedliche Sensoren. Ein aktueller Forschungsansatz ist das sogenannte Federated-Clustering. Dabei werden nicht alle Clients in einem globalen Modell gebündelt, sondern es werden ähnliche Clients in sogenannten Cluster-

Modellen gebildet. Ähnliche Clients könnten bspw. Maschinen oder Motoren eines bestimmten Typs sein, welche charakteristische Eigenschaften teilen.

In einem ECC zeichnet sich FL dadurch aus, dass die Inferenz (Ausführung des Modells) immer auf den Edge-Devices der Clients und die Modellbildung immer auf einem (Cloud-) Server stattfindet. Zudem kann mittels FL die Performanz von Client-Modellen dateneffizient gesteigert werden, ohne dass dazu Client-spezifische Daten vorliegen müssen. Die Dateneffizienz lässt sich durch die Übertragungsgenauigkeit messen. Die Übertragungsgenauigkeit beschreibt die Genauigkeit, die auf einem Client erreicht werden kann, wenn dieser mit dem globalen Modell anstelle eines lokalen Modells initiiert wurde. Die Übertragungsgenauigkeit gibt also an, wie gut ein globales Modell die Informationen verschiedener lokaler Modelle, die auf unterschiedlichen Geräten trainiert wurden, integriert. Zusätzlich kann die Steigerung der Genauigkeit bei einem weiteren Training mit einer geringen Menge lokaler Daten gemessen werden. Eine Steigerung der Genauigkeit gibt dabei Hinweise auf eine gute Generalisierbarkeit des Modells.

Im Folgenden werden verschiedene Beispiele zur Anwendung von FL in der Praxis präsentiert.

Praxisbeispiel: Überwachung des Betriebszustands von Elektromotoren

Ein in Erforschung befindliches Anwendungsbeispiel bei der Firma Bosch ist die Überwachung des Betriebszustandes von Elektromotoren, wie sie in vielfältigen Maschinen verbaut sind. Den Zustand der Motoren rechtzeitig und möglichst genau bestimmen zu können, ist für Kunden relevant. Gleichzeitig dürfen keine Rohdaten der Motoren von den Kunden geteilt werden. FL bietet eine Möglichkeit, neue Motoren mit einer funktionsfähigen Software zu bespielen, kontinuierlich Wissen über den Betriebszustand zwischen den Motoren zu teilen und die Clientmodelle kontinuierlich aktuell zu halten.

Praxisbeispiel: Föderiertes Lernen für die visuelle Qualitätsinspektion in der Fertigung

Neben den industriellen Anwendungsfällen, die sich auf Sensor- und Maschinendaten fokussieren, können Techniken des FL auch für die visuelle Qualitätsinspektion in der Fertigung eingesetzt werden. In einem praktischen Beispiel wurden ähnliche USB-Sticks hergestellt und durch FL ein gemeinsames Modell zur Qualitätsprüfung erstellt. Dabei mussten die einzelnen Trainingsdaten der USB-Sticks nicht zwischen den verschiedenen Produktionsstandorten oder Geräten ausgetauscht werden. Dies simuliert den Fall, dass mehrere Hersteller oder Standorte ein gemeinsames Modell trainieren, ohne einander die Rohdaten zur Verfügung stellen zu müssen. Das trainierte Modell wies eine hohe Genauigkeit auf und zeigt, dass die Methoden des FL auch für visuelle Aufgaben hilfreich sein können (Hegiste et al., 2022).

Praxisbeispiel: Föderiertes Lernen für die Verarbeitung von Smart-Living-Daten

Ein Beispiel für die praktische Umsetzung des FL-Ansatzes bietet das Projekt SECAI. Im Projekt SECAI wird FL für die Verarbeitung von Daten aus der Smart-Living-Domäne mit dem Ansatz des Edge-Cloud-Computing kombiniert (s. Abbildung 11). In der Praxis sieht die Umsetzung die Einteilung von Wohneinheiten innerhalb eines Gebäudes vor, in denen auf Raumebene gesammelte Daten zur individuellen Modellierung genutzt werden. Zu den verwendeten Daten gehören unter anderem die Raumtemperatur und die Außentemperatur. Auf der Edge-Ebene werden diese Daten lokal verarbeitet, um möglichst energieeffiziente Heizpläne für alle Wohnungen eines Gebäudes zu berechnen, die anschließend an die Mieterinnen und Mieter zurückgespielt werden. Die

Modelle auf Gebäudeebene werden dann gesammelt und übergreifend generalisiert. Die generalisierten Gewichte (Weights) werden zur Optimierung zurück auf die Ebene der einzelnen Wohnungen übertragen.

FL kommt im Projekt aus zwei Perspektiven zum Einsatz: Zum einen wird geprüft, inwiefern der Ansatz einen Mehrwert beim Einsatz von energieeffizienten KI-Komponenten bieten kann, und zum anderen dient er der Wahrung der Privatsphäre. Da die in den Wohnungen gesammelten Daten nie die jeweilige Wohnung verlassen, bleibt die Privatsphäre der Mieterinnen und Mieter geschützt. Beide Aspekte werden im Projekt umgesetzt und wissenschaftlich untersucht.

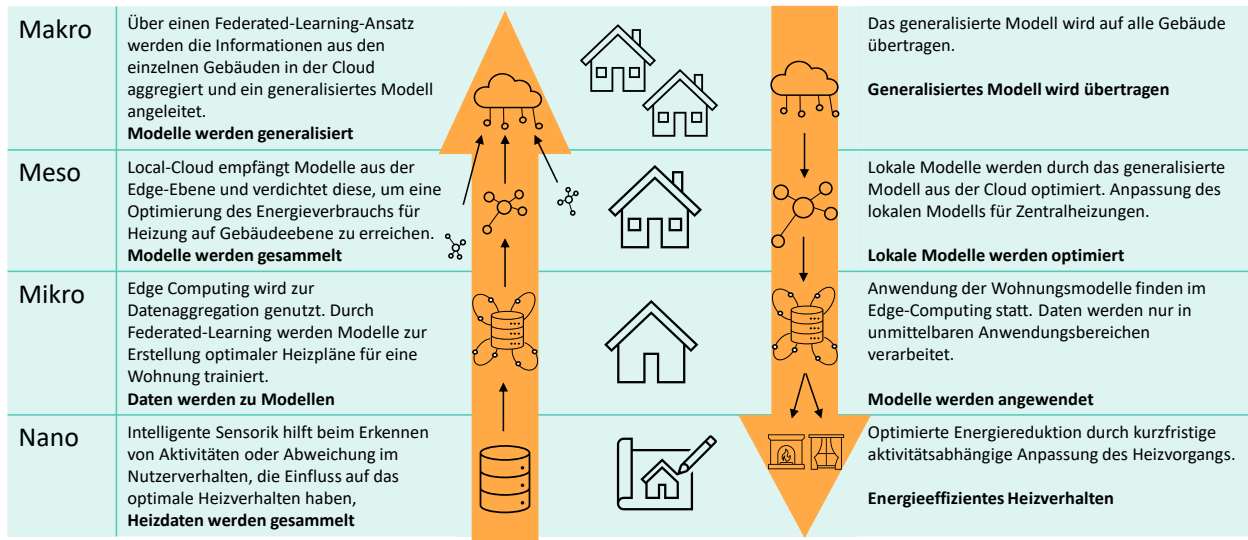


Abbildung 11: Federated Learning im Bereich Smart Living

Praxisbeispiel: Föderiertes Lernen zur Anomaliedetektion in Kommunikationsnetzwerken auf Basis von Netzwerkparametern

Zusätzlich zu den Anwendungen im Kontext des Produktionsumfelds und im Bereich Smart-Living kann FL im ECC auch zur Gewährleistung einer sicheren und zuverlässigen Datenübertragung im Netzwerkbereich eingesetzt werden. Dabei werden einzelne Modelle zur Anomaliedetektion in verschiedenen Teilen des Netzwerkes trainiert und mittels FL zusammengeführt. Die Trainingsdaten aus den jeweiligen Netzwerken müssen dabei nicht ausgetauscht werden und bleiben privat. Trotzdem profitiert das Gesamtmodell von der größeren Menge an Trainingsdaten aus den unterschiedlichen Netzwerken. Dies ist insbesondere für die Anomalieerkennung von Vorteil, da bestimmte Anomalien nur sehr selten auftreten und daher oftmals zu wenig Trainingsdaten für eine modellbasierte Detektion vorliegen. Abbildung 12 zeigt die Architektur des Netzwerkes, mit dem die föderierte Anomalieerkennung für die Datenübertragung im ECC getestet wurde.

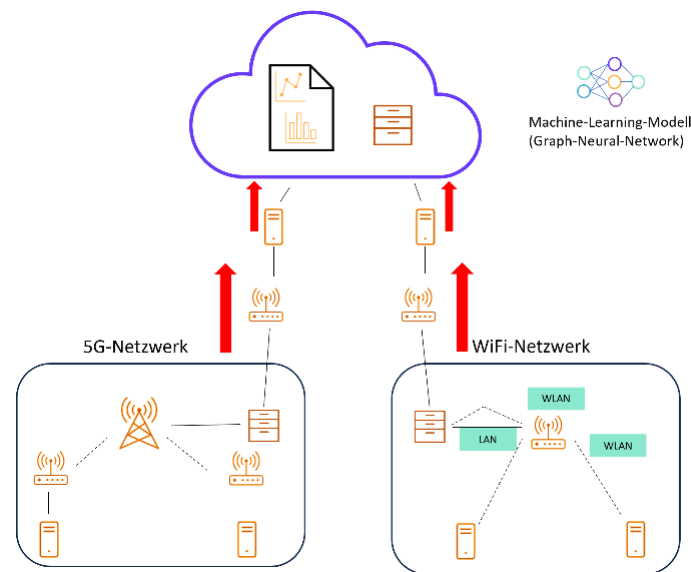


Abbildung 12: Federated-Learning zur Anomaliedetektion in Kommunikationsnetzwerken

3.5 Bewertung von KI-Planung

Anwendungsbereich	Autonom agierende Systeme
Bewertungskriterien	Durchlaufzeit der Planung, Reduktion des Energieverbrauchs
Kennzahlen und Metriken	Zeit, Energieverbrauch, Plankosten (z. B. zur Approximation der Planungsqualität)
<ul style="list-style-type: none"> Die Bewertung des Einsatzes von KI-Planung in einem ECC ist notwendig, um zu bestimmen, wann die mit zusätzlichem Aufwand verbundene Aufteilung des Planungsproblems Vorteile mit sich bringt (z. B. reduzierter Zeitaufwand oder reduzierte Plankosten). Je nach Komplexität des Planungsproblems kann dieser zusätzliche Mehraufwand nicht durch die Vorteile der verteilten Berechnung ausgeglichen werden (Trade-Off). Klassische KI-Planungsprobleme sind in der Regel nicht dafür geeignet, um verteilt gelöst zu werden und müssen somit für die verteilte Planung angepasst werden. Die Trade-Offs beinhalten somit die anfallenden Kosten für die Anpassung sowie den Aufwand der Zusammenführung der einzelnen Planungsergebnisse gegenüber dem effektiven Nutzen bei der Berechnungskomplexitätsreduktion. Im Rahmen der Forschung in Richtung hierarchischer Planung könnten mit voranschreitenden Technologien Ansätze entwickelt werden, die eine automatische Anpassung des Planungsproblems für die verteilte Ausführung ermöglichen. In diesem Fall müssten die Trade-Offs der Nutzung von verteilter Planung neu bewertet werden. 	

KI-Planung zielt auf die Lösung eines Zustandstransformationsproblems ab, um eine Abfolge von Schritten zu finden, mit der es möglich ist, von einem Ausgangszustand zu einem gewünschten Zielzustand zu gelangen (Green, 1981; Haslum, 2006). Der Ausgangszustand beschreibt dabei ein diskretes Abbild einer Planungsumgebung durch eine Menge von Fakten. In einem Produktionsszenario beschreibt der Ausgangszustand z. B. die Produktionsmittel und die aktuell verfügbaren Ressourcen. Der Zielzustand stellt ebenfalls ein diskretes Abbild der Planungsumgebung dar, fokussiert aber die gewünschten Veränderungen, wie z. B. neu hergestellte Produkte. Zusätzlich zum Anfangs- und Zielzustand wird eine Menge von sogenannten Operatoren benötigt.

Ein Operator führt innerhalb des diskreten Abbildes der Planungsumgebung zu einer Zustandsveränderung. Innerhalb eines Produktionsszenarios wäre ein Operator z. B. der Vorgang, ein Loch in ein Werkstück zu bohren.

Im Rahmen eines ECC können die darin angebotenen KI-Planungsdienste sowohl für die Planung interner Prozesse des Kontinuums als auch externer Prozesse verwendet werden. Einer der Vorteile eines solchen Kontinuums ist die Möglichkeit, Rechenressourcen flexibel zu verteilen und somit Berechnungen je nach Bedarf auf mehrere Knoten des Netzwerkes aufzuteilen. Bei dieser Ressourcenallokation handelt es sich um einen internen Prozess, der durch KI-Planungsmethoden unterstützt werden kann. Bei einem solchen Planungsproblem ist der Zielzustand das Erreichen einer optimalen Verteilung von Berechnungen und Daten auf die einzelnen Instanzen im Kontinuum. Bei größeren Netzwerken, komplexen Einschränkungen (z. B. datenschutzrechtlicher Art) oder möglichst echtzeitnahen Systemen, muss diese zwingend automatisiert werden. Eine Anwendung von KI-Planungsmethoden kann diese Automatisierung ermöglichen.

Abseits der internen Prozesse des Kontinuums bieten sich KI-Planungsmethoden an, um externe Prozesse zu adressieren. So können beispielsweise Steuerungsprozesse in Fertigungsumgebungen von am Kontinuum teilnehmenden Industriepartnern betrachtet werden. Diese Technik wird im Bereich des Geschäftsprozessmanagements (engl. Business-Process-Management, BPM) bereits erfolgreich eingesetzt, um den Automatisierungsgrad und die Flexibilität zu erhöhen (Marrella, 2018; Schultheis, Jilg et al., 2024). Bei diesem Anwendungsfall ist der Zielzustand ein gewünschtes Produkt mit bestimmten Eigenschaften. Das zugrundeliegende Planungsproblem besteht in diesem Szenario darin, eine Folge von ausführbaren Fertigungsaktionen zur Herstellung des gewünschten Produkts zu finden.

Einer der Nachteile bei der KI-basierten Lösung von Planungsproblemen ist die hohe Rechenkomplexität, die insbesondere bei komplexeren Planungsproblemen negativ skaliert und dem Einsatz dieser Technologie im Wege steht. Durch den Einsatz in einem ECC können diese Probleme adressiert und zum Teil ausgeglichen werden. So kann beispielsweise die Berechnung auf mehrere Knoten im Kontinuum aufgeteilt oder von Anfang an auf einen leistungsstärkeren Knoten verlagert werden. Dabei ist es auch möglich, das Planungsproblem an sich und nicht nur die Berechnung aufzuteilen. Dies kann erreicht werden, indem Ansätze aus dem Forschungsfeld des hierarchischen Planens angewendet werden. Dabei wird ein Prozess erst auf abstrakter Ebene geplant und in mehreren Stufen konkretisiert, um die Komplexität zu verringern. Dieser Ansatz kann von einem ECC profitieren, da die verschiedenen granularen Planungsprobleme auf verschiedenen Ebenen des ECC gelöst werden können.

Eine weitere Möglichkeit, die Nachteile und Probleme von KI-Planungsansätzen zu mindern, ist die Kombination von KI-Planung mit anderen KI-Methoden, wie z. B. mit dem Fallbasierten Schließen (siehe Abschnitt 3.6). Dabei kann auch hier das Lösen des Planungsproblems verteilt werden, um in diesem speziellen Fall die Stärken von verschiedenen KI-Methoden zu kombinieren. Auch dieser Lösungsansatz kann von den Eigenschaften eines ECC profitieren, indem die verschiedenen KI-Methoden auf verschiedene Knoten des Kontinuums verteilt werden.

Die Potenziale, die ein ECC für die Optimierung von KI-Planung bietet, eignen sich auch dafür, ein solches Kontinuum und somit auch Edge-Computing-Technologien in Bezug auf die oben genannten Vorteile zu bewerten. Dabei kann die singuläre Ausführung eines KI-Planungssystems mit der verteilten Ausführung auf mehreren Edge-Knoten sowie die Kombination mit anderen KI-Methoden verglichen werden. Hierbei kann z. B. zwischen der Qualität des generierten Plans sowie der dabei entstandenen Kosten unterschieden werden. Die Vorteile von Edge-Computing-

Technologien und der Ausführung in einem ECC können sich dabei durch eine höhere Qualität oder niedrigere Kosten bei der Planerstellung auszeichnen. Die Qualität eines durch KI-Planung generierten Plans kann z. B. durch die Länge des Plans (Anzahl der separaten Schritte) oder die Kosten der geplanten Arbeitsschritte bewertet werden. Bei der Bewertung der Kosten der Planerstellung können z. B. die Durchlaufzeit der Planungsphase oder die entstandenen Stromkosten herangezogen werden.

Praxisbeispiel: Produktionsplanung

Ein in einer Produktionsstätte installiertes Edge-Gerät eines ECC erkennt, dass ein Produktionsprozess nicht wie geplant ausgeführt werden kann (beispielsweise auf Basis der Ergebnisse von Analyseprozessen). Daher stößt es eine adaptive Umplanung des Produktionsprozesses an. Da es aber nicht über genug Rechenkapazität für die Anwendung eines KI-Planungssystems verfügt, beauftragt es ein weiteres Edge-Gerät im gleichen Kontinuum damit, einen unvollständigen groben Plan mittels CBR zu generieren. Anschließend vervollständigt und verfeinert das Edge-Gerät den Plan und gibt den auszuführenden Prozess zurück an die Workflow-Engine.

3.6 Bewertung von Methoden des Maschinellen Lernens – Fallbasiertes Schließen

Anwendungsbereich	Analytics / KI, Autonom agierende Systeme
Bewertungskriterien	Ergebnisqualität, Geschwindigkeit, Energieeffizienz
Kennzahlen und Metriken	Qualität der Lösung, Berechnungsdauer, Energieverbrauch
<ul style="list-style-type: none"> ▪ Die Bewertung des Einsatzes von Case-Based-Reasoning (CBR) ist notwendig, um zu bestimmen, wann die Verwendung verteilter Ansätze im Vergleich zu einem traditionellen Ansatz von Vorteil ist (Trade-Off). ▪ Trade-Offs beinhalten die anfallenden Kosten zur Beschaffung der Edge-Geräte sowie den Aufwand der Integration und der Entwicklung der verteilten Ansätze gegenüber dem effektiven Nutzen bei der Berechnungskomplexitätsreduktion. ▪ Im Rahmen der Forschung im Bereich des verteilten CBR könnten mit voranschreitenden Technologien Ansätze entwickelt werden, die eine automatische Verteilung der Fallbasis und anschließende verteilte Ähnlichkeitsberechnung erlauben. In diesem Fall müssten die Trade-Offs der verteilten Nutzung von CBR neu bewertet werden. 	

Fallbasiertes Schließen (engl. Case-Based-Reasoning, CBR) ist eine KI-Methode zur erfahrungsbasierten Problemlösung (Bergmann et al., 2009; Bergmann et al., 2020). Probleme und ihre entsprechenden Lösungen werden als Fälle gespeichert, die die Grundlage für die Bearbeitung neuer Probleme bilden. Auf dieses Erfahrungswissen wird bei der Adressierung neuer Probleme zurückgegriffen. Um geeignete Fälle zu identifizieren, wird das Kriterium der Ähnlichkeit verwendet, mit dem die Eignung eines Falls für eine neue Problemstellung approximiert wird. Die Lösung des Falls kann entweder direkt wiederverwendet werden oder für ähnliche, aber nicht identische Situationen adaptiert werden. Neue Problemlösungen werden auf ihre Eignung evaluiert und bei erfolgreicher Anwendung ebenfalls als Fall gespeichert, sodass es sich um einen kontinuierlich lernenden Ansatz handelt. Im Gegensatz zu datengetriebenen Methoden des klassischen Maschinellen Lernens handelt es sich bei CBR um einen wissensgetriebenen Ansatz. Durch die direkte Wiederverwendung von Erfahrungswissen werden keine großen Datensätze zum Training benötigt, sodass auf Basis weniger Daten geeignete Lösungen gefunden werden können.

Im Rahmen eines industriellen ECC kann CBR sowohl für die Analyse als auch für die Steuerung von Produktionsprozessen verwendet werden. Analyseprozesse befassen sich mit der Untersuchung und Auswertung von Daten, die im Kontext der Produktion oder des ECC auftreten, sowie der Identifikation der Ressourcennutzung. Der Einsatz von CBR bietet sich insbesondere im Kontext der Fehleranalyse für IoT-Daten an, wie beispielsweise im Kontext der Vorausschauenden Wartung (engl. Predictive-Maintenance) oder bei der Erkennung von Datenqualitätsproblemen (engl. Data-Quality-Issues) (Malburg et al., 2023). Aufgrund des Umfangs und der Komplexität von IoT-Daten ist die Klassifikation von Fehlerfällen aufwändig, sodass üblicherweise nur wenige Trainingsdaten zur Verfügung stehen. Durch die beschriebenen Vorteile von CBR ist diese Technologie für diesen Anwendungsfall prädestiniert. Da es sich bei der Analyse von IoT-Zeitreihendaten um komplexe Berechnungen handelt (Schultheis, Malburg et al., 2024), bietet sich die Ausführung in verteilten Umgebungen wie dem ECC an. Neben dem Performanzgewinn können Fehleranalysen direkt an den möglicherweise betroffenen Maschinen stattfinden, sodass die Problemanalyse in weniger komplexe Teilprobleme zerlegt und direkt am jeweiligen Gerät nach Fehlern sowie Methoden zur Fehlerbehebung gesucht wird. Als mögliche Kennzahl lässt sich beispielsweise der Energieverbrauch heranziehen, der für diese verteilte Analyse mit einer klassischen Analyse, die ausschließlich zentral in der Cloud läuft, verglichen werden kann. Zudem kann auch der reduzierte Datenverkehr durch die direkte Analyse an der Edge sowie die benötigte, durch die verteilte Berechnung reduzierte Zeit bis zur Lösungsfindung betrachtet werden.

Neben dem Einsatz für die Analyse kann CBR auch zur Prozesssteuerung verwendet werden. Steuerungsprozesse befassen sich mit der dynamischen Ressourcenzuweisung und der flexiblen, nachhaltigen sowie robusten Planung und Ausführung von Produktionsprozessen. Als klassische Methode dafür wird die KI-Planung (siehe Abschnitt 3.7) verwendet. Da es sich dabei um ein komplexes Problem handelt, wird in der Fallbasierten Planung (engl. Case-Based-Planning, CBP) ein Planungsproblem unter Zuhilfenahme von Fallwissen adressiert (Bergmann et al., 2020). Dabei können Fälle beispielsweise als Kontrollwissen oder zur Bestimmung ähnlicher Lösungen eingesetzt werden. Dies ermöglicht eine weitere Reduktion der Berechnungskomplexität für den bereits in Abschnitt 3.7 beschriebenen Einsatz in der intelligenten Steuerung. Der Effizienzgewinn durch die Kombination von KI-Planung und CBR kann ebenfalls über den Energieverbrauch als Kennzahl sowie zusätzlich über die benötigte Berechnungsdauer für die Erstellung eines fertigen Plans bestimmt werden.

Praxisbeispiel: Case-Based Reasoning für Anwendungen im Bereich Predictive Maintenance

Ein Beispiel aus der Praxis für Kontrollprozesse ist in Abschnitt 3.7 vorgestellt. Dieser Anwendungsfall kann ebenfalls mittels des kombinierten Ansatzes aus CBR und KI-Planung adressiert werden. Für die Analyseprozesse kann Predictive-Maintenance (Schultheis, 2024) als Anwendungsfall herangezogen werden. Dabei handelt es sich um ein Konzept im Bereich der Industrie 4.0, bei dem die proaktive Wartung von Maschinen auf Basis von Sensordaten untersucht wird. Das Ziel ist es, unnötige Eingriffe zu vermeiden und rechtzeitig vor Fehlern zu intervenieren, um Produktionszeiten zu optimieren. Auf Basis der Analyse von IoT-Zeitreihendaten sollen anomale Muster identifiziert werden, die Rückschlüsse auf die Art des Problems sowie den erwarteten Ausfallzeitpunkt zulassen. Somit soll eine Intervention rechtzeitig vor dem Ausfall stattfinden, aber auch nicht zu früh, sodass unnötige Wartungsarbeiten vermieden werden. Ist an jede der Maschinen ein Edge-Gerät angegliedert, so überprüft dieses die Maschinendaten auf Anomalien. Mittels CBR werden vorhandene Muster für Fehlerfälle mit den aktuellen Daten abgeglichen und beginnende Fehlerverläufe klassifiziert. Für die einzelnen Maschinen erfolgt dies direkt auf den Edge-

Geräten, während in der Cloud nur mögliche, von mehreren Maschinen abhängige Fehlerfälle ebenfalls auf Basis von Fallwissen analysiert werden. Somit wird nicht nur die Analyse verteilt, sondern zudem das Fallwissen nur an die Instanzen gegeben, für die dieses relevant ist. Der Datenverkehr wird somit auf die Klassifikationsergebnisse sowie die aufgrund von Abhängigkeiten zu analysierenden Daten beschränkt. Darüber hinaus ermöglicht die durch die Verteilung erreichte Komplexitätsreduktion eine möglichst echtzeitnahe Analyse.

3.7 Ansätze zur Sicherstellung von Privatsphäre, Datensouveränität und Vertrauen

Anwendungsbereich	Unternehmensübergreifender Daten- und Informationsaustausch
Bewertungskriterien	Datensouveränität, Vertrauenswürdigkeit
Kennzahlen und Metriken	Qualitative Betrachtung – Nutzung von Technologien und Standards
<ul style="list-style-type: none"> ▪ Die Bewertung von Datensouveränität und Privatheit ist von zentraler Bedeutung, um die Akzeptanz der Nutzenden zu steigern und Dateninhabende zu befähigen, ihre Daten souverän teilen zu können. ▪ Eine quantitative Betrachtung von Datensouveränität, Privatheit und Vertrauen ist aktuell noch nicht etabliert und kann daher zum jetzigen Zeitpunkt nur qualitativ bzw. binär/dichotom erfolgen. ▪ Die Nutzung offener Standards und verteilter Architekturen trägt dazu bei, langfristig die Kontrolle über die Daten beim Erzeuger zu halten, über Verifizierbarkeit das Vertrauen zu erhöhen oder obsolet zu machen und auch zukünftig Regularien zu erfüllen. 	

Im Zuge der Bewertung von ECS spielen neben Performanz- und Nachhaltigkeitskriterien auch weitere Aspekte zunehmend eine Rolle. Besonders wichtig sind dabei Privatsphäre, Datensouveränität und Vertrauen. Vor diesem Hintergrund werden diese Konzepte im Folgenden vorgestellt. Sie helfen zum einen in der Designphase von ECS und bilden zum anderen auch die Grundlage bei der Evaluation dieser Systeme. Anschließend folgt die detaillierte Betrachtung und Beschreibung dieser Konzepte innerhalb der Projekte DEER, ESCOM und EASY.

Im November 2023 verabschiedete der Rat der Europäischen Union den Data Act (EU Data Act) mit dem Ziel, einen fairen Zugang zu und Vorschriften für die faire Nutzung von Daten einzuführen. Der EU Data Act ist am 11. Januar 2024 in Kraft getreten und wird ab dem 12. September 2025, nach einer Übergangsfrist von 20 Monaten, direkt anwendbares Recht (Generaldirektion Kommunikation, 2024). Der EU Data Act ist dabei ein integraler Bestandteil der europäischen Datenstrategie und trägt dazu bei, die Datensouveränität europäischer Unternehmen weiter zu stärken. Neben dem EU Data Act sind weitere juristische Aspekte wie die DSGVO bei der Bewertung von ECS relevant, die in Kapitel 3.8. ausführlich beschrieben werden. Diese Verordnungen bilden den gesetzlichen Rahmen für die Mehrfachnutzung von Daten in der EU.

Im Zuge der Bewertung von ECS werden die Konzepte von Privatsphäre und Datensouveränität häufig als Synonyme – zum Schutz sensibler Daten – verwendet. Ein differenziertes Verständnis der Konzepte ist jedoch für die Bewertung von ECS höchst relevant: Privatsphäre zielt auf den Schutz der eigenen, sensiblen Daten ab. Datensouveränität nimmt eine ganzheitlichere Perspektive ein und beschreibt verschiedene Aspekte der individuellen oder gesellschaftlichen Befähigung zum selbstbestimmten Umgang mit u. a. Big-Data (SMART DATA Forum). Datensouveränität ist die Fähigkeit, Regeln zur Datennutzung zu formulieren, Datenaustausch zu beeinflussen und dabei eigenständig entscheiden zu können, mit wem Daten geteilt werden sollen. Während Privatsphäre also in erster Linie auf den Schutz personenbezogener Daten abzielt, wird unter

Datensouveränität darüber hinausgehend verstanden, dass Dateninhaberinnen und -inhaber bestimmen können, wo ihre Daten gespeichert werden, wer welche Daten erhält, wie diese verarbeitet werden dürfen sowie an welchen Zweck und an welche Auflagen die Nutzung gebunden sein soll (Schütte et al., 2018).

Neben Datensouveränität und Privatsphäre ist Vertrauen ein zentrales Konzept bei der Entwicklung von ECS, insbesondere dann, wenn Daten über Organisationsgrenzen hinweg geteilt werden. Ein organisationsübergreifendes Teilen von Daten erschwert es, die Kontrolle über diese Daten und damit auch die Datensouveränität zu erhalten. Vertrauen ist die mit positiven Erwartungen gepaarte Bereitschaft, sich auf die Datennutzung entsprechend der vorher vereinbarten Richtlinien zu verlassen (Doney & Cannon, 1997; Rousseau et al., 1998). Insbesondere in organisationsübergreifenden ECS ist es herausfordernd, ein hohes Vertrauensniveau zwischen den Systemteilnehmern zu gewährleisten, um Datenaustausche zu ermöglichen.

In diesem Zusammenhang erhalten verifizierbare Berechtigungsnachweise (engl. Verifiable-Credentials) sowie der damit kombinierbare Sicherheitsansatz des „Zero-Trust“ wachsende Aufmerksamkeit. Letzterer basiert auf dem Prinzip, dass generell keinen Nutzenden bzw. keinem Gerät vertraut wird – unabhängig davon, ob sie sich innerhalb oder außerhalb eines spezifischen Netzwerkes befinden. Stattdessen wird jeder Zugriff auf (Daten-) Ressourcen streng überprüft und zunächst authentifiziert. Der Zero-Trust-Ansatz setzt dabei auf eine kontinuierliche Überwachung und Validierung aller Interaktionen, um sicherzustellen, dass nur autorisierte Nutzende und Geräte Zugang zu sensiblen Daten und Systemen erhalten. Folgend werden Anwendungsbeispiele zur Herstellung von Privatsphäre, Datensouveränität und Vertrauen in ECS erläutert.

Praxisbeispiel: Self-Sovereign-Identities und Zero-Knowledge-Proofs zur prototypischen Integration von Kleinanlagen in den Redispatch-Prozess

Unser Energiesystem muss sich vor dem Hintergrund der notwendigen Einsparung von CO₂-Emissionen massiv transformieren: Angebotsseitig führt der Ausbau erneuerbarer Energien dazu, dass die Einspeisung unter dem Einfluss von nicht steuerbaren Wetterbedingungen immer volatiler wird. Darüber hinaus gehen immer mehr kleine und Kleinanlagen – wie Wärmepumpen, Batteriespeicher und Photovoltaikanlagen – ans Netz und erhöhen sowohl die Stromnachfrage als auch die Dezentralität des Stromangebots. Vor diesem Hintergrund stehen im Zuge der Energiewende zukünftig viele dezentrale Erzeugende einer wachsenden Menge von dezentralen Abnehmenden gegenüber. Für Übertragungs- und Verteilnetzbetreiber wird deshalb der Aufwand für das Management des Stromnetzes (bspw. bei auftretenden Engpässen im Netz) immer komplexer. Um entsprechende Engpässe aufzulösen bzw. um ihnen vorzubeugen, stehen Netzbetreibern unterschiedliche Instrumente zur Verfügung, unter denen das DEER-Projekt insbesondere den sogenannten Redispatch-Prozess betrachtet (Heeß et al., 2025).

Redispatch-Maßnahmen werden ergriffen, wenn beispielsweise im Norden Deutschlands (zu) viel Strom erzeugt und im Süden (zu) viel Strom nachgefragt wird, sodass die Übertragungsleitungen, beziehungsweise Systemkomponenten, an ihre Belastungsgrenzen stoßen. Vorbeugend (d. h. vor dem Engpass) können dann Erzeugende im Norden vom Netz genommen und dafür Erzeugende, die südlich (also hinter dem Engpass) liegen, dazugeschaltet werden. Zur Sicherstellung der Netzstabilität kann dies auch dazu führen, dass bspw. ein Windrad abgeschaltet und ein konventionelles Kraftwerk zugeschaltet wird, und damit einerseits die Ausfallkosten der nachhaltigen Energie und andererseits höhere Emissionen in Kauf zu nehmen.

Im Projekt DEER wird daher erprobt, wie dieser Situation durch die aktive Einbindung von vorhandenen Kleinanlagen auf Verbrauchsseite, wie Wärmepumpen, PV-Anlagen, Heimspeichern oder Elektroautos, begegnet werden kann. Anstatt erneuerbare Energie zukünftig abzuregeln, kann so das Flexibilitätspotenzial dieser Anlagen genutzt werden. Das bedeutet bspw. nachfrageeitig einerseits den Verbrauch der Anlagen bei Bedarf zu drosseln und andererseits, Ladevorgänge von Batterien so zu planen, um netzdienlich Strom zu speichern. Somit soll die Nachfrage nach Strom generell besser und flexibler gesteuert werden. Das Projekt zielt dabei zunächst darauf ab, die technische Machbarkeit einer solchen Option, die dann freiwillig von den Eigentümerinnen und Eigentümern dieser Anlagen gewählt werden kann, technisch zu konzeptionieren, prototypisch zu implementieren und in einem Feldtest zu evaluieren.

Dazu wird im Projekt DEER ein Multi-Agenten-System entwickelt, das sich durch automatisierte Datenverarbeitung und -übermittlung auf Anlagenebene auszeichnet. Hierzu werden auf Grundlage von Wettervorhersagen und historischen Nutzungsdaten sogenannte Baselines auf den Geräten selbst („on-the-Edge“) erzeugt. Diese Baselines sind erwartete Energienutzungsprognosen, auf deren Grundlage sich dann wiederum Flexibilitätspotenziale ableiten lassen. Diese Potenziale werden über cloudbasierte Aggregatoren gebündelt und an die Netzbetreiber weitergegeben. Im Falle eines prognostizierten Netzengpasses können die Netzbetreiber Flexibilitäten abrufen und so aktiv einer Überlastung des Gesamtsystems entgegenwirken.

Vor dem Hintergrund des oben beschriebenen Prozesses ist es notwendig, Daten über den Stromverbrauch der Kleinanlagen zu teilen. Viele dieser Daten bieten tiefe Einblicke in das Nutzungsverhalten der Kleinanlagen, oft durch private Haushalte. Entsprechend sind diese Daten hochsensibel und zum Schutz der Privatsphäre der jeweiligen Dateneigentümerschaft gut zu schützen. Netzbetreiber müssen jedoch in der Lage sein, die Korrektheit und den Ursprung der geteilten Daten transparent verifizieren zu können, um Plausibilitätstests der einzelnen Angebote durchführen zu können, um wiederum die Stabilität des Stromsystems zu gewährleisten.

Um diesen Herausforderungen zu begegnen, werden im Rahmen des Projekts DEER Kleinanlagen in einem System mit Wallets ausgestattet, die die Konzepte Datensouveränität, Privatsphäre und Vertrauen berücksichtigen. In diesen Wallets lassen sich Verifiable-Credentials speichern und unter Wahrung der Datensouveränität gegenüber Dritten präsentieren. Gemeinsam mit Verifiable-Presentations bilden Wallets und Verifiable-Credentials dabei die Hauptsystemkomponenten von Self-Sovereign-Identities, einem Paradigma des digitalen Identitätsmanagements. Self-Sovereign-Identities erlauben es, zertifizierte Eigenschaften einem verifizierenden Akteur (Verifier) zu präsentieren, sodass dieser diese Eigenschaften überprüfen kann (Babel, Willburger et al., 2025). Ein analoges Pendant dieses Systems ist der Identitätsnachweis, in dem Bürgerinnen und Bürger ihre Identität mit einem gültigen Ausweisdokument präsentieren können und ein Verifier die Gültigkeit des Ausweises überprüfen kann.

Im Multi-Agenten-System für den Redispatch werden Self-Sovereign-Identities eingesetzt, um Anlagen mit Daten auszustatten, mit denen nachgewiesen werden kann, dass eine spezifische Anlage für den Redispatch Prozess geeignet – d. h. präqualifiziert – ist. Für die koordinierte Abwicklung des Datenaustauschs werden die Daten durch Aggregatoren zusammengefasst und optimierte, aggregierte Flexibilitätsangebote erstellt. Zusätzlich werden diese Daten den Netzbetreibern in einer verifizierbaren Weise zur Verfügung gestellt. So entsteht in den Vertrauens- und Datenaustauschbeziehungen ein sogenannter Trust-Diamond (s. Abbildung 13).

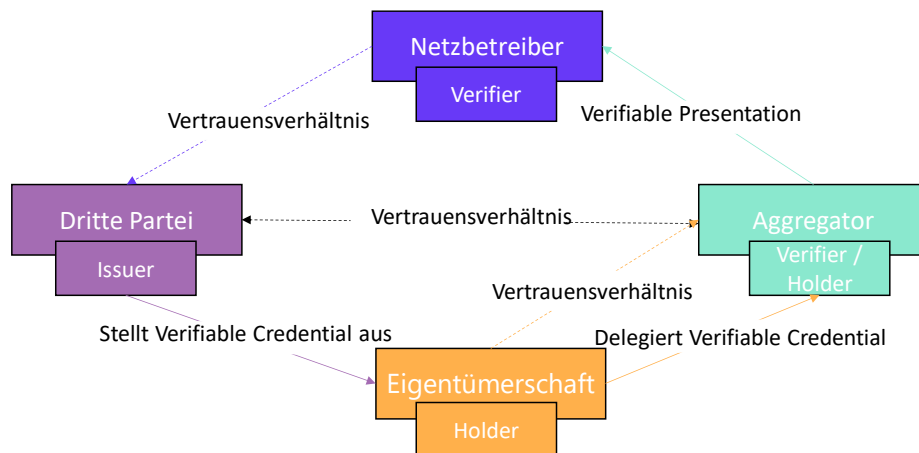


Abbildung 13: Verifizierbare Flexibilitätangebote und Datenschutz (Babel, Ehaus et al., 2025)

Im ersten Schritt stellt eine sogenannte Trusted-Third-Party, wie beispielsweise ein zertifizierter Anlagenhersteller, ein Verifiable-Credential an die Eigentümerschaft der Anlage aus. Das Credential wird im Wallet, das als Softwarekomponente auf der Anlage installiert wird, gespeichert. In einem initialen Onboarding-Prozess erhält der Aggregator die notwendige Autorisierung und Stammdaten, um Flexibilitätangebote der Eigentümerschaft zu bündeln und verifizierbar anzubieten. Für jeden Angebotsprozess wird automatisch ein Flexibilitätangebot auf Anlagenebene erstellt und an den Aggregator übertragen. Der wiederum aggregiert die Angebote vieler einzelner Anlagen und bietet den Netzbetreibern entsprechende (aggregierte) Flexibilitätspotenziale an, wobei überprüft wird, ob die einzelnen Angebote jeweils von zertifizierten Anlagen erstellt wurden, indem die Netzbetreiber auf Signaturen der Issuer referenzieren können. Durch diesen Ansatz werden die geteilten Daten zu den Anlagen verifizierbar, sodass das Risiko von Ausfällen durch bspw. nicht verfügbare Anlagen oder Betrug durch Überangebote minimiert werden kann.

Neben der Verifizierbarkeit ist der Schutz der privatsphärerelevanten Daten der Anlagen, die mit dem Aggregator bzw. den Netzbetreibern geteilt werden, ein essenzieller Bestandteil des Projekts. Diese Daten sind privatsphärerelevant, da sie detaillierte Rückschlüsse über die Stromnachfrage spezifischer Anlagen und damit über das Verhalten von Haushalten (wie z. B. Arbeitszeiten, Urlaube etc.) ermöglichen können. Der zielgerichtete Einsatz von Zero-Knowledge-Proofs hilft dabei, den notwendigen Austausch von Daten und Informationen zwischen Aggregator und Netzbetreibern zu senken und gleichzeitig die Verifizierbarkeit der Daten und Informationen zu erhalten. Zero-Knowledge-Proofs sind kryptografische Verfahren, mit denen Eigentümerinnen und Eigentümer von Informationen deren Korrektheit beweisen können, ohne die Information selbst preisgeben zu müssen: So kann beispielsweise nachgewiesen werden, dass eine Anlage innerhalb eines Netzgebietes liegt, ohne die genauen Koordinaten der Anlage selbst offenzulegen. Bei der Integration von Kleinstanlagen in den Redispatch-Prozess kann mit dieser Technik also nachgewiesen werden, dass eine Anlage für den Redispatch relevante Attribute erfüllt, ohne die Attribute selbst preisgeben zu müssen. Das erarbeitete Konzept im Projekt DEER ermöglicht es, aggregierte Flexibilitätangebote zu erstellen bei gleichzeitigem Erhalt der Verifizierbarkeit der Datenintegrität, die den Angeboten zugrunde liegt. Das Konzept legt die Grundlage, um bei Flexibilitätsabruf die Steuerung der Anlagen zu starten und flexible Kleinstanlagen in den Redispatch-Prozess zu integrieren.

Einsatz des Gaia-X-Trust-Frameworks für Komponenten-Service-Systeme in souveränen Edge-Cloud-Umgebungen

Unternehmen stehen vor einem doppelten Wandel, der sogenannten „Twin Transition“: Sie müssen ihren globalen ökologischen Fußabdruck verkleinern und gleichzeitig auf einem zunehmend digitalisierten Markt wettbewerbsfähig sein. Auch Komponentenhersteller von Werkzeugmaschinen, wie Spindel- und Kugelgewindetriebehersteller, haben Potenzial, mit digitalen Geschäftsmodellen und Nachhaltigkeit als Innovationstreiber ihre Wettbewerbsfähigkeit zu steigern. Bisher mangelt es jedoch am Kontakt zu Erstanwendenden und deren an der Edge aufgezeichneten (Prozess-)Daten, um Komponenten zu verbessern und datengetriebene Services für die Anwendenden anzubieten (Verein Deutscher Werkzeugmaschinenfabriken e.V., 2020). Hierbei stellt sich die Herausforderung, dass sowohl Endanwendende als auch Komponentenhersteller aufgrund der Angst vor Wissensabfluss nicht bereit sind, ihre Daten zu teilen. Zudem ergeben sich Herausforderungen bezüglich der Interoperabilität beim Austausch von Daten und der Provisionierung von Services im ECS externer Kunden. Somit steht der Mehrwert der Datenverarbeitung und Analyse dem damit verbundenen Aufwand und dem Verlust von Wettbewerbsvorteilen durch Preisgabe von privaten Daten, die inhärent Knowhow enthalten, entgegen.

Das Projekt ESCOM arbeitet durch einen anbieterunabhängigen Zugang zu digitalen Diensten von Komponentenherstellern, die in Edge-Cloud-Umgebungen eingesetzt werden an einer Lösung für diese Probleme. ESCOM stützt sich dabei auf erprobte Technologien wie die Federation-Services der Gaia-X-Initiative und das Verwaltungsschalenkonzept (VWS) der Plattform Industrie 4.0, um digitale Zwillinge von physischen Komponenten zu modellieren und den Austausch von Informationen über den gesamten Produktlebenszyklus hinweg ohne Verlust der Privatsphäre und Datensouveränität zu ermöglichen.

Durch die Integration von Komponentenservices in offenen, souveränen ECS über vertrauensvolle Gaia-X-Ökosysteme verschiebt sich die Marktperspektive von einer angebotsorientierten hin zu einer nachfrageorientierten Sichtweise. Komponentenhersteller können dadurch ein größeres Marktpotenzial erschließen, insbesondere durch Co-Innovationen mit Maschinenherstellern (OEMs). Konkret werden ebenfalls die zuvor beschriebenen Self-Sovereign-Identities in Zusammenhang mit Verifiable-Credentials und Presentations genutzt, um die Teilnehmenden im Ökosystem und die angebotenen Services überprüfbar zu beschreiben und die Einhaltung von (rechtlichen) Vorgaben zu bescheinigen. Diese dezentrale Identitätsstruktur wird vom World Wide Web Consortium (W3C) standardisiert (Credentials Community Group [W3C], 2021; Sporny et al., 2025) und eignet sich gut für die ebenfalls dezentralisierten und verteilten ECS.

Auch Gaia-X als Initiative zur Entwicklung eines technischen Frameworks für rechtssichere und souveräne Datenökosysteme und Gegenpol zu zentralisierten und monopolistischen Plattformmodellen orientiert sich an der oben beschriebenen dezentralen Infrastruktur und den entsprechenden Standards. Der Kernaspekt und das Werteversprechen von Datenökosystemen ist hierbei, dass offene Standards genutzt werden, die die Interoperabilität in verteilten ECS gewährleisten und Vertrauen schaffen. Hierdurch werden viele neue Geschäftsmodelle und Anwendungsfälle unter anderem im Bereich der produzierenden Industrie und des Maschinenbaus ermöglicht.

Praxisbeispiel: Souveräne Komponentenservices in der Anwendung

Als Anwendungsbeispiele werden im Projekt ESCOM die Bereitstellung mehrerer datenbasierten (KI-)Services für Maschinenkomponenten in der produzierenden Industrie betrachtet. Die Endan-

wendenden sammeln über maschineninterne und externe Sensoren Daten während ihrer Produktionsprozesse, die wertvolle Informationen für mögliche Zustandsanalysen der Komponenten, vorausschauende Instandhaltung und Vorschläge zur Prozessoptimierung liefern können. Ein konkretes Beispiel aus dem Projekt ist die Zustandsüberwachung und Lebensdauerabschätzung von Kugelgewindetrieben über die Berechnung einer Reibmomentenkennlinie mithilfe von Herstellerwissen. Zur Umsetzung dieses Anwendungsfalls trotz der zuvor beschriebenen Hindernisse werden Datenökosysteme und die Verwaltungsschale als Umsetzung des digitalen Zwillinges zur standardisierten Bereitstellung von Daten genutzt, um Services von Komponentenherstellern von Fräsmaschinen (Spindel, Kugelgewindetriebe) bei den Anwendenden einzusetzen.

Die an der Maschine gesammelten Daten können direkt auf nahegelegenen Edge-Devices verarbeitet werden und es kann eine Signatur/Hash erzeugt werden, mit der die Integrität der Daten im Laufe der Zeit verifiziert werden kann. Bei Fehlern oder bössartigen Änderungen an den Daten in folgenden Verarbeitungsschritten ändert sich die Signatur. Eine Veränderung der Daten kann entsprechend durch den Vergleich mit dem direkt an der Edge erstellten Hash erkannt werden. Da die Daten internes Knowhow enthalten, werden die oben beschriebenen Vertrauensmodelle aus dem Kontext von Gaia-X-Ökosystemen verwendet. Der Aufbau auf dem Gaia-X-Framework ermöglicht es, durch die interoperable Beschreibung von Services, Infrastrukturen und Vertragspartnern eine rechtliche Haftbarkeit und Vertrauensbasis zu schaffen. Zudem werden die teilbaren Services und Daten mit verifizierbaren Policies und Zugriffsregeln ausgestattet.

Um auf diesen Daten einen Zustandsüberwachungs-Service auszuführen, ohne dass die Daten die Domäne der Endanwendenden verlassen, ist der Ansatz des Compute-to-Data (Ocean Protocol Foundation Ltd., 2024) im Pontus-X-Ökosystem³ vielversprechend. Hierbei werden die Algorithmen zu den Daten gesendet und im Edge-Cloud-Netzwerk der Anwendenden in abgeschlossenen Containern ausgeführt. Dadurch kann die Datensouveränität sichergestellt werden, da nur die Ergebnisse der Berechnung das Unternehmen in Richtung des Komponentenherstellers verlassen. Zusätzlich setzt das Pontus-X-Ökosystem auf eine dezentrale Infrastruktur auf Basis der Distributed-Ledger-Technologie (DLT). Diese Infrastruktur nutzt ein verteiltes Netzwerk unabhängiger Knoten (Nodes), die gemeinsam den Systemzustand verwalten, Transaktionen validieren und die Integrität der Datenraumdienste sicherstellen, ohne dass eine zentrale Instanz erforderlich ist. Die Dezentralität verhindert, dass einzelne Akteure die Hoheit über das Gesamtsystem erlangen und schafft so eine Vertrauensbasis für alle beteiligten Unternehmen und Institutionen. Die beschriebenen Komponenten und Aspekte des Anwendungsfalls im ECS, die Privatsphäre, Datensouveränität und Vertrauen sicherstellen, sind in Abbildung 14 dargestellt.

³ <https://www.pontus-x.eu/>

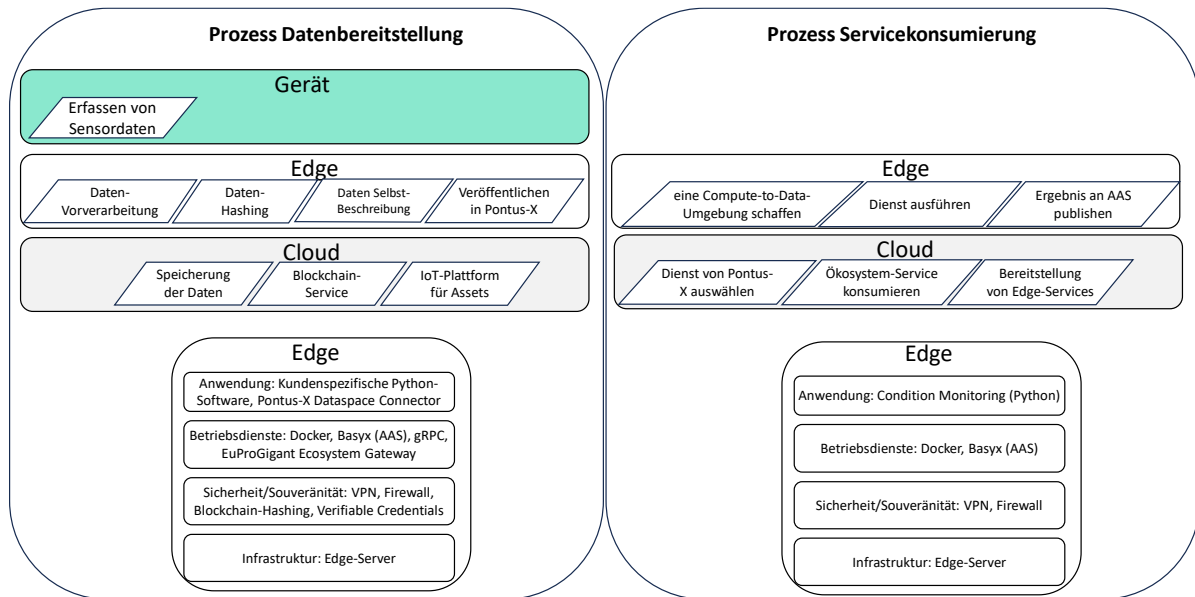


Abbildung 14: Datenbasierte Services und Datensouveränität

Privatheit und Sicherheit zur Schaffung von Zero-Trust-Architekturen in ECS

Privatheit und Sicherheit in ECS sind komplexe Themen, die eine gründliche Analyse und sorgfältig ausgewählte Schutzmechanismen erfordern. Im Folgenden wird exemplarisch auf Zero-Trust-Architekturen eingegangen. Sowohl das Bundesamt für Sicherheit in der Informationstechnik (BSI) (Bundesamt für Sicherheit in der Informationstechnik, 2023) als auch das National Institute of Standards and Technology (NIST) (Rose et al., 2020) heben die Bedeutung von Zero-Trust hervor, da es eine effektive Strategie zur Minimierung von Sicherheitsrisiken darstellt.

Die Implementierung von Zero-Trust-Architekturen in ECS erfordert eine sorgfältige Abwägung, insbesondere im industriellen Kontext. Einerseits bieten Zero-Trust-Prinzipien durch kontinuierliche Verifizierung und Segmentierung wirksamen Schutz für sensible Daten. Andererseits stehen Unternehmen vor Herausforderungen, da viele Legacy-Systeme nicht ohne weiteres in diese Architektur integriert werden können. Diese Systeme sind oft nicht für die strikten Sicherheitsanforderungen von Zero-Trust ausgelegt, was die Umsetzung erschwert.

Neben der Integration von Legacy-Systemen muss zudem der individuelle Prozess beachtet werden. Echtzeitanwendungen, z. B. in der industriellen Fertigung, können möglicherweise nicht alle Anforderungen einer modernen Cloud-Zero-Trust-Architektur einhalten, ohne den Prozess selbst durch den gestiegenen Kommunikationsaufwand zu gefährden. Hier sind besonders die Vorteile von ECS erneut hervorzuheben: Durch die Verlagerung von Authentifizierungsprozessen an die Edge kann die allgemeine Systemsicherheit durch dezentrale Zero-Trust-Prinzipien verbessert werden (Creutz & Dartmann, 2023). Maschinen, Steuerungen und Co. entscheiden so selbst über Zugriffe anhand globaler Regeln und reduzieren den Kommunikationsaufwand und die Latenzen.

Da es keine übergreifenden Standards für Zero-Trust-Architekturen gibt, müssen Unternehmen individuelle Lösungen entwickeln, die ihren spezifischen Sicherheitsanforderungen und Bedürfnissen gerecht werden. Ein entscheidender Schritt hierbei ist eine umfassende Anforderungsanalyse, die das spezifische Umfeld und die Bedürfnisse des Unternehmens berücksichtigt. Diese Analyse ermöglicht eine gezielte Bewertung des ECS und hilft, maßgeschneiderte Sicherheitsstrategien zu entwickeln, die sowohl die Privatheit als auch die Leistungsfähigkeit der Systeme sicherstellen.

Praxisbeispiel: AI on Device Lüftungsdemonstrator mit Zero-Trust-Architektur

Im Projekt EASY wird ein Lüftungsdemonstrator entwickelt (s. Abbildung 15). Der Lüftungsdemonstrator zeigt auf, wie die Luftqualität in industriellen Umgebungen wie Lagerhallen und Fertigungsräumen kontinuierlich überwacht und bei Bedarf automatisch angepasst werden kann. Technologisch basiert er auf einem SCD30-Sensor gekoppelt an ein vom Umwelt-Campus Birkenfeld erstelltes Entwicklungsboard, dessen Kern ein ESP8266 bildet⁴. Der SCD30 liest Umweltdaten über Temperatur, Feuchtigkeit und CO₂-Gehalt der Luft aus, auf deren Grundlage der Lüftungsstand bestimmt wird. Zum Sammeln von Trainingsdaten sind die beiden äußeren Klappen mit jeweils zwei magnetbasierten Reed-Schaltern ausgestattet. Durch diese kann zwischen den Zuständen „geschlossen“, „halb-offen“ und „geöffnet“ unterschieden werden. Die CO₂-Zuführung wird momentan noch über eine kohlenstoffhaltige Flüssigkeit erreicht. Exemplarisch genügt dieser Ansatz, zur verbesserten Steuerung soll dieses System jedoch zu einem späteren Zeitpunkt durch eine genauere CO₂-Zufuhr ersetzt werden. Im weiteren Projektverlauf sollen ebenfalls durch die einfache Replizierbarkeit des Aufbaus zusätzliche Raumkonfigurationen mit abweichender Größe erstellt werden, in denen ein verteilter Lernansatz genutzt wird, um die genauen Zeitreihenmuster der Lüftungsszenarien besser erkennen zu können.

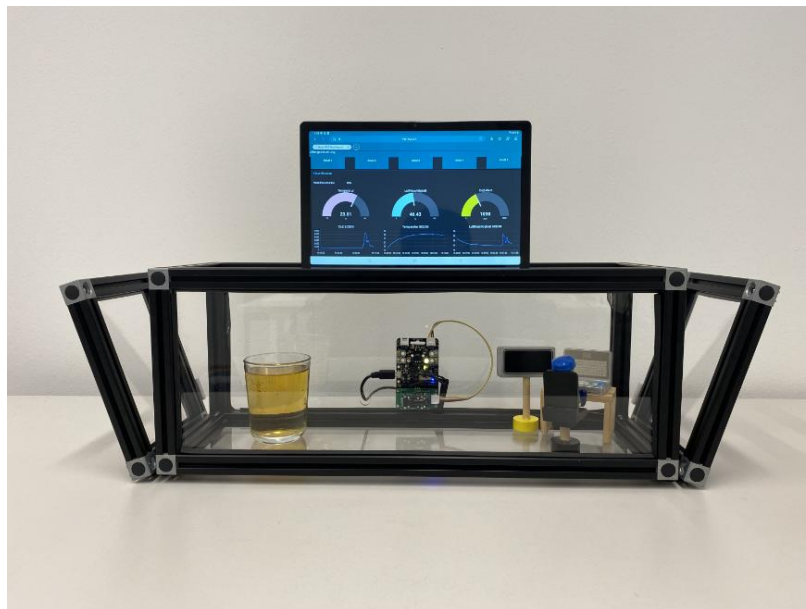


Abbildung 15: Lüftungssteuerung und verteiltes Lernen

Das Nutzerinterface, dargestellt in Abbildung 16, zeigt eine Auswahl der verschiedenen Räume (zum jetzigen Zeitpunkt ist nur ein Raum vorhanden). Über die im Interface abgebildeten Gaugen und Zeitreihengraphen werden die Sensorwerte dargestellt und eine Einschätzung über deren Höhe geliefert. Des Weiteren werden Informationen über den aktuellen und den ermittelten Stand der Lüftung sowie Informationen über die Größe des Raumes und den Energieverbrauch von einzelnen Komponenten geliefert. Zusätzlich wird den Nutzenden ermöglicht, Übertragungseigenschaften wie die Messfrequenz und die Größe der übermittelten Datenpakete zu bestimmen.

⁴ SCD30: hochpräzises CO₂-Sensormodul; ESP8266: 32-Bit-Mikrocontroller

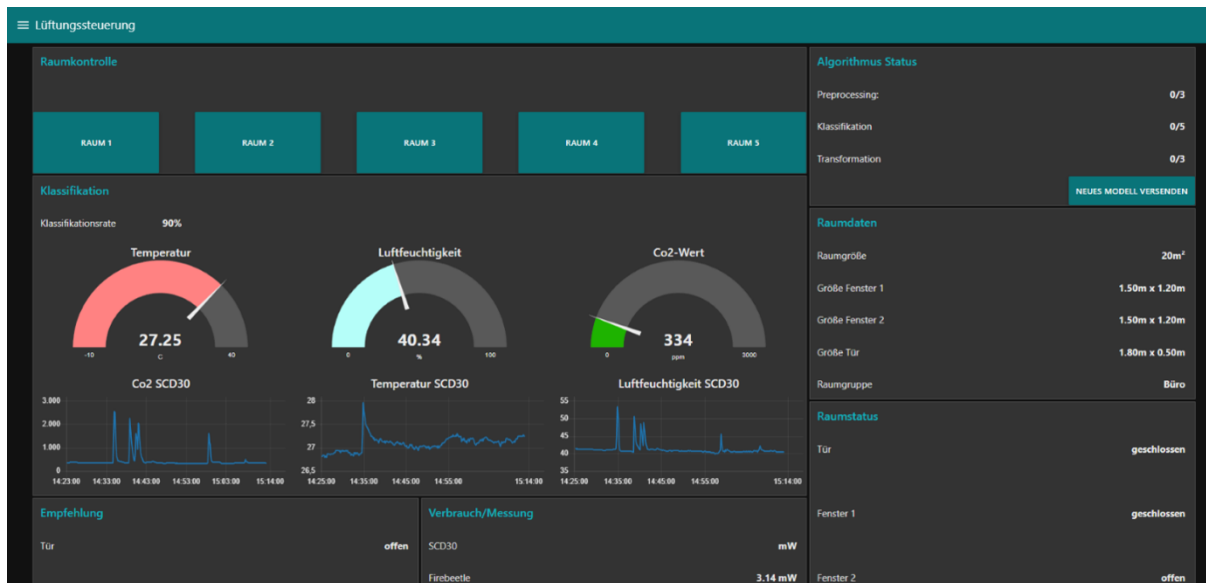


Abbildung 16: Luftqualitätsüberwachung und Anomaliedetektion

Der Algorithmus basiert auf einem maschinellen Lernansatz, bei dem über die Sensorik gemessene Lüftungsdaten gesammelt werden, aus welchen ein zu den Räumlichkeiten passendes Modell zur Überwachung der Luftqualität und Erkennung von Anomalien geschaffen wird. Das System soll eigenständig die Lüftungsverhältnisse prüfen und im Bedarfsfall Anpassungen durch das Ventilationssystem vornehmen können. Grundlage dafür ist ein maschinelles Lernmodell, welches gefördert durch mehrere im Gebäude platzierte „AI-on-Device-Boards“ trainiert wird. Ein zentraler Aspekt des Systems ist die Fähigkeit, Anomalien zu detektieren und auf diese unmittelbar zu reagieren. Zu möglichen Anomalien zählen beispielsweise Rauch im Brandfall oder die Freisetzung von Gefahrenstoffen.

Ein weiteres Ziel ist die Nutzung des vom Umwelt-Campus entwickelten „AI-on-Device“-Boards in Verbindung mit FL, welches ein lokales Netzwerk aufspannt, zu dem sich mehrere Boards verbinden, um gemeinsam im Bedarfsfall neue Raumlufmodelle zu erstellen. Dabei berechnen die einzelnen Nodes im Netzwerk, soweit ihre Beteiligung verifiziert ist, Teilmodelle auf Basis ihrer lokalen Daten, welche im Anschluss auf dem Zielsystem aggregiert und zu einem gemeinsamen Modell zusammengefasst werden.

Besonderes Augenmerk wird auf die Verbindung des Lüftungssystems mit Prinzipien der Privatheit und Sicherheit gelegt, insbesondere einer dezentralen Zero-Trust-Architektur, basierend auf einem bereits publizierten Ansatz (Creutz & Dartmann, 2023). Die Implementierung des Zero-Trust-Prinzips schützt somit den gesamten förderierten Lernprozess vor Eingriffen von außen, was insbesondere im industriellen Kontext, etwa in Lagerhallen und Fertigungsräumen, essenziell ist. Ein ungeschützter Zugang könnte zur Einspeisung fehlerhafter Daten führen, die das Lüftungssystem manipulieren und unerwünschte Ereignisse auslösen könnten.

Zu den möglichen Folgen solcher Manipulationen gehört das gezielte Aussetzen von erkannten Fehlerfällen und die Aktivierung von Sicherheitsmaßnahmen, wenn das Lüftungssystem mit einem Erkennungssystem, wie einer Brandschutzanlage, gekoppelt ist. Zudem könnte eine fehlerhafte Raumlufanpassung den Maschinenbetrieb stören, was schwerwiegende Störungen des Fertigungsprozesses zur Folge hätte. Eine willentliche Aussetzung des Überwachungssystems

könnte schwer behebbare Störungen des gesamten Fertigungsbetriebs zur Folge haben. Abbildung 17 verdeutlicht das Konzept. Authentifizierten Clients wird der Zugang gewährt, während Clients ohne passende Verifizierung abgeblockt werden.

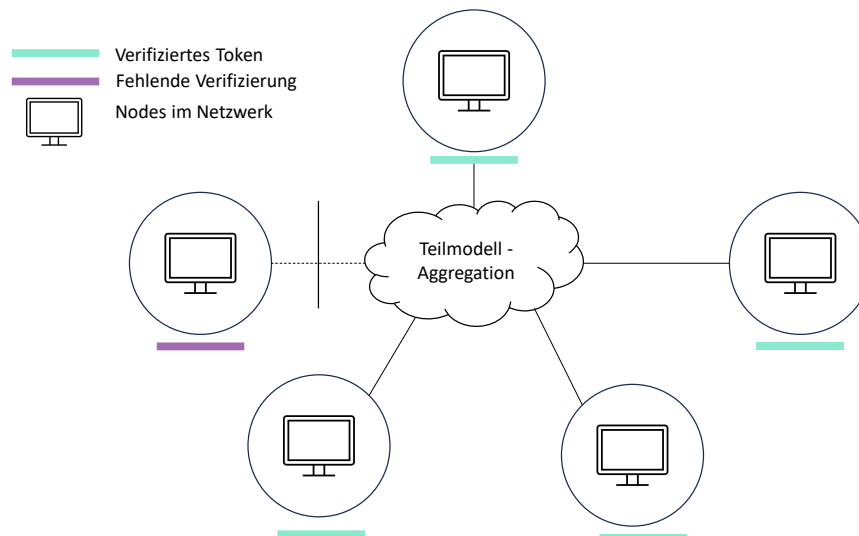


Abbildung 17: Zero-Trust und föderiertes Lernen für Luftqualität

Um die Sicherheit zu erhöhen, muss sich jeder Client-Edge-Knoten, bevor er in den föderierten Lernprozess integriert wird, über eine Cloud-Infrastruktur, welche Teil des dezentralen Zero-Trust-Systems ist, authentifizieren, um bei der Modellberechnung mitwirken zu können. Sowohl das AI-on-Device-Board als auch die Kommunikationsebene der einzelnen Nodes sowie ein Prototyp des Lüftungsdemonstrators existieren bereits. Auch eine Open-Source-Simulation des Zero-Trust-Ansatzes ist schon öffentlich zugänglich (Creutz & Dartmann, 2023).

Wie bereits beschrieben, steht bisher nur eine Raumkonfiguration zur Verfügung. Das System soll um mehrere Raumkonfigurationen erweitert werden und ein System entwickelt werden, das die Raumluftbestimmung über diese Raumkonfigurationen föderiert lernt. Außerdem werden verschiedene Ansätze zur Bestimmung des Lüftungszustandes untersucht. Geplant sind zum jetzigen Zeitpunkt Dynamic-Time-Warping (DTW) und Long-Short-Term-Memory (LSTM).

3.8 Bewertung der Kosten

Anwendungsbereich	Alle Arten kommerzieller ECS
Bewertungskriterien	Kosten und Erlöse
Kennzahlen und Metriken	Kosten
<ul style="list-style-type: none"> ▪ Zur Evaluation, ob sich ein auf ECS basierendes Geschäftsmodell langfristig und gegebenenfalls im Vergleich zu Cloud-Systemen rechnet, müssen die Kosten des ECS erfasst werden. ▪ Durch ihren verteilten Charakter existiert eine Vielzahl verschiedener Kostenkriterien entlang des Kontinuums von Edge zu Cloud. ▪ Kostenmanagement im ECS ist ein integrativer und kultureller Prozess, der die Mitwirkung verschiedener Stakeholder erfordert. 	

Die Kosten eines ECS stellen ein wichtiges Entscheidungskriterium zur Umsetzung einer Edge-Cloud-Anwendung dar. Ebenso kann die Gestaltung des Systems von Kosten-Nutzen-Erwägungen

gen abhängen. Der verteilte Charakter von ECS führt zu einer erhöhten Komplexität bei der Erfassung und Bewertung relevanter Kostenkriterien. Während in Cloud-Systemen üblicherweise die Betriebskosten im Vordergrund stehen, ergeben sich durch den Einsatz von Edge-Geräten vermehrt Investitionsausgaben, beispielsweise für die Beschaffung und Installation.

Die durch die Umsetzung eines ECS entstehenden Kosten lassen sich grob in fünf Kostenkategorien unterteilen (Kilic & Engelhardt, 2025). Eine Übersicht wird in Abbildung 18 gezeigt. Einrichtungskosten umfassen die Anschaffungs- und Installationskosten für Edge-Geräte sowie Lizenzgebühren, Konfigurationskosten und Schulungskosten entlang des Cloud-Edge-Kontinuums. Während des Betriebs entstehen Kosten für die physische Wartung und Energiekosten von Geräten und Netzwerk an der Edge sowie laufende Gebühren für die Nutzung von Cloud-Diensten und Datenübertragung in die Cloud. Leistungs- und Effizienzfaktoren können sowohl bei Edge- als auch bei Cloud-Komponenten indirekt Kosten verursachen. Der Einsatz von Edge-Geräten kann aufgrund begrenzter Rechenkapazitäten zu längeren Verarbeitungszeiten und eingeschränkter Funktionalität führen. Durch die Nutzung von Cloud-Services entstehen Produktivitätsverluste durch Latenzzeiten und Bandbreitenprobleme, die insbesondere bei Echtzeitanwendungen zu Verzögerungen führen können. Zu den Sicherheits- und Compliancekosten gehören die Ausgaben für Maßnahmen, die die physische Sicherheit und Verfügbarkeit der Geräte, Maßnahmen zur Cybersicherheit sowie die Einhaltung von Compliance-Vorgaben im Bereich der Edge gewährleisten. Auf Seiten der Cloud ist der Cloud-Anbieter für physische Schutzmaßnahmen verantwortlich. Zusätzliche Schutzmaßnahmen wie Virtual-Private-Networks, Firewalls und Verschlüsselungsmechanismen müssen jedoch teilweise zusätzlich bezahlt werden. Zudem existieren in ECS Kostenrisiken. Diese sind durch ihre Eintrittswahrscheinlichkeit multipliziert mit den vorgesehenen Kosten zu bemessen. Diese Kosten treten im Idealfall nicht auf. Zu den Kostenrisiken gehören Kosten, die durch eine unvorhergesehene notwendige Skalierung von Edge-Hardware oder Cloud-Diensten auftreten können. Weiterhin stellen notwendige Zahlungen durch Datensicherheitsvorfälle oder Haftungsansprüche (s. Abschnitt 3.9) ein Kostenrisiko dar.

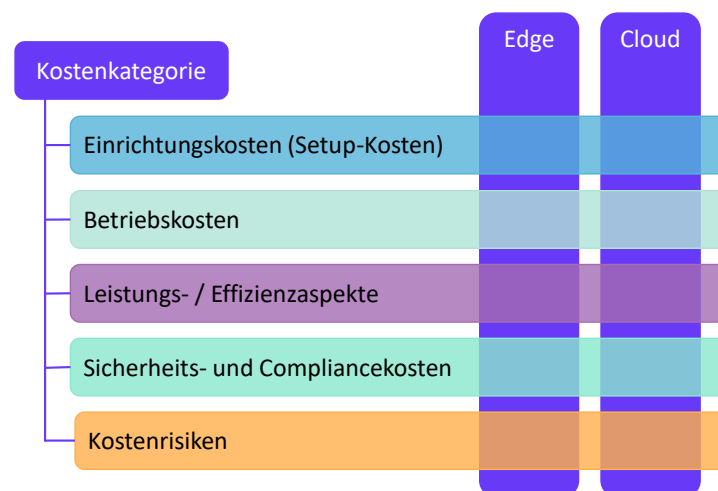


Abbildung 18: Kostenkategorien in ECS nach Kilic und Engelhardt (2025)

Auf Basis der Diversität möglicher Kostenfaktoren werden für das Kostenmanagement in ECS die Beiträge verschiedener Stakeholder benötigt. Es bedarf daher einer interdisziplinären Zusammenarbeit verschiedener Personen in den Organisationen. Zudem sollte eine Kultur geschaffen werden, die ein Bewusstsein über Kostenfaktoren wie beispielsweise Lizenzgebühren vermittelt. Weitere Informationen über das Kostenmanagement in ECS sowie eine Checkliste für Kostenfaktoren sind in der Publikation von Kilic und Engelhardt (2025) zu finden.

3.9 Ansätze zur rechtlichen Bewertung

Anwendungsbereich	Alle Arten kommerzieller ECS
Bewertungskriterien	Einhaltung regulatorischer Pflichten
Kennzahlen und Metriken	Rechtliche Einschätzung von Datenschutz, Datenhoheit, Haftung, Datenbereitstellungspflicht
<ul style="list-style-type: none"> ▪ Die rechtliche Bewertung von ECS ist notwendig, um eine langfristige rechtskonforme Nutzung zu ermöglichen und rechtliche Risiken bei der Umsetzung und Nutzung eines ECS in seinem Anwendungskontext zu minimieren. ▪ Bei der rechtlichen Bewertung eines ECS ist oftmals eine individuelle Prüfung vertraglicher Regelwerke zur Bewertung der Implikationen für Haftung und Datenhoheit notwendig. ▪ Aus den rechtlichen Bewertungen können sich technische Implikationen, beispielsweise hinsichtlich der Gestaltung technischer Sicherheitsmaßnahmen oder der Umsetzung von Standards und Schnittstellen zur Datenbereitstellung ergeben. 	

Der Einsatz von ECS in der Praxis ist mit vielfältigen rechtlichen Herausforderungen verbunden. Besonders im Fokus steht die rechtssichere Verarbeitung personenbezogener Daten sowie sensibler Unternehmensinformationen. Von zentraler Bedeutung ist in diesem Zusammenhang auch die Wahrung der Datenhoheit innerhalb komplexer, vernetzter Systemarchitekturen. Zunehmend relevant werden haftungsrechtliche Aspekte: Je stärker die technischen und organisatorischen Abhängigkeiten zwischen verschiedenen Akteuren in Cloud-Edge-Szenarien ausgeprägt sind, desto dringlicher stellt sich die Frage, wer im Fall von Systemausfällen, Störungen oder Sicherheitsvorfällen für etwaige Schäden einzustehen hat.

Datenschutzrechtliche Anforderungen in Edge-Cloud-Szenarien

Die Einhaltung datenschutzrechtlicher Anforderungen stellt im Kontext von ECS eine besondere Herausforderung dar. Diese Systeme zeichnen sich häufig dadurch aus, dass mehrere Akteure – wie Gerätehersteller, Cloud-Dienstleister und Nutzende – an der Erhebung, Verarbeitung und Übermittlung von Daten beteiligt sind. Dabei werden Daten häufig über Unternehmens- und sogar Landesgrenzen hinweg ausgetauscht, was die Anforderungen an eine rechtskonforme Verarbeitung deutlich erhöht.

Eine zentrale Rolle spielt dabei die Frage der datenschutzrechtlichen Verantwortlichkeit, wie sie in Art. 4 Nr. 7 der EU-Datenschutz-Grundverordnung (DSGVO) definiert ist. Der Verantwortliche ist derjenige, der allein oder gemeinsam über die Zwecke und Mittel der Datenverarbeitung entscheidet. Als primärer Adressat der datenschutzrechtlichen Pflichten trägt er die Verantwortung dafür, dass die gesetzlichen Anforderungen eingehalten werden. Diese Entscheidungsbefugnis macht den Verantwortlichen zur Schlüsselfigur bei der Sicherstellung der Datenschutzkonformität. Im Gegensatz dazu handelt der Auftragsverarbeiter ausschließlich auf Weisung des Verantwortlichen und ist nicht befugt, eigenständig über Zwecke und Mittel der Verarbeitung zu entscheiden. In Edge-Cloud-Szenarien wird die Zuordnung von Verantwortlichkeiten besonders komplex, wenn mehrere Akteure gemeinsam über die Zwecke und Mittel der Datenverarbeitung bestimmen. In einem solchen Fall gelten sie als gemeinsam Verantwortliche im Sinne von Art. 26 DSGVO und sind verpflichtet, klare Vereinbarungen über ihre jeweiligen Verantwortungsbereiche zu treffen, um die datenschutzrechtliche Konformität sicherzustellen. Zudem stellt sich die Herausforderung, dass sowohl Betreiber der Edge-Geräte als auch der Cloud-Dienstleister je nach Konstellation unterschiedliche Rollen einnehmen können – sei es als Verantwortliche, gemeinsam Verantwortliche oder Auftragsverarbeitende. Eine klare Abgrenzung dieser Rollen ist daher

wichtig, um Rechtsunsicherheiten zu vermeiden und die Einhaltung der datenschutzrechtlichen Vorgaben zu gewährleisten.

Zusätzliche datenschutzrechtliche Anforderungen ergeben sich bei der grenzüberschreitenden Verarbeitung personenbezogener Daten. Die Übermittlung solcher Daten in Drittstaaten, also Staaten außerhalb der EU und des EWR (López & Razauskas, 2024), unterliegt den strengen Vorgaben der Art. 44 ff. DSGVO. Zu den Voraussetzungen zählen insbesondere die Verwendung von Standardvertragsklauseln oder die Berücksichtigung von Angemessenheitsbeschlüssen, die sicherstellen, dass personenbezogene Daten in einem Drittstaat einen Schutz genießen, der mit den europäischen Datenschutzstandards vergleichbar ist.

Aber auch innerhalb der EU kann es ergänzende nationale Regelungen geben, die bei der Planung und Umsetzung solcher Systeme zu beachten sind. Dies gilt insbesondere für die Verarbeitung besonders sensibler Daten wie Gesundheitsdaten. Nach Art. 9 Abs. 4 DSGVO können die Mitgliedstaaten zusätzliche Bedingungen und Einschränkungen für die Verarbeitung solcher Daten festlegen, was in der Praxis zu erheblichen Unterschieden im Datenschutzrecht führen kann.

Schließlich spielt die Datensicherheit bei ECS eine zentrale Rolle. Sie wird maßgeblich durch Art. 25 DSGVO geprägt, der „Datenschutz durch Technikgestaltung“ und „datenschutzfreundliche Voreinstellungen“ fordert. Verantwortliche sind verpflichtet, technische und organisatorische Maßnahmen (TOM) zu treffen, um Risiken für die Rechte und Freiheiten der Betroffenen zu minimieren. Dazu gehören Maßnahmen wie Verschlüsselung, Pseudonymisierung und Zugriffskontrolle. In verteilten Systemen sind zudem die sichere Übertragung von Daten zwischen Edge- und Cloud-Ebene sowie der Schutz vor Angriffen auf unter Umständen weniger gesicherte Edge-Geräte entscheidend, um dem Stand der Technik zu entsprechen und die Anforderungen der DSGVO zu erfüllen.

Haftungsrechtliche Überlegungen

ECS zeichnen sich durch eine mehrschichtige Datenverarbeitung aus, die sowohl dezentral auf Edge-Geräten als auch zentral in der Cloud erfolgt. Diese Struktur impliziert die Beteiligung mehrerer Akteure, z. B. Betreibende von Edge-Geräten und Cloud-Dienstleister, und wirft zwangsläufig die Frage auf, wer im Falle eines Systemausfalls oder eines Cybersicherheitsvorfalls für den entstandenen Schaden haftet. Ein zentraler Anknüpfungspunkt für die Haftungsfrage ist das Vertragsrecht. Verträge zwischen Nutzenden oder Betreibenden von Edge-Geräten und Cloud-Anbietern regeln häufig die Haftungsverteilung für Schäden wie Systemausfälle, Datenverluste oder Sicherheitsverletzungen. Die rechtliche Beurteilung solcher Ereignisse hängt maßgeblich von den vertraglichen Vereinbarungen ab. Sichert der Cloud-Anbieter beispielsweise Verfügbarkeiten oder Sicherheitsstandards zu und werden diese nicht eingehalten, kann dies zu Schadensersatzansprüchen führen. Ebenso treffen Betreibende von Edge-Geräten aufgrund ihrer dezentralen Rolle spezifische Pflichten, wie etwa die Gewährleistung der Integrität und Verfügbarkeit der lokalen Datenverarbeitungsprozesse. Insgesamt ist festzuhalten, dass die Beurteilung von Haftungsszenarien in ECS stets einer Einzelfallbetrachtung bedarf. Dabei sind auch individualvertragliche Haftungsbeschränkungen oder Haftungsausschlüsse zu berücksichtigen.

Datenhoheit

Datenhoheit bezeichnet die tatsächliche Kontrolle und die rechtliche Möglichkeit, den Zugang zu Daten, ihre Verwendung und ihre Weitergabe zu bestimmen. Sie umfasst sowohl die technische

als auch die organisatorische Befugnis, Daten effektiv zu verwalten und ihre Verwendung zu kontrollieren. Im rechtlichen Kontext schließt der Begriff die Absicherung dieser Kontrolle durch vertragliche Vereinbarungen oder gesetzliche Rahmenbedingungen ein. In ECS stellt sich die Frage der Datenhoheit in besonderem Maße, da Daten dezentral auf Edge-Geräten erzeugt und zentral in der Cloud verarbeitet werden. Diese Architektur führt zu Unsicherheiten darüber, wer Zugriff auf die Daten hat und wer letztlich über deren Verwendung entscheidet.

Ein zentrales Problem ergibt sich aus der Vielzahl der beteiligten Akteure: Geräteherstellender, Cloud-Anbieter und Nutzende können jeweils die Kontrolle über die Daten beanspruchen. Ohne klare vertragliche Regelungen drohen Konflikte zwischen den Interessen der Beteiligten. Während Cloud-Anbieter häufig die Speicherung und Verwaltung der Daten übernehmen, sollte die Kontrolle idealerweise bei den Nutzenden verbleiben, sofern dies klar vereinbart ist. Da es keine absoluten Rechte an Daten gibt, die eine umfassende rechtliche Verfügungsgewalt garantieren, müssen Zugriffs- und Nutzungsrechte vertraglich geregelt werden. Solche Vereinbarungen regeln, wer auf die Daten zugreift und sie zu welchen Zwecken verarbeiten darf. Gleichzeitig müssen alle technischen Maßnahmen zur Sicherung der Datenhoheit, wie Verschlüsselung oder Zugriffskontrollen, mit den vertraglichen Regelungen abgestimmt werden. Nur so können Widersprüche vermieden und eine effektive Kontrolle über die Daten gewährleistet werden.

Die Bedeutung des Data-Act im Kontext von ECS

Ziel des Data-Act ist es, den Zugang zu und die Nutzung von Daten aus vernetzten Produkten und Diensten zu erleichtern. Durch die Schaffung rechtlicher und technischer Rahmenbedingungen soll die Basis für datengetriebene Innovationen und Geschäftsmodelle gestärkt werden. Insbesondere für Edge-Cloud-Szenarien bietet der Data-Act das Potenzial, die Zusammenarbeit zwischen den beteiligten Akteuren zu verbessern und den Datenfluss effizienter zu gestalten.

Ein zentrales Anliegen des Data-Act ist die Erhöhung der Datenverfügbarkeit. Nutzenden und Dritten soll ein rechtssicherer Zugang zu Daten ermöglicht werden, die in vernetzten Produkten oder Diensten anfallen. Diese Regelung könnte dazu beitragen, den Datenfluss zwischen Endgeräten und Cloud-Systemen zu optimieren, was die Entwicklung neuer datengetriebener Anwendungen und Dienste in solchen hybriden Architekturen erleichtern würde. Darüber hinaus könnte der Data-Act dazu beitragen, die Rechtsbeziehungen zwischen den beteiligten Akteuren zu konkretisieren. In Edge-Cloud-Konstellationen, in denen Daten häufig von unterschiedlichen Parteien wie Herstellern und Dienstleistern verarbeitet werden, verlangt der Data-Act künftig eine klare vertragliche Abstimmung. Dies könnte nicht nur die Rechtssicherheit erhöhen, sondern auch die Zusammenarbeit zwischen den beteiligten Akteuren erleichtern. Ein weiterer zentraler Aspekt des Data-Act ist die Förderung der Interoperabilität zwischen verschiedenen Systemen und Anbietern. Anbieter werden verpflichtet, technische Standards einzuhalten, die den Wechsel zwischen Cloud-Anbietern erleichtern und die Integration verschiedener Systeme unterstützen sollen. Diese Maßnahme kann dazu beitragen, Datensilos zu reduzieren und die Nutzung heterogener Plattformen zu vereinfachen, was die Flexibilität und Effizienz von Cloud-Edge-Szenarien erhöhen würde.

4 Aktuelle Herausforderungen bei der Bewertung von Edge-Cloud-Systemen

Bei der Bewertung von Edge-Cloud-Systemen (ECS) treffen Anwendende auf verschiedenste Problemstellungen und Herausforderungen. Ein Bewusstsein für diese Herausforderungen ist entscheidend, um diese im eigenen Bewertungsprozess zu berücksichtigen und etwaige Maßnahmen ableiten zu können. Gleichmaßen zeigen die aktuellen Herausforderungen weitere Forschungs- und Entwicklungsbedarfe auf. Aktuelle Herausforderungen bei der Bewertung von ECS (s. Abbildung 19) werden an dieser Stelle kurz zusammengefasst und im Laufe des Abschnittes tiefergehend erläutert.



Abbildung 19: Herausforderungen bei der Bewertung von Edge-Cloud-Systemen

Allgemein ist die Evaluierung von ECS in hohem Maße abhängig vom spezifischen Anwendungsfall, da je nach Zielstellung unterschiedliche Anforderungen an Hardware, Netzwerktopologien oder die Skalierbarkeit gestellt werden. Aus wirtschaftlicher Sicht sind die Leistungsvorteile von ECS mit den im Vergleich zu zentralisierten Systemen erhöhten Kosten gegenüberzustellen. Ebenso muss sichergestellt werden, dass neben den anfallenden Kosten ebenso die Komplexität eines ECS durch die umsetzende Organisation verwaltet werden kann, um die gegebenen Potenziale zu realisieren. Weiterhin ist es oftmals nicht möglich, die Evaluation eines ECS unabhängig von den Rahmenbedingungen durchzuführen. Vielmehr hängt die Entscheidung für die Nutzung von Edge, Cloud oder einer hybriden Lösung maßgeblich von der existierenden IT-Strategie des Unternehmens ab.

Durch den stetigen Wandel rechtlicher und gesellschaftlicher Erwartungen an den Umgang mit Daten entstehen kontinuierlich steigende Anforderungen an ECS, die durch organisatorische und technische Maßnahmen erfüllt werden müssen. Beispielhaft sind hier die Datenschutzgrundverordnung und der Data-Act zu nennen. Weiterhin ergibt sich ein sogenanntes „Moving-Target“ bei der Bewertung der rechtlichen Kriterien, sodass potenziell ein kontinuierliches Monitoring des rechtlichen Umfeldes erfolgen muss. Die Implementierung von ECS erfordert ein hohes Knowhow im Bereich der Informatik und Netzwerktechnik, was auf Seiten der Anwendenden und Systemintegratoren oftmals nicht vorhanden ist und die Hinzunahme weiterer IT-Expertinnen und -Experten bei der Umsetzung erfordert. Auch bei der Einschätzung weiterer weicher Kriterien wie etwa der Datensouveränität und der Vertrauenswürdigkeit von ECS fehlt es derzeit an relevantem

Knowhow der beteiligten Stakeholder, um eine informierte Entscheidung über die Systemauslegung zu treffen.

Eine weitere Herausforderung stellt der Transfer von Laborergebnissen in die Praxis dar. Oftmals unterscheiden sich die Praxisbedingungen stark von der idealisierten Laborwelt, sodass sich beispielsweise das Verhalten von Methoden im realen Einsatz nur schwer vorab einschätzen lässt. Darüber hinaus wird die für eine Bewertung notwendige Transparenz im Edge-Cloud-Kontinuum (ECC) durch fehlende einheitliche Standards für Leistungsparameter, mangelnde Dokumentation der Ressourcennutzung, Heterogenität der Hardware und unterschiedliche Bilanzierungsmodelle erheblich erschwert. Insbesondere führt dies zu Hindernissen bei der effizienten Planung und Verwaltung von Edge-Ressourcen durch Unternehmen.

Die Auslegung der Bewertung ist stark vom Anwendungsfall abhängig

Eine zentrale Charakteristik von ECS ist deren hohe Heterogenität. Systemkonfigurationen werden üblicherweise auf konkrete Anwendungsfälle zugeschnitten und greifen insbesondere auf Ebene der „Device-Edge“ auf spezialisierte Hardware zurück. Beispiele hierfür sind spezialisierte IoT-Geräte oder Kamerasysteme in der industriellen Produktion. Gleichmaßen ergeben sich zwischen den ECS Unterschiede in den verwendeten Netzwerktopologien, Sicherheits- und Souveränitätsanforderungen sowie hinsichtlich der Anforderungen in den Bereichen Skalierbarkeit, Latenz oder Flexibilität.

ECS müssen daher im Kontext konkreter Anwendungsfälle und deren Anforderungen evaluiert werden. Dies führt einerseits zu einer hohen Komplexität im Evaluationsprozess. Beispielsweise müssen für eine spezifische Anwendung folgende Entscheidungen getroffen werden:

- Die Auswahl zu evaluierender Bewertungskriterien.
- Die Definition konkreter Metriken.
- Die Auswahl und Anwendung von Mess- und Bewertungsverfahren in Abhängigkeit von der Systemkonfiguration.
- Die Abwägung verschiedener Bewertungskriterien gegeneinander im Rahmen einer multikriteriellen Entscheidungsfindung.

Andererseits führt die hohe Heterogenität von ECS zu einer geringen Übertragbarkeit bereits existierender Messungen und Bewertungen auf verwandte Systemkonfigurationen oder Problemstellungen.

Abwägung von Komplexität und Kosten gegenüber dem Leistungsversprechen von ECS

Wenngleich ECS verschiedene Vorteile gegenüber zentralisierten Architekturen bieten, sollten zur Einführung in Unternehmen erwartete Performanzgewinne und Nutzen gegenüber der gesteigerten Komplexität und den daraus resultierten Kosten abgewogen werden. Denn nicht für alle Anwendungen stellt sich die Nutzung eines ECS als sinnvoll heraus. Für diese Abwägung existiert jedoch keine einfach zu berechnende Kennzahl. Vielmehr muss, wie bereits erwähnt, eine Reihe von Kriterien gegeneinander abgewogen werden.

Einer der Hauptgründe, ECS in Betracht zu ziehen, ist die Aussicht auf eine gesteigerte Effizienz durch dezentrale Datenverarbeitung. Dazu zählt beispielsweise ein potenzieller Performanzgewinn, der durch die Verlagerung von Rechenleistung näher an den Datenursprung erzielt wird. Dies ermöglicht es, Latenzzeiten zu reduzieren, schnellere Reaktionszeiten zu erreichen und

möglicherweise die Datenlast auf zentrale Rechenzentren zu verringern. Besonders in zeitkritischen Anwendungen kann diese Steigerung der Leistung einen Wettbewerbsvorteil bedeuten.

Demgegenüber stehen allerdings eine Reihe von Komplexitätshürden, die sich üblicherweise aus der Dezentralisierung der ECS ergeben und mit denen oftmals signifikante Kosten einhergehen. Hierzu gehören zunächst die Investitionskosten für Beschaffung und Integration von Edge-Geräten, Netzwerkinfrastruktur und weiterer benötigter Komponenten. Diese Dezentralisierung der ECS führt zudem dazu, dass zwecks Datenintegrität und -verfügbarkeit Lösungen für Sicherheits- und Kommunikationsprotokolle implementiert werden müssen. Zusätzlich sind grundlegende KI-Verfahren und die daraus entwickelte Software für die verteilte Ausführung auf Edge-Cloud-Instanzen anzupassen. Während des Betriebs ergeben sich erhöhte Bedarfe bei der Verwaltung, Synchronisierung und Koordination zwischen den einzelnen Instanzen im ECC. Diese Faktoren können Aspekte wie die reduzierte Latenz und lokale Datenverarbeitung negieren und das Systemverhalten beeinträchtigen. Da Edge-Geräte oft in schwer zugänglichen oder entfernten Standorten installiert sind, gestaltet sich die Wartung und Aktualisierung der Hardware als schwierig und kostspielig. Dies betrifft nicht nur die physische Wartung, sondern auch Software-Updates, die über das Netzwerk verteilt werden müssen. Solche Updates müssen oft ohne Ausfallzeiten durchgeführt werden, was zusätzliche Komplexität erzeugt. Darüber hinaus können sich Edge-Standorte in Umgebungen befinden, in denen die Netzwerkkonnektivität unzuverlässig ist, was die Stabilität der Systeme weiter beeinträchtigen kann.

Zusammenfassend besteht die Herausforderung darin, abzuwägen, ob der Nutzen von Edge-Computing, wie Echtzeitverarbeitung, geringere Latenzzeiten und reduzierte Bandbreitennutzung, die gesteigerte Komplexität und die damit verbundenen Kosten rechtfertigt. Diese Abwägung ist stark abhängig von den spezifischen Anwendungsfällen und den Anforderungen des Unternehmens. Für das Verfahren der KI-Planung benötigt es beispielsweise sehr komplexe Planungsdomänen und die Integration hierarchischer Ansätze, damit Vorteile einer verteilten Ausführung zur Geltung kommen.

Integration der Mehrwerte von Edge-Computing in die Organisationsstrategie

Bei der Bewertung von Edge-Cloud-Anwendungen stellt die strategische Ausrichtung auf Edge- oder Cloud-Computing eine wesentliche Herausforderung dar, da die Wahl zwischen Edge-, Cloud- oder hybriden Lösungen maßgeblich von der Zielsetzung und den spezifischen Anforderungen abhängt.

Ein zentraler Aspekt bei der Bewertung von Edge-Cloud-Anwendungen ist die Abhängigkeit von der Cloud-Strategie des anwendenden Unternehmens, welche die Integration und Nutzung von Edge- und Cloud-Komponenten maßgeblich prägt. Die Wahl, entweder hybride Modelle oder reine Edge-Ansätze zu verfolgen, beeinflusst dabei direkt, wie Datenverarbeitung und Ressourcen zwischen Edge und Cloud verteilt werden. Diese Entscheidung bringt spezifische Herausforderungen mit sich und erfordert eine fundierte Analyse sowie eine präzise Abstimmung mit der langfristigen IT-Strategie des Unternehmens. Insbesondere die Möglichkeit einer parallelen Nutzung von Edge-Computing und Cloud-Diensten eröffnet zwar umfassende Potenziale, birgt jedoch zugleich auch Risiken.

Zusammengefasst ist der strategische Umgang mit der Cloud eine wesentliche Herausforderung bei der Bewertung von Edge-Cloud-Anwendungen. Eine durchdachte Integration der Edge-

Cloud-Ansätze ist daher von zentraler Bedeutung, um deren Vorteile gezielt zu nutzen, potenzielle Risiken zu minimieren und die IT-Strategie des anwendenden Unternehmens optimal zu unterstützen.

Kontinuierliche Weiterentwicklung der regulatorischen und gesellschaftlichen Anforderungen an Privatheit und Datensouveränität

Im Zuge der digitalen Transformation haben Ansätze wie „Cloud-first“ oder „Cloud-only“ dazu geführt, dass nahezu alle digitalen Prozesse in der Cloud stattfinden – von der Informationsverarbeitung bis hin zum Datenaustausch. Dabei wird die Gefahr akzeptiert, die Kontrolle über persönliche Daten, deren Verarbeitung und Verbreitung zu verlieren (Fassnacht et al., 2023). Mit der kontinuierlichen Weiterentwicklung der regulatorischen und gesellschaftlichen Anforderungen an Privatsphäre und Datensouveränität wächst jedoch das Bewusstsein für die Notwendigkeit neuer Lösungen. Insbesondere in komplexen Ökosystemen spielt die Kontrolle über die eigenen Daten eine immer relevanter werdende Rolle. ECS bieten hier eine vielversprechende Möglichkeit: Sie ermöglichen es, Datenverarbeitung und -verwaltung nicht mehr ausschließlich in der Cloud, sondern direkt auf den Anlagen, also "on-the-Edge", durchzuführen. Dadurch kann die Kontrolle über persönliche Daten zurückgewonnen und den Anforderungen an Datenschutz und Datensouveränität besser entsprochen werden (Stadelbacher, 2020).

Bei der Entwicklung dieser ECS können Gesetze, Normen und Standards sowie gesellschaftlichen Anforderungen relevante Leitlinien liefern. Jedoch sind – u. a. aufgrund der Dynamik und vergleichswisen Neuartigkeit des Konzepts Datensouveränität und des dezentralen Datenteilens – die relevanten Gesetze, Normen und Standards noch nicht etabliert oder befinden sich zum Zeitpunkt der Veröffentlichung noch in einer Übergangsfrist; so wird beispielsweise der EU Data-Act erst nach einer Übergangsfrist von 20 Monaten am 12. September 2025 direkt anwendbares Recht, wenngleich seine Prinzipien heute schon zur Bewertung von ECS herangezogen werden könnten. Neben den aktuellen Entwicklungen in der Rechtsprechung verändern sich auch die gesellschaftlichen Erwartungen an den Schutz der eigenen Privatsphäre kontinuierlich.

Insgesamt zeigt sich, dass die kontinuierliche Weiterentwicklung der regulatorischen und gesellschaftlichen Anforderungen an Datensouveränität und Informationssysteme eine dynamische und komplexe Herausforderung darstellt. Unternehmen müssen sich ständig an neue Vorschriften und gesellschaftliche Erwartungen anpassen, um ihre Daten sicher und transparent zu verwalten und den Schutz der persönlichen Daten zu gewährleisten (Fassnacht et al., 2023). Die Entwicklung von ECS im Einklang mit den genannten Leitlinien kann einen wesentlichen Beitrag dazu liefern, Datensouveränität „by-design“ umzusetzen.

Qualifikation von Systemintegratoren und Anwendenden

Die Bewertung von Edge-Cloud-Anwendungen stellt sowohl die Umsetzenden als auch die Endanwendenden vor große Herausforderungen. Bei der Umsetzung industrieller Produktionsanlagen sind eine große Zahl von Stakeholdern aus unterschiedlichen Domänen involviert, die sowohl die Anforderungen an das technische System als auch dessen Umsetzung beeinflussen. Expertinnen und Experten aus den Bereichen Systemintegration, Sondermaschinenbau und Anlagenplanung verantworten hierbei maßgeblich die Konstruktion und Inbetriebnahme der Anlage, haben jedoch in der Regel wenig Erfahrung in der Netzwerktechnik oder Informatik, um die Anforderungen an Edge-Cloud-Datenverarbeitungssysteme abzuleiten. Häufig müssen dediziert IT-

Expertinnen und Experten hinzugezogen werden, um im Dialog mit Integratorinnen und Integratoren und Endanwendenden Anforderungen an die Datenverarbeitungssysteme abzuleiten, ein ECS auszulegen und entsprechend zu konfigurieren. Sowohl die Konfiguration als auch der physische Anschluss der Netzwerkteilnehmer wird hierbei durch die häufige Notwendigkeit des physischen Zugangs zu den Geräten deutlich erschwert. Die häufig sehr strengen Firewall-Vorschriften der anwendenden Unternehmen stellen eine weitere Hürde zur erfolgreichen Konfiguration von ECS dar, da Netzwerkteilnehmer zwischen dem Inter- oder Intranet (der Cloud) und lokalen Netzen (z. B. Roboternetzen) kommunizieren müssen. In industriellen Anlagen stellt die Integration mit bestehenden Systemen, z. B. MES- oder ERP-Systemen, die Integration vor eine weitere Herausforderung, da bestehende industrielle Datenverarbeitungssysteme in der Regel nicht auf das dynamische Deployment im ECC ausgelegt sind, sondern als reine Cloud- oder lokale Anwendungen konzipiert wurden. Die Umsetzung realer, lauffähiger Systeme erfordert somit häufig große Mengen an Schnittstellencode, der nur schwer wiederverwendbar ist.

Quantifizierung von weichen Kriterien wie Datensouveränität

Bei der Bewertung von weichen Kriterien in ECS, wie dem Konzept der Datensouveränität, ist eine Quantifizierung oftmals nicht möglich und deshalb häufig zunächst eine qualitative Betrachtung des Systems hinreichend. Datensouveränität ist die Fähigkeit, Regeln zur Datennutzung zu formulieren, Datenaustausch zu beeinflussen und dabei eigenständig entscheiden zu können, mit wem Daten geteilt werden sollen (Scherenberg et al., 2024). Effektiv implementierte Datensouveränität kann das benötigte Vertrauensniveau senken und damit interorganisationalen Datenaustausch fördern, indem Dateninhabende die Kontrolle behalten und Datenempfangende die Daten verifizieren können. Sobald die Datensouveränität nicht mehr gewahrt werden kann, also die Kontrolle über die eigenen Daten verloren geht, steigt das benötigte Vertrauensniveau zwischen den Teilnehmenden des Systems.

Zur Bewertung der Datensouveränität eines ECS und der Grenzen ihrer Implementierung fehlen jedoch aktuell noch etablierte Frameworks, mit denen sich die eingesetzten Technologien und deren Wechselwirkungen evaluieren lassen. Ohne solche Frameworks ist es schwierig, die verschiedenen Ansätze und Technologien objektiv zu vergleichen und ihre Eignung für spezifische Anforderungen zu beurteilen. Diese Situation führt dazu, dass die Auswahl eines geeigneten Systems, Standards oder einer Technologie eine hohe Expertise und tiefgehendes Fachwissen im Bereich der Datensouveränität und ihrer Implementierung erfordert. Für Entscheidungsträger und Organisationen, die sich mit der Implementierung von datensouveränen Lösungen befassen, bedeutet dies einen erhöhten Aufwand und eine größere Unsicherheit bei der Auswahl der optimalen Lösung.

Es ist daher dringend notwendig, dass in Zukunft umfassende Bewertungsrahmen entwickelt werden, um die Vergleichbarkeit und Transparenz von datensouveränen Systemen zu verbessern und die Entscheidungsprozesse zu erleichtern. Um dieser Herausforderung zu begegnen, ist die Entwicklung einheitlicher Frameworks und Metriken daher dringend geboten.

Übertragbarkeit von Laborergebnissen in die Praxis

Bei Edge-Cloud-Technologien handelt es sich um (neue und) komplexe Systeme, die je nach Zusammensetzung, Hardware und Anwendungsfall ein unterschiedliches Verhalten aufweisen können. Daher ist es nicht immer leicht, die Übertragbarkeit von Metriken zur Beurteilung solcher Systeme in die Praxis vorherzusagen. Viele Verfahren, Methoden und Modelle wurden bisher

lediglich für ECS im Labor- und Forschungsumfeld getestet, es existieren allerdings noch nicht immer entsprechende Realisierungen im Praxisumfeld. Dies ist bei der Beurteilung zu berücksichtigen.

Einige Aspekte zur Beurteilung von ECS können sehr gut unter Laborbedingungen oder im Forschungskontext simuliert werden, aber bei anderen Aspekten sind die realen Bedingungen durchaus signifikant unterschiedlich und eine Simulation ist nicht immer möglich. Dadurch wird es erschwert, die Generalisierbarkeit von Modellen oder entwickelten Methoden zu testen oder zu evaluieren. Inwieweit sich beispielsweise die Netzbelastung für die Datenübertragung oder der Energieverbrauch bestimmter Anwendungen in einem realen Szenario von einem Szenario unter Laborbedingungen unterscheidet, ist im Vorfeld oft unklar und schwer abzuschätzen. Insbesondere bei der Verwendung von KI-Methoden ist es sehr wichtig, reale Daten mit einzubeziehen, damit die Modelle auch in der Praxis für den jeweiligen Anwendungsfall verwendet werden können. Daher ist bei der Bewertung von ECS immer zu beachten, dass sich die betrachteten Methoden, Modelle und Metriken bei der Verwendung in der Praxis gegebenenfalls anders verhalten können als unter Laborbedingungen oder in der Simulation.

Transparenz der Systembeschreibung

Die Überwachung und Transparenz von Systemen im ECC stellt die Industrie vor erhebliche Herausforderungen. Einer der Hauptgründe dafür ist das Fehlen einheitlicher Standards zur Beschreibung von Leistungsparametern von Netzwerkteilnehmern. Falls relevante Parameter wie Latenzen, Abtastraten oder Rauschverhalten in Datenblättern enthalten sind, werden diese nicht in standardisierter, maschinenlesbarer Form angegeben und können in der Regel nicht zur Laufzeit vom entsprechenden Gerät abgefragt werden.

Ein weiteres Problem liegt in der mangelnden Transparenz bei der Ressourcennutzung der verschiedenen Systemkomponenten. Oft werden Informationen darüber nicht ausreichend dokumentiert, was es Unternehmen erschwert, ihre Ressourcen effizient zu planen und zu verwalten. Hinzu kommt, dass zeitkritische Parameter wie Latenzen häufig nicht dokumentiert werden. Diese fehlenden Informationen führen zu zusätzlichem Aufwand in den Bereich des Anlagenbaus und der Systemintegration, die daher die Latenzen und Reaktionszeiten selbst ermitteln müssen.

Die Heterogenität der eingesetzten Hardware und Rechenressourcen im ECC erschwert ebenfalls die Überwachung. Die Vielzahl unterschiedlicher Technologien behindert die Konsistenz und Skalierbarkeit von Überwachungslösungen. Gleichzeitig besteht oft eine Abhängigkeit von Closed-Source-Hardware, die von kommerziellen Anbietern stammt und deren Kommunikationsverhalten nicht optimiert oder angepasst werden kann. Diese Abhängigkeit schränkt die Flexibilität und Anpassungsfähigkeit der Systeme erheblich ein. Zusätzlich erschweren unterschiedliche Bilanzierungsmodelle für die Erstellung, den Betrieb und die Nutzung von Systemkomponenten eine ganzheitliche Betrachtung der Ressourcennutzung. Ohne standardisierte und transparente Parameter, die sowohl Hardware- als auch Softwareanforderungen abdecken, fehlen Unternehmen die Informationen, die sie für fundierte Entscheidungen und eine effiziente Ressourcennutzung benötigen.

5 Handlungsempfehlungen

Im Einklang mit dem vielfach von Führungspersonen referenzierten Prinzip „You can't manage what you can't measure“ (deutsch: „Was man nicht messen kann, kann man nicht steuern“) stellt eine Bewertung die Grundlage zur Validierung des Wertversprechens und zur Durchführung möglicher Optimierungen von Edge-Cloud-Systemen (ECS) dar. Nur wenn eine zielgerichtete Bewertung von ECS durchgeführt wird, können die auf konzeptueller Ebene formulierten Vorteile von ECS wie ein reduzierter Energieverbrauch, eine verbesserte Latenz oder eine erweiterte Datensouveränität in der Praxis belegt werden. Die Bewertung eines ECS ist entsprechend relevant, um potenziellen Kunden die Mehrwerte der entwickelten Lösung aufzeigen zu können. Weiterhin stellen Bewertung und Messung der Eigenschaften von ECS sowie die daraus abgeleiteten qualitativen und quantitativen Ergebnisse die Grundlage für die Entscheidungsfindung, beispielsweise hinsichtlich der Verteilung von Ressourcen im Edge-Cloud-Kontinuum (ECC), dar.

Allerdings zeigen sich im Rahmen der Arbeiten der Early-Adopter von ECS vielfältige Herausforderungen bei der Bewertung von ECS, die im vorangegangenen Abschnitt zusammengefasst wurden. Um diese zu adressieren und Anwendende künftig bei der Bewertung zu unterstützen, formuliert dieser Abschnitt konkrete Handlungsempfehlungen für verschiedene Akteursgruppen, aufgeteilt in zwei Handlungsfelder. Die Handlungsempfehlungen wurden gemeinsam durch die Expertinnen und Experten in der Task-Force definiert und ausgearbeitet und stellen somit ein zentrales Resultat der Arbeiten in der Task-Force dar. Tabelle 2 fasst die genannten Handlungsempfehlungen zusammen.

Im ersten Handlungsfeld werden Empfehlungen für die praktische Durchführung einer Bewertung von ECS präsentiert. Diese adressieren Akteure, die aktuell oder in naher Zukunft unmittelbar in die Bewertung von ECS involviert sind. Das erste Handlungsfeld richtet sich somit an Edge-Computing-Anwendende, Plattformbetreiber, Systemintegratoren sowie Software- und Hardwarehersteller. Sie erhalten Hinweise, wie eine Bewertung von ECS in Unternehmen etabliert werden kann und welche Bewertungskriterien dabei insbesondere zu berücksichtigen sind. Konkret umfassen die Empfehlungen die Abwägung von Aufwand und Nutzen, die Schaffung interner Bewertungsstandards über den gesamten Lebenszyklus oder die Einbeziehung von Aspekten wie Resilienz, Datenschutz und regulatorische Anforderungen in die Bewertungsstrategie.

Das zweite Handlungsfeld umfasst Empfehlungen, um die strukturellen Rahmenbedingungen zur Bewertung von ECS zu verbessern und damit eine nachhaltigere Befähigung der Bewertung von ECS zu erreichen. Die Empfehlungen richten sich dabei nicht nur an unmittelbar in die Bewertung Involvierte, sondern auch an Akteure, die einen mittelbaren Einfluss auf die Voraussetzungen der Bewertung von ECS nehmen können. Somit werden insbesondere Forschungseinrichtungen und politische Entscheidungsträger adressiert. Zu den Empfehlungen zählen konkret die Förderung des Aufbaus notwendiger Kompetenzen zur Bewertung, die Etablierung von Transparenzstandards und die semantische Standardisierung, die Verbesserung der Erklärbarkeit von (KI-) Systemen sowie die Schaffung von Angeboten, die die Einstiegshürden für ECS reduzieren. Die Handlungsempfehlungen werden folgend in dedizierten Abschnitten umfassender erläutert. Die primären Adressaten der Handlungsempfehlungen werden in den jeweiligen Abschnitten grafisch hervorgehoben.

Handlungsfeld	Handlungsempfehlung	Stakeholder					
		Edge-Computing-Anwender	Plattformbetreiber und Systemintegratoren	Hardwarehersteller	Softwarehersteller	Forschung	Politik
Planung und Durchführung der Bewertung	Aufwände und Nutzen der Messung abwägen	x	x	x	x		
	Interne Bewertungsstandards entlang des gesamten Lebenszyklus schaffen	x	x				x
	Resilienz der ECS bewerten	x	x				
	Neue regulatorische Aspekte in der Bewertung berücksichtigen	x	x				
	Aspekte der Privatsphäre und Souveränität verstärkt betrachten	x	x	x	x		
Nachhaltige Befähigung der Bewertung	Verpflichtende Transparenzstandards etablieren			x	x	x	x
	Semantische Modelle für die Modellierung von Produktionsressourcen standardisieren	x	x	x	x		x
	Erklärbarkeit von (KI-) Systemen verbessern		x		x	x	x
	Notwendige Kompetenzen schaffen	x	x			x	x
	Einstiegshürden für Edge-Computing und dessen Bewertung verringern	x	x			x	x

Tabelle 2: Handlungsempfehlungen für die Bewertung von Edge-Cloud-Systemen

5.1 Handlungsempfehlungen für die praktische Durchführung einer Bewertung von Edge-Cloud-Systemen

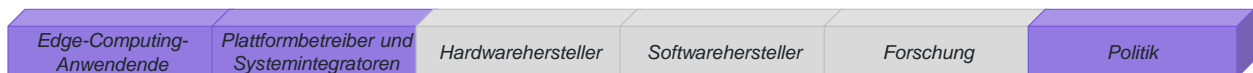
Dieser Abschnitt beschreibt Handlungsempfehlungen, um eine Bewertung von ECS im Sinne des aktuellen Erkenntnisstandes durchzuführen. Hierzu werden einerseits Hinweise für den Messablauf und die Etablierung der Bewertung innerhalb eines Unternehmens gegeben und andererseits spezifische Kriterien für die Bewertung von ECS hervorgehoben.

Aufwände und Nutzen der Messung abwägen



Mit der Messung und Bewertung von ECS geht ein vom Umfang abhängiger Overhead bei Planung, Umsetzung und Betrieb von ECS einher. Dies ist insbesondere der Fall, wenn keine Erfahrungswerte zur Verfügung stehen. Daher ist es wichtig, bei der Planung der aufzunehmenden Messwerte, Kennzahlen und Metriken den Aufwand der Messung im Vergleich zum Nutzen zu berücksichtigen. Insbesondere Messverfahren, die eine Installation neuer Messgeräte oder Software zur Durchführung experimenteller Evaluationen bedürfen, sollten hinsichtlich deren Notwendigkeit motiviert werden. Hinweise auf die Notwendigkeit zur Evaluation eines bestimmten Faktors liefert das Wertversprechen der durch das ECS realisierten Lösung.

Interne Bewertungsstandards entlang des gesamten Lebenszyklus schaffen

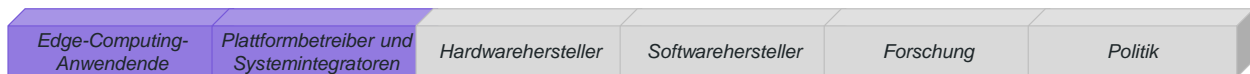


Unternehmen und Systemintegratoren sollten die im Whitepaper vorgestellten Bewertungskriterien aktiv nutzen und priorisieren, um die eigenen (geplanten) Edge- und Cloud-Deployments systematisch zu analysieren. Eine regelmäßige Selbsteinschätzung anhand dieser Kriterien ermöglicht es, Schwachstellen zu identifizieren, Potenziale zu erkennen und gezielte Maßnahmen zur Weiterentwicklung abzuleiten.

Wenngleich die relevanten Bestandteile der digitalen Lieferkette (u. a. Geräte für Endnutzende, Netzwerk, Datenzentren) derzeit noch nicht während ihres gesamten Lebenszyklus bemessen werden, ist dies für eine ganzheitliche Analyse von ECS notwendig. Zukünftig können Elemente wie der digitale Produktpass dazu beitragen, eine Lebenszyklusanalyse von Bestandteilen von ECS durchzuführen. Um die Bewertungen gleicher Komponenten für verschiedene Einsatzszenarien innerhalb eines Unternehmens wiederverwendbar zu machen, sollten vorausschauend standardisierte, unternehmensweite Metriken und Regeln zur Bewertung geschaffen werden. Ein Beispiel für die Notwendigkeit standardisierter Metriken stellt die Bemessung des Treibhauspotenzials dar. So wird zukünftig der Standort der ECS mittelbar durch den lokalen Strommix einen Einfluss auf die Emissionen des Systems besitzen. Sollte dieser für die relevanten Komponenten eines ECS standortübergreifend zur Verfügung stehen, kann dessen Bemessung als Unternehmensstandard definiert werden. Durch die standardisierte Beschreibung und Bewertung von Systemkomponenten können wiederverwendbare Profile erstellt werden. Diese Profile ermöglichen eine flexible Kombination und Zusammenführung bei der Simulation bestimmter Eigenschaften von ECS und reduzieren somit wiederkehrende Mess- und Bewertungsaufwände.

Wichtig ist dabei, den gesamten Lebenszyklus zur Erfassung der Ökobilanz einer Komponente zu betrachten. Die Umweltwirkung eines Produkts ist beispielsweise kumuliert entlang der Phasen Herstellung, Nutzung und Entsorgung zu betrachten. Die Herstellungsphase umfasst den Herstellungsaufwand von Design, Hardware, Software und die im Bereich elektronischer Komponenten tendenziell vernachlässigbaren Transportaufwände. Die Nutzungsphase umfasst Umweltwirkungen durch den Einsatz von Energie und anderen Wirkmitteln wie beispielsweise Kältestoffen. Die Entsorgungsphase umfasst die Aufwände zur Wiederverwendung von Systemkomponenten und enthaltenen Rohstoffen gemäß der R-Strategien (DIN) und die anfallenden Mengen an Elektroschrott unter Berücksichtigung der Recyclingquote. Ein Rahmenwerk für die Ökobilanz bzw. Lebenszyklusanalyse liefern DIN EN ISO 14040 bzw. 14044. Eine Voraussetzung dafür ist die Verfügbarkeit der notwendigen Daten. Entsprechend gehen mit der Schaffung von organisationsinternen Standards auch immer eine Betrachtung und gegebenenfalls Verbesserung des Managements von Nachhaltigkeitsdaten einher.

Resilienz der ECS bewerten



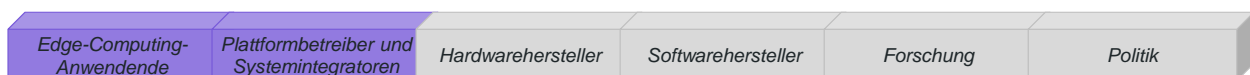
Die Resilienz von Softwaresystemen gilt als die Fähigkeit, trotz widriger Umstände wie Hardwareausfällen, Softwarefehlern, hoher Datenlasten oder Cyber-Angriffen weiterhin korrekt zu funktionieren. Ein resilientes System kann solche Ereignisse antizipieren, absorbieren, sich anpassen und sich schnell von diesen erholen.

Auch bei der Bewertung von ECS kann die Resilienz aufgrund der Vorteile der Edge ein weiteres wichtiges Entscheidungskriterium darstellen. So kann die Edge-Komponente als Bestandteil des verteilten Systems einzelne Aufgaben im „Stand-Alone-Modus“ durchführen. Dadurch kann eine größere Unabhängigkeit von zentralen Cloud-Diensten erreicht werden und beispielsweise eine kontinuierliche Datenverarbeitung und -analyse durchgeführt werden, selbst wenn die Verbindung zur Cloud temporär unterbrochen ist.

Ein weiterer wichtiger Aspekt der Resilienz von ECS ist die sichere Verfügbarkeit der Datenübertragungstrecken. ECS sollten so konzipiert sein, dass sie Daten in akzeptabler Latenzzeit übermitteln können, auch unter Belastung oder bei Netzwerkstörungen. Hierfür können redundante Kommunikationswege – gegebenenfalls auch mit verschiedenartigen Technologien, wie 5G/6G in Campus- oder öffentlichen Netzen, WiFi, Bluetooth, LiFi – oder intelligente Lastverteilungen eingesetzt werden, um sicherzustellen, dass Daten auch bei unerwarteten Ereignissen zuverlässig und schnell übertragen werden.

Insgesamt kann also der Bedarf nach Resilienz eines Softwaresystems einen wesentlichen Motivationsfaktor und eine Argumentationshilfe für den Einsatz von ECS darstellen. Nur eine anwendungsspezifische Bewertung des Faktors Resilienz kann die Vorteile von ECS aufzeigen.

Neue regulatorische Aspekte in der Bewertung berücksichtigen



Im Bereich der rechtlichen Bewertung ergeben sich durch die in den letzten Jahren entwickelten EU-Vorschriften im Bereich der Datenwirtschaft verstärkte Anforderungen an ECS. Aus dem

Data-Act und dem AI-Act folgen unmittelbare Ansprüche an die technische Gestaltung von ECS und die Rechenschaftspflichten von Unternehmen im Bereich der Entwicklung und Anwendung von ECS. Mit dem Data-Act müssen Daten, die durch vernetzte Produkte oder durch mit den Produkten verbundene Dienste generiert werden, den nutzenden Unternehmen einfach, sicher und direkt zur Verfügung gestellt werden. Die Verantwortung für die Sicherstellung der Datenverfügbarkeit liegt dabei beim Hersteller des vernetzten Produkts beziehungsweise Dienstes. Diese produktbezogenen Daten dürfen weder durch den Nutzenden oder durch eine vom Nutzenden beauftragte Drittpartei zur Entwicklung von Konkurrenzprodukten genutzt werden.

Mitunter können im Rahmen von ECS auch KI-Systeme entworfen werden, die unter die Kategorie der Hochrisiko-KI-Systeme des AI-Act fallen. Dies ist beispielsweise der Fall, wenn die KI-Systeme Sicherheitskomponenten von Produkten oder selbst Produkte im Bereich von beispielsweise Maschinen und Anlagen, Medizintechnik oder kritischer Infrastruktur darstellen. Für diese KI-Systeme existiert ein umfassender Pflichtenkatalog, der insbesondere für den KI-Anbieter Maßnahmen entlang der Phasen Entwicklung, Zulassung und Inverkehrbringen definiert. Die Pflichten der Betreiber sind in diesem Bereich hingegen überschaubar und umfassen unter anderem Informationspflichten.

Basierend auf der EU-Ökodesign-Verordnung wird der digitale Produktpass als ein zentrales Element zur Förderung der Kreislaufwirtschaft und der Nachhaltigkeit von Produkten gesehen, indem er umfassende Transparenz über die Materialien, die Herkunft und die Herstellungsprozesse eines Produkts schafft. Verschiedene Komponenten von ECS fallen in diese Verordnung. Kernaspekt des digitalen Produktpasses ist die technische Bereitstellung der zur Transparenz notwendigen Informationen. Derzeit ist jedoch noch nicht einheitlich definiert, wie ein digitaler Produktpass technisch realisiert werden kann.

Entsprechend gilt es für Unternehmen zukünftig, ihre individuelle Rolle auf Basis der neuen Verordnungen einzuschätzen und zu prüfen, ob die gemeinsam entwickelten Systeme in diese Verordnungen fallen. Sollte dies der Fall sein, sind je nach Systemauslegung die notwendigen technischen oder organisatorischen Maßnahmen zu treffen, um die Rechtskonformität der Lösungen sicherzustellen. Dazu gehört beispielsweise die Umsetzung notwendiger Schnittstellen für den Abruf von Daten oder die Einführung eines Risikomanagementsystems für den Einsatz von KI. Andersherum bieten ECS auch neuartige Chancen zur Realisierung der regulatorischen Anforderungen. So können präzisere Daten über den Produktentstehungsprozess erhoben und in den digitalen Produktpass integriert oder Daten einfacher zugänglich gemacht werden.

Aspekte der Privatsphäre und Souveränität verstärkt betrachten



Der Übergang zu ECS bietet Unternehmen die Chance, ihre Sicherheitskonzepte an aktuelle Herausforderungen anzupassen, insbesondere durch die Implementierung von Zero-Trust-Architekturen. Zero-Trust basiert auf dem Prinzip, dass keine Benutzenden oder Systeme pauschal als vertrauenswürdig angesehen werden. Stattdessen erfolgen eine kontinuierliche Überprüfung und Minimierung von Zugriffsrechten. Das macht dieses Modell besonders geeignet, die Sicherheit in verteilten, komplexen IT-Landschaften wie Edge-Cloud-Umgebungen zu erhöhen.

Die Einführung von Zero-Trust ist jedoch anspruchsvoll. Zu den zentralen Herausforderungen gehören die Vielzahl von Datenformaten, die Interoperabilität zwischen Systemen und die Integration bestehender Legacy-Infrastrukturen. Viele Anbieter setzen auf proprietäre Technologien, die eine nahtlose Migration erschweren. Dies kann zu Insellösungen führen, die die Vorteile eines einheitlichen Sicherheitsmodells einschränken. Auch Legacy-Systeme stellen eine Hürde dar, da sie häufig nicht für die Anforderungen moderner Sicherheitsarchitekturen konzipiert wurden.

Trotz dieser Herausforderungen ist die Einführung von Zero-Trust-Architekturen zu empfehlen, da sie eine zukunftssichere Alternative zu traditionellen, perimeterbasierten Ansätzen darstellen. Zero-Trust-Architekturen bieten die Möglichkeit, Sicherheitsrisiken zu reduzieren und den Schutz vor Bedrohungen zu verbessern, insbesondere in dynamischen und skalierenden IT-Landschaften wie ECS.

Die Bedeutung von Zero-Trust wird zudem durch politische Rahmenbedingungen unterstrichen. In den USA hat Ex-Präsident Biden im Mai 2021 ein Dekret verabschiedet, das alle Bundesbehörden dazu verpflichtet, bis Ende 2024 auf Zero-Trust-Architekturen umzusteigen (Biden, 2021). Dieses Dekret setzt einen deutlichen Impuls und zeigt, dass Zero-Trust nicht nur eine technologische Entscheidung, sondern auch ein wesentlicher Bestandteil nationaler Sicherheitsstrategien ist.

Vor diesem Hintergrund sollten Unternehmen und andere Einrichtungen den Übergang zu Zero-Trust als strategische Priorität betrachten, um Sicherheit in zunehmend komplexen IT-Landschaften zu gewährleisten und aktuellen wie zukünftigen Anforderungen gerecht zu werden. In der Entwicklung von ECS wird darüber hinaus auch die zielgerichtete Berücksichtigung von Aspekten der Privatsphäre und vor allem des Konzepts der Datensouveränität immer relevanter. Während Privatsphäre auf den Schutz personenbezogener Daten fokussiert, ist Datensouveränität die Fähigkeit, Regeln zur Datennutzung zu formulieren, Datenaustausch zu beeinflussen und dabei eigenständig entscheiden zu können, mit wem Daten geteilt werden sollen.

Die technische Integration von Mechanismen zur Sicherstellung von Datensouveränität findet zunehmend Anwendung in verschiedenen Forschungsprojekten. Unternehmen sollten Ansätze, die aktuell erforscht werden, aktiv unterstützen und dabei verfolgen, wie sich die Regulatorik sowie technische Möglichkeiten weiterentwickeln. Die Sicherstellung der eigenen Datensouveränität kann eine wichtige Grundlage darstellen, um eigenes Wissen zu schützen und zu kontrollieren, welche Parteien in welchem Umfang auf Daten und damit Wissen zugreifen können. Vor diesem Hintergrund sollten Maßnahmen zur Etablierung und Sicherstellung von Datensouveränität – sowohl auf regulatorischer als auch auf unternehmerischer Ebene – strategische Priorität genießen. Entsprechende Projekte müssen konsequent vorangetrieben und die Rolle von Edge-Computing für Datenschutz und Datensouveränität weiter untersucht werden.

5.2 Handlungsempfehlungen zur nachhaltigen Befähigung von Organisationen zur Umsetzung von Bewertungen der Edge-Cloud-Systeme

Dieser Abschnitt präsentiert Handlungsempfehlungen, um Organisationen nachhaltig zu einer Bewertung von ECS zu befähigen. Hierzu sind neben den an der Bewertung von ECS beteiligten Akteuren insbesondere die Politik zur Schaffung der notwendigen Rahmenbedingungen sowie die Forschung zur Entwicklung einfacher Bewertungsansätze gefragt. Konkrete Ansatzpunkte zur Erreichung dieses Ziels werden folgend beschrieben.

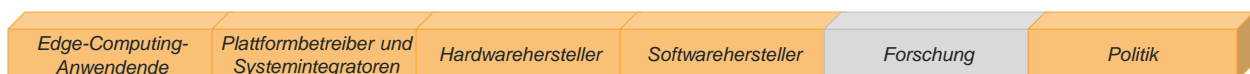
Verpflichtende Transparenzstandards etablieren



Um die Herausforderungen im Bereich der Transparenz und des Monitorings der Komponenten von ECS anzugehen, bedarf es einer stärkeren und gegebenenfalls verpflichtenden Standardisierung der technischen Beschreibungen. Die Schaffung von Standards für die technische Systembeschreibung von Netzwerkteilnehmern wie Sensoren, Aktoren oder Netzwerkhardware, aber auch z. B. Softwareplattformen für die Datenverarbeitung, ermöglicht die Umsetzung großflächig einsetzbarer, schlüsselfertiger Softwarelösungen für das Monitoring von Edge-Cloud-Kommunikationssystemen. Zu den relevanten Informationen gehören einerseits Leistungsparameter wie Latenzen, Durchsatz oder Abstraten und andererseits Angaben über die Ressourcennutzung. Eine Übersicht über relevante Kennzahlen zur Umweltbilanzierung liefern Gröger et al. (2024).

Die relevanten Informationen sollten von den Komponenten selbst maschinenlesbar und in einem standardisierten Datenformat zur Verfügung gestellt werden und zur Laufzeit abfragbar sein. Neben der Festlegung, was auf welche Art und Weise preisgegeben werden soll, sollten ebenso standardisierte Bilanzierungsmodelle zur Bemessung der jeweiligen Kennzahlen definiert werden, um die Vergleichbarkeit unter den Komponenten eines ECS sicherzustellen. Mit der Einführung des digitalen Produktpasses wird zukünftig die Transparenz über Informationen der Umweltwirkung von Produkten, darunter beispielsweise Elektronikkomponenten, verpflichtend. Allerdings erfolgt die Umsetzung der digitalen Produktpässe aufgrund divergierender Anforderungen teilweise branchen- oder sogar produktspezifisch. Die daraus resultierende Datenheterogenität erschwert die Zusammenführung zu einer Systemsicht. Verschiedene Initiativen arbeiten derzeit an der Standardisierung von digitalen Produktpässen. Hier ist es wichtig, die Interoperabilität zwischen den resultierenden Standards, gegebenenfalls auch mittels politischer Einwirkung, zu gewährleisten. Dabei sollte ein Diskurs hinsichtlich der neuartigen, durch ECS gegebenen, Möglichkeiten zur Beschreibung von Charakteristika entlang des Produktlebenszyklus stattfinden. So können beispielsweise in Echtzeit Produktionsdaten und Strommix zusammengeführt werden, um den realen CO₂-Fußabdruck eines Produkts zu ermitteln, oder Produkte während ihres gesamten Lebenszyklus verfolgt werden, um bessere Recycling- und Wiederverwendungsstrategien zu ermöglichen.

Semantische Modelle für die Modellierung von Produktionsressourcen standardisieren

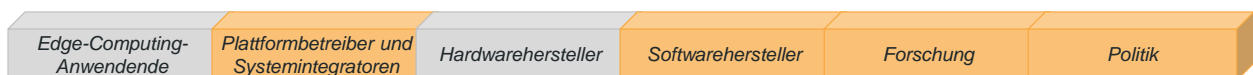


Die Möglichkeiten der verteilten Ausführung und die hohe verfügbare Rechenkapazität durch Einbindung verschiedener Instanzen bieten signifikante Vorteile für die Anwendung von KI-Planungsverfahren in einem Edge-Cloud-Kontinuum (ECC). Dabei können KI-Planungsverfahren sowohl für die Planung von Berechnungsprozessen innerhalb des ECC als auch für Steuerungs- und Produktionsprozesse bei den teilnehmenden Partnern genutzt werden. Eine große Herausforderung bei der Anwendung solcher Verfahren ist die Beschaffung und Modellierung der benötigten semantischen Informationen, die die verfügbaren Ressourcen beschreiben. Ohne dass eine detaillierte semantische Beschreibung der Ressourcen vorliegt, können KI-Planungsverfahren nicht angewandt werden.

In aktuellen Ansätzen werden standardmäßig separate und eigenständige semantische Modelle für Planungssysteme modelliert, die einer an das jeweilige Planungssystem angepassten Struktur folgen. Bei der Erstellung eines solchen semantischen Modells handelt es sich um einen sehr arbeitsaufwändigen Schritt. Die verwendeten Ressourcen werden normalerweise in vielen verschiedenen Unternehmen baugleich eingesetzt, sodass auch weitere Methoden von semantischen Modellierungen dieser Ressourcen profitieren können. Deshalb bietet die Nutzung generischer semantischer Modelle Potenziale, von denen nicht nur die KI-Planung, sondern auch andere KI-Verfahren oder Methoden profitieren können. Zur Erstellung solcher generischen Modelle ist jedoch eine Stakeholder-übergreifende Standardisierung der Modelle nötig. Die Struktur sollte dabei unabhängig von spezifischen Herstellern oder Betreibern sein. Dementsprechend erfordert eine solche Standardisierung die Zusammenarbeit von Plattformbetreibern, Systemintegratoren, Hardware- sowie Softwareherstellern und Betreibern der Produktionsressourcen, um generische und gleichzeitig passende Modelle zu identifizieren. Vor diesem Hintergrund müssen die eben genannten Stakeholder bereit sein, semantische Informationen bezüglich der Ressourcen auszutauschen und eine gemeinsame Standardisierung der Informationen anzustreben.

Standardisierung ist über die semantischen Modelle hinaus auch allgemein ein wichtiger Aspekt für die sichere und nachhaltige Umsetzung zukünftiger Edge-Cloud-Umgebungen. Einheitliche Schnittstellen und Protokolle verhindern Vendor-Lock-ins und ermöglichen somit den Wechsel zwischen verschiedenen Cloud-Anbietern oder erlauben flexible „Mix-and-Match“-Strategien. Dies schafft nicht nur technologische Unabhängigkeit, sondern erhöht auch die Flexibilität bei der Anpassung an neue Anforderungen.

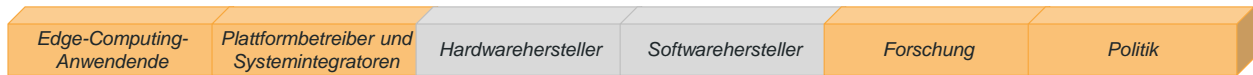
Erklärbarkeit von (KI-) Systemen verbessern



Die Implementierung erklärbarer, lokaler KI-Systeme in einem Edge-Cloud-Umfeld bietet eine Vielzahl an Vorteilen, insbesondere hinsichtlich niedriger Latenz, Datenschutz und transparenterer Entscheidungsfindung. Eine vielversprechende Technologie zur Verbesserung der Erklärbarkeit in diesem Kontext ist die Verwendung von Retrieval-Augmented-Generation (RAG)-Architekturen. Diese Architekturen ermöglichen es, relevante Informationen effizient aus einer lokalen Wissensdatenbank abzurufen und in den Entscheidungsprozess des Modells einzubeziehen, wodurch die Transparenz der KI-Systeme erhöht wird. Häufig sind die Entscheidungsprozesse der RAG-Modelle jedoch nur schwer nachzuvollziehen. Um diese Herausforderung zu adressieren, sollten zudem Methoden zur Interpretation von Modellen genutzt werden. Sie helfen dabei, die für die Entscheidungsfindung relevanten Teile des Eingabedatensatzes sowie den Entscheidungsweg des Modells verständlicher zu machen. Ein erklärbares KI-System sollte daher nicht nur die richtige Antwort liefern, sondern nach Möglichkeit auch darstellen können, wie diese Antwort zustande gekommen ist.

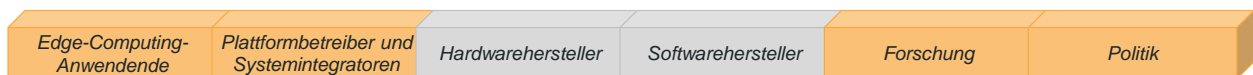
Darüber hinaus sollte die Erklärbarkeit nicht nur die lokale Verarbeitung betreffen, sondern auch die Cloud-Infrastruktur berücksichtigen. Hier ist es entscheidend, auch die Entscheidungen, die in der Cloud getroffen werden, nachvollziehbar zu gestalten. Durch die Implementierung von zusätzlichen Mechanismen zur Fehlerdiagnose und transparenten Nachvollziehbarkeit kann das Vertrauen in das System weiter gestärkt werden.

Notwendige Kompetenzen schaffen



Um die Bewertung und Optimierung von ECS effektiv durchzuführen, ist es entscheidend, die erforderlichen Kompetenzen auf Seiten der Integratoren und Anwender systematisch zu erfassen und weiterzuentwickeln. Die Entwicklung eines individualisierbaren Kompetenzkatalogs, der die spezifischen Fähigkeiten und Kenntnisse beschreibt, die für die Implementierung und Bewertung von ECS notwendig sind, kann eine einsetzbare Handlungshilfe für Unternehmen darstellen. So kann beispielsweise auf Basis des Kompetenzkatalogs gezielt in die Weiterbildung der Mitarbeitenden investiert werden. Ein Kompetenzkatalog im Bereich der Bewertung von ECS sollte beispielsweise Aspekte wie Edge-spezifische technische Kenntnisse, Verständnis von Netzwerktechnologien, Kenntnisse über Datenschutz und Datensouveränität sowie Erfahrungen mit KI-Methoden abdecken.

Einstiegshürden für Edge-Computing und dessen Bewertung verringern



Aufgrund ihrer hohen Komplexität stellt die Umsetzung und Bewertung von ECS weiterhin, insbesondere für KMU, eine umfassende Herausforderung dar. Anders als Cloud-Services können ECS nicht mit geringen Aufwänden bestellt oder gekündigt werden. Entsprechend gilt es, andere Wege zu finden, um die Einstiegshürden, insbesondere für kleinere und mittlere Unternehmen (KMU), weiter zu verringern. Hierzu ist insbesondere die Politik, aber auch die angewandte Forschung gefragt, niedrighschwellige Angebote zu entwickeln. Einige mögliche Maßnahmen werden folgend dargestellt.

Förderung von Proof-of-Concepts und Workshopangeboten zur Evaluation von ECS

Workshops und Proof-of-Concepts (PoCs) sind für Unternehmen im Bereich Edge- und Cloud-Deployments von entscheidender Bedeutung, da sie eine risikoarme Möglichkeit bieten, neue Technologien zu evaluieren und an die spezifischen Anforderungen der Fertigung anzupassen. Besonders mittelständische Unternehmen stehen oft vor der Herausforderung, mit begrenzten Ressourcen digitale Innovationen einzuführen und gleichzeitig wettbewerbsfähig zu bleiben. In Workshops können dafür Kompetenzen aufgebaut werden, welche den Unternehmen die gezielte Identifikation von Anwendungsfällen ermöglichen. Mit einem PoC kann die Machbarkeit und der Mehrwert neuer Technologien in einer kontrollierten Umgebung getestet werden, bevor größere Investitionen getätigt werden. Dies schafft Sicherheit, reduziert Umsetzungsrisiken und ermöglicht fundierte Entscheidungen für die digitale Transformation. Die Politik kann hier unterstützend wirken, indem sie solche PoCs verstärkt fördert. Ein aktuelles Beispiel ist das von der EU im Rahmen des IPCEI-CIS geförderte Projekt „Kicks for Edge“⁵.

⁵ <https://kicksforedge.siemens.cloud/>

Entwicklung eines wissenschaftlichen Reifegradmodells für ECS

Insbesondere im Kontext der Bewertung und Optimierung von Edge- und Cloud-Deployments in der Produktion stellt die Entwicklung von praxistauglichen Metriken und Reifegradanalysen eine zentrale Forschungsaufgabe dar. Zusätzlich zu den in der vorliegenden Orientierungshilfe vorgestellten Methoden zur Analyse und Bewertung fehlt ein umfassendes Reifegradmodell, das Unternehmen dabei unterstützt, ihren aktuellen Entwicklungsstand systematisch zu erfassen und gezielt zu verbessern.

Hier kann die Forschung durch die Integration von praxisnahen Messsystemen und standardisierten Bewertungsansätzen einen entscheidenden Beitrag leisten, um Unternehmen eine transparente Orientierungshilfe und konkrete Handlungsempfehlungen für den Übergang zu intelligenten, flexiblen Produktionssystemen zu geben. Damit diese Entwicklungen vorangetrieben werden können, ist es notwendig, dass die Politik entsprechende Förderprogramme aufsetzt, um Forschung und Innovation in diesem Bereich nachhaltig zu stärken.

Entwicklung eines Leitfadens zur anwendungsfallspezifischen Bewertung von ECS

Die Entscheidung zwischen Edge-, Cloud- oder hybriden Edge-Cloud-Lösungen stellt Unternehmen, insbesondere KMU, vor Herausforderungen. Das Ziel ist es, die spezifischen Anforderungen und Möglichkeiten verschiedener Anwendungsfälle so darzustellen, dass KMU auf Grundlage dieser eine eigenständige Bewertung und Selektion der Edge-, Cloud- oder hybriden Edge-Cloud-Lösungen vornehmen können. Die Unternehmen sollten in die Lage versetzt werden, spezifische Anwendungen eigenständig einer ersten fundierten Bewertung zu unterziehen, um auf dieser Grundlage geeignete Lösungen identifizieren zu können.

Neben positiven Effekten wie der Steigerung der Akzeptanz von Edge-Computing-Technologien und der Unterstützung einer reibungslosen Integration in bestehende Systeme liegt der Hauptzweck darin, Risiken und Probleme der jeweiligen Technologien zu verstehen. Insbesondere sollte der Einsatz ungeeigneter Technologien oder Architekturen für die jeweiligen Anwendungsfälle verhindert werden, die ineffiziente Prozesse, hohe Kosten oder eine mangelnde Skalierbarkeit verursachen und dadurch die langfristigen Ziele von Unternehmen gefährden könnten. Der Leitfaden würde somit als strategisches Werkzeug fungieren, um den Unternehmen eine eigenständige Bewertung der Lösungen auf Grundlage von vordefinierten Kriterien zu ermöglichen. Hierbei sollte auch berücksichtigt werden, welche Parameter überhaupt von einem Unternehmen, in Anbetracht von dessen Rolle im ECS beeinflusst werden können.

Dafür ist es zwingend notwendig, dass die Entwicklung eines Leitfadens (Bewertungsrahmen) die Unternehmen dabei unterstützt, unterschiedliche Anwendungsfälle zu gruppieren/kategorisieren und entsprechende spezifische Bewertungskriterien (z. B. Echtzeitfähigkeit, Datenvolumen und Sicherheitsanforderungen) in einem modularen Bewertungskriterienkatalog zu definieren. Es sollte demnach ein an das Praxisfeld angepasster Bewertungskatalog auf wissenschaftlicher Basis erstellt werden. Im gesamten Erstellungsprozess ist die Einbindung verschiedener Stakeholder wie Unternehmen, Plattformbetreiber, Hersteller und Forschung ein zentrales Element, um sowohl die Relevanz als auch die Qualität sicherzustellen.

6 Referenzen

- Ahvar, E., Orgerie, A.-C. & Lebre, A. (2022). Estimating Energy Consumption of Cloud, Fog, and Edge Computing Infrastructures. *IEEE Transactions on Sustainable Computing*, 7(2), 277–288. <https://doi.org/10.1109/TSUSC.2019.2905900>
- AIOTI. (2024). *IoT and Edge Computing Carbon Footprint Measurement Methodology*. <https://aioti.eu/wp-content/uploads/AIOTI-Carbon-Footprint-Methodology-Report-R3-Final.pdf>
- Aslanpour, M. S., Gill, S. S. & Toosi, A. N. (2020). Performance evaluation metrics for cloud, fog and edge computing: A review, taxonomy, benchmarks and standards for future research. *Internet of Things*, 12, 100273. <https://doi.org/10.1016/j.iot.2020.100273>
- Babel, M., Ehaus, M., Heeß, P., Körner, M.-F., Schick, L. & Strüker, J. (2025). Introducing the Trust Diamond for Energy Flexibility Provision: On the Tension of Data Verifiability and Privacy. Vorab-Onlinepublikation. <https://doi.org/109837>
- Babel, M., Willburger, L., Lautenschlager, J., Völter, F., Guggenberger, T., Körner, M.-F., Sedlmeir, J., Strüker, J. & Urbach, N. (2025). Self-sovereign identity and digital wallets. *Electronic Markets*, 35(1). <https://doi.org/10.1007/s12525-025-00772-0>
- Baudoin, C. R. (2018). The Impact of Data Residency on Cloud Computing. In L. Barolli (Hrsg.), *32nd IEEE International Conference on Advanced Information Networking and Applications workshops: IEEE WAINA 2018 : 16-18 May 2018, Kraków, Poland : proceedings* (S. 430–435). IEEE. <https://doi.org/10.1109/WAINA.2018.00124>
- Bergmann, R., Althoff, K.-D., Minor, M., Reichle, M. & Bach, K. (2009). *Case-Based Reasoning: Introduction and Recent Developments*. http://wi.cs.uni-frankfurt.de/webdav/publications/2009_KI_CBR.pdf
- Bergmann, R., Minor, M., Bach, K., Althoff, K.-D. & Muñoz-Avila, H. (2020). Fallbasiertes Schließen: Handbuch der Künstlichen Intelligenz. In G. Görz, U. Schmid & T. Braun (Hrsg.), *Handbuch der Künstlichen Intelligenz* (S. 343–394). De Gruyter. <https://doi.org/10.1515/9783110659948-009>
- Biden, J. R. (2021). *Improving the Nation's Cybersecurity*. <https://www.federalregister.gov/d/2021-10460>
- Brockhaus, B. (2024). *Modellierung von Edge-Computing-Clustern zur Ermittlung ressourcenezufizienter Softwareverteilungen in der Produktion*. TU Darmstadt. https://tuprints.ulb.tu-darmstadt.de/26739/1/2024_Dissertation_Benjamin_Brockhaus.pdf
- Brockhaus, B., Gast, F. & Weigold, M. (2024). Modeling of an Edge Computing Cluster for Optimized Distribution of Tasks in Production Environments. In T. Bauernhansl, A. Verl, M. Liewald & H.-C. Möhring (Hrsg.), *Lecture Notes in Production Engineering. Production at the Leading Edge of Technology* (S. 44–53). Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-47394-4_5
- Bundesamt für Sicherheit in der Informationstechnik. (2023). *Zero Trust*. <https://www.bsi.bund.de/DE/Themen/Unternehmen-und-Organisationen/Informationen-und-Empfehlungen/Zero-Trust/zero-trust.html>

- Cha, S.-C., Hsu, T.-Y., Xiang, Y. & Yeh, K.-H. (2019). Privacy Enhancing Technologies in the Internet of Things: Perspectives and Challenges. *IEEE Internet of Things Journal*, 6(2), 2159–2187. <https://doi.org/10.1109/JIOT.2018.2878658>
- Coutinho, E. F., Carvalho Sousa, F. R. de, Rego, P. A. L., Gomes, D. G. & Souza, J. N. de (2015). Elasticity in cloud computing: a survey. *annals of telecommunications - annales des télécommunications*, 70(7-8), 289–309. <https://doi.org/10.1007/s12243-014-0450-7>
- Credentials Community Group. (2021, 11. November). *A Primer for Decentralized Identifiers: An introduction to self-administered identifiers for curious people*. <https://w3c-ccg.github.io/did-primer/>
- Creutz, L. & Dartmann, G. (2023). Decentralized Policy Enforcement in Zero Trust Architectures. In *2023 IEEE Future Networks World Forum (FNWF)* (S. 1–6). IEEE. <https://doi.org/10.1109/FNWF58287.2023.10520563>
- Das, A., Patterson, S. & Wittie, M. (2018). EdgeBench: Benchmarking Edge Computing Platforms. In A. Sill & J. Spillner (Hrsg.), *11th IEEE/ACM International Conference on Utility and Cloud Computing companion - UCC-C 2018: 17-20 December 2018, Zurich, Switzerland : proceedings* (S. 175–180). IEEE. <https://doi.org/10.1109/UCC-Companion.2018.00053>
- DIN (Hrsg.). *Modell der R-Strategien: Circular Thinking in Standards*. <https://www.din.de/de/forschung-und-innovation/themen/circular-economy/normenrecherche/modell-der-r-strategien>
- Doney, P. M. & Cannon, J. P. (1997). An Examination of the Nature of Trust in Buyer–Seller Relationships. *Journal of Marketing*, 61(2), 35–51. <https://doi.org/10.1177/002224299706100203>
- Engel, T. (2023). GEMIMEG-II — How metrology can go digital. *Measurement Science and Technology*, 34(10), 104002. <https://doi.org/10.1088/1361-6501/ace468>
- European Commission. (2017). *New European interoperability framework: Promoting seamless services and data flows for European public administrations*. Publications Office. <https://doi.org/10.2799/78681>
- Fassnacht, M., Benz, C., Heinz, D., Leimstoll, J. & Satzger, G. (2023). *Barriers to Data Sharing among Private Sector Organizations*. <https://scholarspace.manoa.hawaii.edu/server/api/core/bitstreams/61e6bcec-d348-4b2d-b127-3c387dfba7d2/content>
- GEMIMEG-II. (2024). *Abschlussbericht*.
- Generaldirektion Kommunikation. (2024, 11. Januar). *Das Datengesetz tritt in Kraft: Was das für Sie bedeutet*. https://commission.europa.eu/news/data-act-enters-force-what-it-means-you-2024-01-11_de
- Gill, S. S., Chana, I., Singh, M. & Buyya, R. (2018). CHOPPER: an intelligent QoS-aware automatic resource management approach for cloud computing. *Cluster Computing*, 21(2), 1203–1241. <https://doi.org/10.1007/s10586-017-1040-z>
- Green, C. (1981). APPLICATION OF THEOREM PROVING TO PROBLEM SOLVING**This research is a part of Project Defender and was supported by the Advanced Research Projects Agency of the Department of Defense and was monitored by Rome Air Development Center under Contracts AF 30(602)-4147 and F30602-69-C-0056.††This preprint

- is a preliminary version and is subject to modification prior to publication. In *Readings in Artificial Intelligence* (S. 202–222). Elsevier. <https://doi.org/10.1016/B978-0-934613-03-3.50019-2>
- Gröger, J., Behrens, F., Liu, R. & Bunke, D. (2024). *Methodenskizze: Ökobilanz Digitaler Dienstleistungen*. <https://ecodigit.de/veroeffentlichungen>
- Guldner, A., Bender, R., Calero, C., Fernando, G. S., Funke, M., Gröger, J., Hilty, L. M., Hörschemeyer, J., Hoffmann, G.-D., Junger, D., Kennes, T., Kreten, S., Lago, P., Mai, F., Malavolta, I., Murach, J., Obergöker, K., Schmidt, B., Tarara, A., . . . Naumann, S. (2024). Development and evaluation of a reference measurement model for assessing the resource and energy efficiency of software products and components—Green Software Measurement Model (GSMM). *Future Generation Computer Systems*, 155, 402–418. <https://doi.org/10.1016/j.future.2024.01.033>
- Haslum, P. (2006). *Admissible heuristics for automated planning. Linköping studies in science and technology. Dissertation: Bd. 1004*. Department of Computer and Information Science, Linköpings universitet.
- Heeß, P., Holly, S., Körner, M.-F., Nieße, A., Radtke, M., Schick, L., Stark, S., Strüker, J. & Zwede, T. (2025). A multi-agent approach with verifiable and data-sovereign information flows for decentralizing redispatch in distributed energy systems. *Energy Informatics*, 8(1). <https://doi.org/10.1186/s42162-024-00464-7>
- Hegiste, V., Legler, T. & Ruskowski, M. (2022). Application of Federated Machine Learning in Manufacturing. In *2022 International Conference on Industry 4.0 Technology (I4Tech)* (S. 1–8). IEEE. <https://doi.org/10.1109/I4Tech55392.2022.9952385>
- Herlich, M., Pfeiffenberger, T., Du, J. & Dorfinger, P. (2017). *5G-MLab: Developing a Wireless Reliability Measurement Framework*. <https://www.salzburgresearch.at/wp-content/uploads/2017/11/Herlich-5G-MLab-Paper.pdf>
- Horsch, J. (2023). Entscheidungsorientierte Kostenrechnung. In J. Horsch (Hrsg.), *Kostenrechnung* (S. 211–257). Springer Fachmedien Wiesbaden. https://doi.org/10.1007/978-3-658-39545-2_4
- Hu, Z., Chen, S., Rao, H., Hong, C., Huang, O., Xu, X. & Jia, G. (2024). An adaptive service deployment algorithm for cloud-edge collaborative system based on speedup weights. *The Journal of Supercomputing*, 80(16), 23177–23204. <https://doi.org/10.1007/s11227-024-06339-8>
- Hüning, F. (2019). *Embedded Systems für IoT*. Springer Berlin Heidelberg. <https://doi.org/10.1007/978-3-662-57901-5>
- IONOS Redaktion. (2022). *Was ist Jitter? Definition und vorbeugende Maßnahmen*. <https://www.ionos.de/digitalguide/server/knowhow/jitter/>
- ISO. (2015). *ISO/IEC 2382:2015: Information technology — Vocabulary*. <https://www.iso.org/standard/63598.html>
- Karagiannis, G., Campbell, N., Robles, M. I., Wetterwald, M., Voutyras, O., Kung, A., Frost, L., Suci, G., Gyrard, A., Gayko, J., Rennoch, A., Zimmermann, E. C., Carugi, M., Suo, A., Zhang, X., Pitwon, R., Perey, C., Lingbo, K., Bin, S., . . . Kopertowski, Z. (2025). *Report*

- of TWG IoT & Edge: Landscape of Edge Computing Standards | V.2. <https://zenodo.org/records/14804841> <https://doi.org/10.24406/PUBLICA-4537>
- Kilic, Ü. & Engelhardt, S. von. (2025). *EDGE-FINOPS: Ein Leitfaden für das Kostenmanagement von Edge-Cloud-Systemen*. https://www.digitale-technologien.de/DT/Redaktion/DE/Downloads/Publikation/EDGE-Datenwirtschaft/20253101_Leitfaden_Edge.pdf?__blob=publicationFile&v=7
- Kühnapfel, J. B. (2021). *Scoring und Nutzwertanalysen*. Springer Fachmedien Wiesbaden. <https://doi.org/10.1007/978-3-658-34810-6>
- Lenhard, J., Harrer, S. & Wirtz, G. (2013). Measuring the Installability of Service Orchestrations Using the Square Method. In *IEEE 6th International Conference on Service-Oriented Computing and Applications: 16-18 December 2013, Kauai, Hawaii, USA* (S. 118–125). Institute of Electrical and Electronics Engineers. <https://doi.org/10.1109/SOCA.2013.30>
- Li, J., Liang, W., Xu, W., Xu, Z., Jia, X., Zhou, W. & Zhao, J. (2022). Maximizing User Service Satisfaction for Delay-Sensitive IoT Applications in Edge Computing. *IEEE Transactions on Parallel and Distributed Systems*, 33(5), 1199–1212. <https://doi.org/10.1109/TPDS.2021.3107137>
- Linux Foundation (Hrsg.). (2020). *Sharpening the Edge: Overview of the LF Edge Taxonomy and Framework*. https://lfedge.org/wp-content/uploads/sites/24/2020/07/LFedge_Whitepaper.pdf
- López, M. Á. & Razauskas, A. (2024). *Der Europäische Wirtschaftsraum, die Schweiz und der Norden*. <https://www.europarl.europa.eu/factsheets/de/sheet/169/der-europaische-wirtschaftsraum-ewr-die-schweiz-und-der-norden>
- Luber, S. & Donner, A. (2018, 1. August). *Was ist Jitter?* <https://www.ip-insider.de/was-ist-jitter-a-651837/>
- Maciel, P., Dantas, J., Melo, C., Pereira, P., Oliveira, F., Araujo, J. & Matos, R. (2022). A survey on reliability and availability modeling of edge, fog, and cloud computing. *Journal of Reliable Intelligent Environments*, 8(3), 227–245. <https://doi.org/10.1007/s40860-021-00154-1>
- Malburg, L., Schultheis, A. & Bergmann, R. (2023). *Modeling and Using Complex IoT Time Series Data in Case-Based Reasoning: From Application Scenarios to Implementations*. https://ceur-ws.org/Vol-3438/paper_07.pdf
- Marrella, A. (2018). What Automated Planning Can Do for Business Process Management. In E. Teniente & M. Weidlich (Hrsg.), *Lecture Notes in Business Information Processing: Bd. 308. Business Process Management Workshops* (Bd. 308, S. 7–19). Springer International Publishing. https://doi.org/10.1007/978-3-319-74030-0_1
- McMahan, H. B., Moore, E., Ramage, D., Hampson, S. & Arcas, B. A. y. (2016). Communication-Efficient Learning of Deep Networks from Decentralized Data. Vorab-Onlinepublikation. <https://doi.org/10.48550/ARXIV.1602.05629>
- Mo, W., Wang, T., Zhang, S. & Zhang, J. (2020). An active and verifiable trust evaluation approach for edge computing. *Journal of Cloud Computing*, 9(1). <https://doi.org/10.1186/s13677-020-00202-w>

- Mondello, E. (2022). *Corporate Finance*. Springer Fachmedien Wiesbaden. <https://doi.org/10.1007/978-3-658-34408-5>
- Ocean Protocol Foundation Ltd. (Hrsg.). (2024). *Compute to data: Compute to data version 2 (C2dv2)*. <https://docs.oceanprotocol.com/developers/compute-to-data>
- Omar, M., Challal, Y. & Bouabdallah, A. (2012). Certification-based trust models in mobile ad hoc networks: A survey and taxonomy. *Journal of Network and Computer Applications*, 35(1), 268–286. <https://doi.org/10.1016/j.jnca.2011.08.008>
- Rose, S., Borchert, O., Mitchell, S. & Connelly, S. (2020). *Zero Trust Architecture*. <https://doi.org/10.6028/NIST.SP.800-207>
- Rouf, A., Arslan, E. & Charyyev, B. (2024). Latency and Bandwidth Benefits of Edge Computing for Scientific Applications. In *IEEE INFOCOM 2024 - IEEE Conference on Computer Communications workshops (INFOCOM WKSHPS): 20-23 May 2024* (S. 1–6). IEEE. <https://doi.org/10.1109/INFOCOMWKSHPS61880.2024.10620811>
- Rousseau, D. M., Sitkin, S. B., Burt, R. S. & Camerer, C. (1998). Not So Different After All: A Cross-Discipline View Of Trust. *Academy of Management Review*, 23(3), 393–404. <https://doi.org/10.5465/amr.1998.926617>
- Scarselli, F., Gori, M., Tsoi, A. C., Hagenbuchner, M. & Monfardini, G. (2009). The graph neural network model. *IEEE transactions on neural networks*, 20(1), 61–80. <https://doi.org/10.1109/TNN.2008.2005605>
- Scherenberg, F. von, Hellmeier, M. & Otto, B. (2024). Data Sovereignty in Information Systems. *Electronic Markets*, 34(1). <https://doi.org/10.1007/s12525-024-00693-4>
- Schultheis, A. (2024). *Exploring a Hybrid Case-Based Reasoning Approach for Time Series Adaptation in Predictive Maintenance*. https://ceur-ws.org/Vol-3708/paper_19.pdf
- Schultheis, A., Alt, B., Bast, S., Guldner, A., Jilg, D., Katic, D., Mundorf, J., Schlagenhauf, T., Weber, S., Bergmann, R., Bergweiler, S., Creutz, L., Dartmann, G., Malburg, L., Naumann, S., Rezapour, M. & Ruskowski, M. (2024). EASY: Energy-Efficient Analysis and Control Processes in the Dynamic Edge-Cloud Continuum for Industrial Manufacturing. *KI - Künstliche Intelligenz*. Vorab-Onlinepublikation. <https://doi.org/10.1007/s13218-024-00868-3>
- Schultheis, A., Jilg, D., Malburg, L., Bergweiler, S. & Bergmann, R. (2024). Towards Flexible Control of Production Processes: A Requirements Analysis for Adaptive Workflow Management and Evaluation of Suitable Process Modeling Languages. *Processes*, 12(12), 2714. <https://doi.org/10.3390/pr12122714>
- Schultheis, A., Malburg, L., Grüger, J., Weich, J., Bertrand, Y., Bergmann, R. & Serral Asensio, E. (2024). Identifying Missing Sensor Values in IoT Time Series Data: A Weight-Based Extension of Similarity Measures for Smart Manufacturing. In J. A. Recio-Garcia, M. G. Orozco-del-Castillo & D. Bridge (Hrsg.), *Lecture Notes in Computer Science: Bd. 14775. Case-Based Reasoning Research and Development* (Bd. 14775, S. 240–257). Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-63646-2_16
- Schütte, J., Brost, G. & Wessel, S. (2018). *Der Trusted Connector im Industrial Data Space*. <https://doi.org/10.48550/ARXIV.1804.09442>

- Sheybani, N., Ahmed, A., Kinsy, M. & Koushanfar, F. (2025, 10. Februar). *Zero-Knowledge Proof Frameworks: A Systematic Survey*. <http://arxiv.org/pdf/2502.07063v3>
- Slattery, T. (2019). *Netzwerküberlastung: Ursachen und Tools zur Abhilfe*. <https://www.computerweekly.com/de/ratgeber/Netzwerkueberlastung-Ursachen-und-Tools-zur-Abhilfe>
- SMART DATA Forum. *Datensouveränität*. https://www.digitale-technologien.de/DT/Redaktion/DE/Downloads/Smart-Data-Forum/wissen-datensouveraenitaet.pdf?__blob=publicationFile&v=2
- Souza, P., Ferreto, T. & Calheiros, R. (2024). Maintenance Operations on Cloud, Edge, and IoT Environments: Taxonomy, Survey, and Research Challenges. *ACM Computing Surveys*, 56(10), 1–38. <https://doi.org/10.1145/3659097>
- Sporny, M., Longley, D., Chadwick, D., Herman, I. & (Keine Angabe). (2025, 20. März). *Verifiable Credentials Data Model v2.0*. <https://www.w3.org/TR/vc-data-model-2.0/>
- Stadelbacher, S. (2020). Privatheit im Wandel: Aktuelle Entwicklungs- und Diskussionslinien. In S. Stadelbacher (Hrsg.), *Soziologie des Privaten in Zeiten fortgeschrittener Modernisierung* (S. 151–294). Springer Fachmedien Wiesbaden. https://doi.org/10.1007/978-3-658-32015-7_3
- Straub, S. & Bogenstahl, C. (2024). *DIE REGULIERUNG VON KÜNSTLICHER INTELLIGENZ UND DATENWIRTSCHAFT: AUSWIRKUNGEN DER AKTUELLEN EU-GESETZGEBUNG AUF FORSCHUNGS- UND ENTWICKLUNGSPROJEKTE – EINE ÜBERSICHT*. https://www.iit-berlin.de/wp-content/uploads/2024/12/KI_Leitfaden_EU_Recht-2024.pdf
- Strnadl, C. F. (2021). End-to-End-Architekturen zur Datenmonetarisierung im Industrial Internet of Things (IIoT). In D. Trauth, T. Bergs & W. Prinz (Hrsg.), *Monetarisierung von technischen Daten: Innovationen aus Industrie und Forschung* (S. 169–206). Springer Vieweg. https://doi.org/10.1007/978-3-662-62915-4_10
- Suárez-Varela, J., Almasan, P., Ferriol-Galmés, M., Rusek, K., Geyer, F., Cheng, X., Shi, X., Xiao, S., Scarselli, F., Cabellos-Aparicio, A. & Barlet-Ros, P. (2023). Graph Neural Networks for Communication Networks: Context, Use Cases and Opportunities. *IEEE Network*, 37(3), 146–153. <https://doi.org/10.1109/MNET.123.2100773>
- Verein Deutscher Werkzeugmaschinenfabriken e.V. (Hrsg.). (2020). *Marktbericht 2019: Die deutsche Werkzeugmaschinenindustrie und ihre Stellung im Weltmarkt*. www.vdw.de
- Wang, Z., Zhou, Y., Jin, X., Chen, Y. & Lu, C. (2024). An edge server deployment approach for delay reduction and reliability enhancement in the industrial internet. *Wireless Networks*, 30(6), 5743–5757. <https://doi.org/10.1007/s11276-023-03339-z>
- Willner, A. & Gowtham, V. (2020). Toward a Reference Architecture Model for Industrial Edge Computing. *IEEE Communications Standards Magazine*, 4(4), 42–48. <https://doi.org/10.1109/MCOMSTD.001.2000007>
- Zhu, Z., Peng, J., Gu, X., Li, H., Liu, K., Zhou, Z. & Liu, W. (2018). Fair Resource Allocation for System Throughput Maximization in Mobile Edge Computing. *IEEE Access*, 6, 5332–5340. <https://doi.org/10.1109/ACCESS.2018.2790963>