

Mit Glasfaser, AI-IX und Edge in die nächste digitale Epoche

Performanz und Resilienz der digitalen Infrastrukturen für die KI-Ära

Die Verlagerung digitaler Anwendungen in Richtung KI verändert die Anforderungen an Netze und Rechenzentren grundlegend. DE-CIX und GasLINE zeigen, wie Glasfaserinfrastrukturen, Internet Exchanges und latenzoptimierte Verbindungen den reibungslosen Betrieb von KI-Systemen sichern. Die Vision: KI-Inferenz am Edge, europaweit vernetzt, mit maximaler Performanz und minimaler Verzögerung – von Anwendungen für Smart City bis Robotik. Von Ivo Ivanov, DE-CIX, und Wolfram Rinner, GasLINE

„Glasfaser ist das Medium der Zukunft“, das war schon vor vielen Jahren der weitsichtige Leitsatz von GasLINE als Infrastrukturanbieter. Diese Bedeutung der Glasfaser, als das Übertragungsmedium schlechthin, ist zehn Jahre später nach wie vor gültig. „Wir brauchen mehr Glasfaser“ – ein klares Statement von DE-CIX. Glasfasernetze sind essenziell für höchste Bandbreiten, geringe Latenzzeiten und die internationale Vernetzung der eigenen IX-Infrastruktur. Die nächste Generation ist die „AI-readiness“. Glasfasernetze sind immer noch der Enabler für den Fortschritt der Digitalisierung. Künstliche Intelligenz (KI) ist kein temporärer Trend, sie durchzieht die Wirt-

schaft und spielt schon jetzt bei den Menschen auch privat als zukunftsrelevante Anwendung eine immer größer werdende Rolle. Der Beitrag in dieser Cable!vision-Ausgabe thematisiert das Schwerpunktthema Resilienz in Bezug auf die neuen Voraussetzungen für Internet Exchanges, Rechenzentren und Netzinfrastrukturen. Verfügbarkeit und Leistungsfähigkeit der digitalen Infrastrukturen sind substantiell für die KI-Ära. Jeder Chat, der KI anzapft, jedes Prognosemodell oder jeder angesprochene Agent geht mit einer sich entwickelnden digitalen Infrastruktur im Hintergrund einher. Kaum eine andere Technologie hat in so kurzer Zeit einen solchen Durchbruch – und das international – erreicht. Die

Entwicklungen und Forschungen werden weitergehen, Unternehmen setzen auf Künstliche Intelligenz und ChatGPT ist ein Informationstool und eine Art Kontrollinstanz als Sparringspartner geworden. Generative KI hat sich von einer spannenden neuen Anwendung, die zum Experimentieren motivierte, zu einer Technologie entwickelt, die in vielen Branchen genutzt und weiterentwickelt wird. Dazu gehören u. a. das bislang in puncto Digitalisierung eher konservative Gesundheitswesen, die Branchen Fertigung und Handel, sowie Lehrinstitute und die öffentliche Hand. Die digitalen Anwendungen der KI-Ära stellen infrastruktureseitig immer höhere Ansprüche an die Netze und Internet



Abbildung 1: Je nach Anforderung dürfen Daten nicht weiter als 80 Kilometer von einer digitalen Anwendung entfernt ausgetauscht werden

Exchanges. Das ist die Herausforderung für viele Akteure, die eine adäquate digitale Infrastruktur auf- und ausbauen. Es geht zudem um die Resilienz mit ausreichend Kapazität, Redundanz und Robustheit der digitalen Infrastrukturen, um auch die zukünftigen KI-Anwendungen in passender Qualität und Skalierbarkeit zu unterstützen. Sicherheit muss prioritär gewährleistet sein, denn auch personenbezogene Daten werden dabei eingesetzt. Das betrifft insbesondere das Gesundheitswesen und die öffentliche Verwaltung.

Neue KI-Anwendungsbereiche bei Kommunen

Um urbane Herausforderungen managen zu können, wurde vom Unternehmen URBAN.KI, einer deutschen Initiative speziell für Kommunen, ein bundesweiter Ansatz verfolgt. Bis Ende 2025 werden die KI-basierten Produkte 11.000 Städten und Gemeinden zur Verfügung gestellt (Quelle: <https://urban-ki.de/neun-use-cases-zur-zukunft-smarter-staedte-durch-kuenstliche/>).

Zentrale Bereiche wie Stadtentwicklung, Smart City und in dem Kontext speziell Mobilitätssteuerung, Umweltplanung, Ver- und Entsorgung und Verwaltungsprozesse stehen im Fokus.

Das Deutsche Forschungszentrum für Künstliche Intelligenz, Fraunhofer FOKUS und weitere Partner waren an der Entwicklung der Anwendungsfälle in diesem Projekt involviert. Neun Anwendungsgebiete wurden ausgewählt: KI für Stadtplanung mit datenbasierten Infrastrukturen, ein Projekt für eine universelle KI-Engine zur Luftbilddauswertung, KI-Mobilitätsplanung und -steuerung für den On-demand-ÖPNV-Verkehr oder Smart Energie für Umweltplanung. Die Stadt Solingen will der Anzahl an Cyberangriffen auf Kommunen mit einer modularen KI-Sicherheitsplattform entgegenzutreten. Um diese Anwendungen auszuspielen, braucht jede Kommune Netze und auch Rechenzentren in den Regionen, in denen die KI-Anwendungen nah am Nutzer verarbeitet werden.

Latenzzeit ist ein Indikator für die KI-Qualität

„Latency is the new currency“ ist eine Kernaussage, die DE-CIX mit Blick auf

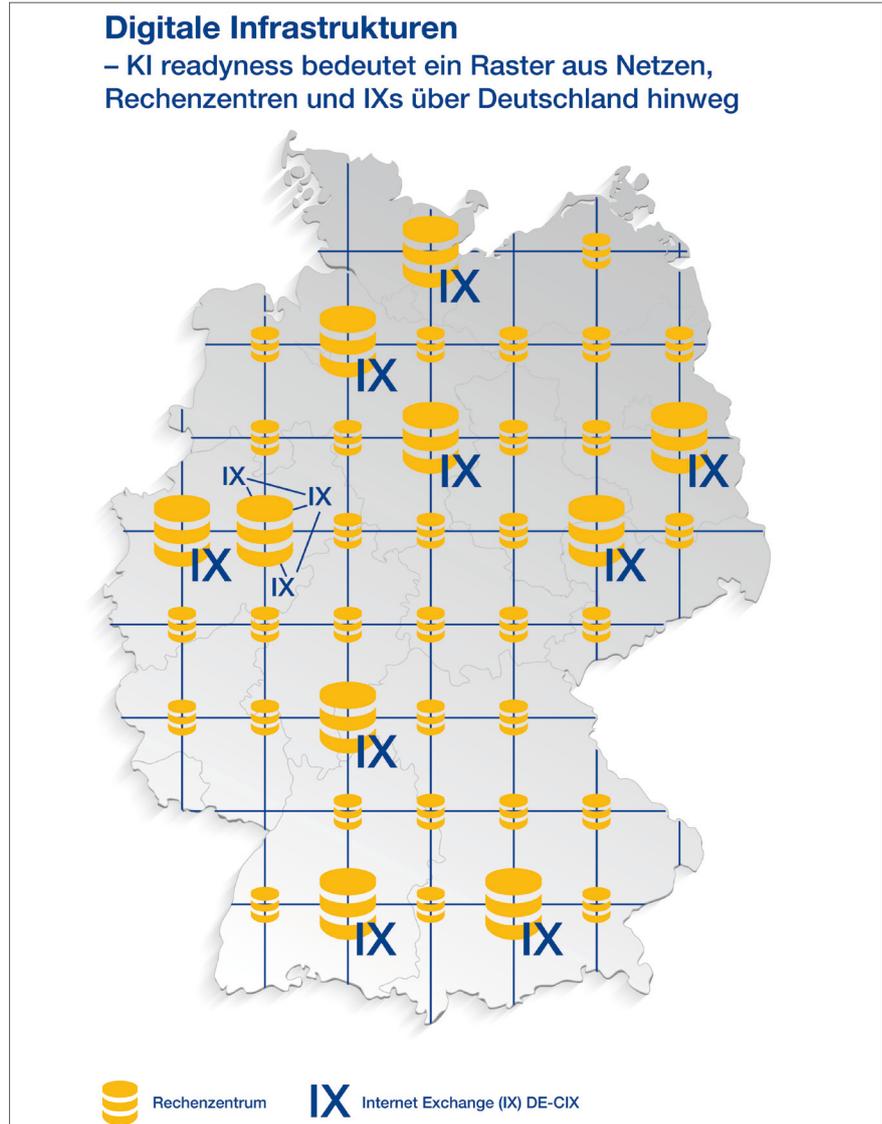


Abbildung 2: Eine rasterartige digitale Infrastruktur ist die Voraussetzung für geringe Latenzzeiten

die rasante Entwicklung der Verkürzung der Latenzzeiten geprägt hat. Latenz, also die Verzögerung bei der Datenübertragung, hängt direkt mit der Lichtgeschwindigkeit zusammen – denn Daten bewegen sich in Glasfaserkabeln nahezu mit Lichtgeschwindigkeit. Internetknoten (IX) sorgen für niedrige Latenzzeiten, da sie einen direkten und effizienten Datenaustausch zwischen Netzwerken ermöglichen – von Content-Providern bis hin zu Endnutzern. Dabei wird deutlich, dass die geografische Struktur der IXs für die Verkürzung von Latenzzeiten immens wichtig ist. Dies ist der Treiber für DE-CIX, die Anzahl an eigenen Standorten immer weiter zu erhöhen. Die Expansion umfasste Deutschland mit den Metropolen und geht international mit vielen Ländern

weiter. Aktuell betreibt das Unternehmen 60 Standorte in verschiedenen Modellen rund um den Globus.

Digitale Anwendungen und Inhalte müssen so nah wie möglich am Nutzer verarbeitet werden. So benötigt die Produktion in der Industrie 4.0 Latenzzeiten im einstelligen Millisekunden-Bereich. Das bedeutet für die Netze und die Verteilung der Zugangspunkte, dass Daten in einem Radius von 80 Kilometern verarbeitet werden müssen (siehe Abbildung 1).

Über Deutschland hinweg ist eine rasterartige Infrastruktur aus Glasfaser und ergänzend anderen Netztypen, Rechenzentren (von Hyperscalern bis zum Edge-RZ) und Internet Exchanges weiter aufzubauen, um solche geringen Latenzzeiten möglich zu machen (Abbildung 2).

Diese visualisierte, rasterartige Verteilung der digitalen Infrastruktur am Beispiel Deutschlands wird in der Zukunft entstehen müssen. Bestehende Infrastrukturen werden dabei genutzt und neue ausgebaut. Carrier, Internet Service Provider und Internet Exchange-Betreiber arbeiten gemeinsam daran. Es geht um das „Neue Internet“.

Auch die Wirtschaft ist herausgefordert: Digitale Werte, sogenannte „Digital Assets“, gilt es für neue Umsatzströme zu erschaffen. Die dabei entstehenden Datenströme brauchen verlässliche, ausfallsichere Netze, da eine Downtime, vor allem bei zeitsensitiven KI-Anwendungen, zu vermeiden ist. Dieser Anspruch umfasst ebenso dezentrale KI-Anwendungen und einen länderübergreifenden KI-Einsatz.

Der seit Langem etablierte Messfaktor für Netzwerk-Performance – die Round Trip Delay (RTD) – spielt eine zentrale Rolle für KI-Anwendungen. Besonders für sogenannte „Agentic AI“, also autonome KI-Systeme, die ohne menschliches Eingreifen Entscheidungen treffen und Handlungen initiieren, ist eine extrem geringe RTD unerlässlich. Diese Systeme benötigen eine hochgradig reaktionsschnelle Infrastruktur für Analyse, Steuerung und andere Echtzeitanforderungen.

Auch beim Training von Generativer KI (GenAI) hat die Latenz entscheidenden Einfluss – insbesondere beim Datentransport zwischen verschiedenen Rechenzentren oder den Servern, auf

denen Trainingsdaten gespeichert sind. KI-Modelle und -Agenten werden in Clustern betrieben, die untereinander vernetzt sind.

Die Ergebnisse von Inferenzprozessen müssen mit minimaler Latenz – idealerweise im einstelligen Millisekunden-Bereich – am Zielsystem ankommen, um das reibungslose Zusammenspiel KI-gestützter Geräte, Dienste und Anwendungen wie autonomes Fahren, eHealth oder humanoide Robotik zu gewährleisten.

Humanoide Roboter werden künftig über hochentwickelte, wenn auch künstlich erzeugte Intelligenz verfügen. Um die Anforderungen an Echtzeitkommunikation besser zu verstehen, lassen sich diese mit der menschlichen Wahrnehmung vergleichen, insbesondere mit der Verarbeitung sensorischer Informationen. Das menschliche Gehirn reagiert auf haptische Reize – wie Berührungen oder Temperaturveränderungen – im Schnitt innerhalb von 18 Millisekunden. Visuelle Eindrücke benötigen bis zu 12 Millisekunden für eine Reaktion, während auditive Reize besonders schnell verarbeitet werden: bereits nach etwa 1 Millisekunde.

Wenn Menschen künftig in immersiven Umgebungen mit humanoiden Robotern oder KI-Systemen zusammenarbeiten sollen, dann muss sich die digitale Infrastruktur an diesen biologischen Reaktionszeiten orientieren. Die geringstmögliche Latenz wird zum entscheidenden Maßstab, um eine authentische und reibungslose Interaktion sicherzustellen.

AI-IX – Die nächste Generation

Internet Exchanges spielen eine entscheidende Rolle, um die Latenz für AI-Workloads zu optimieren. Das DE-CIX-Konzept zu „AI-IX“ verbindet KI-Modelle mit den Geräten und Diensten, die sie antreiben. Es ist an die Voraussetzungen für Multi-Modell- und Multi-Agent-Inferencing angelehnt.

Abbildung 3 visualisiert dieses „Digitale Dreieck für KI-Inferenz“.

Das Herzstück des „AI-IX“ bildet eine leistungsstarke, virtuelle Multi-AI- und Cloud-Routing-Technologie, die den dynamischen und latenzarmen Austausch zwischen unterschiedlichen KI-Modellen und Cloud-Umgebungen ermöglicht.

Ziel des „AI-IX“ ist es, ein reibungsloses Zusammenspiel mehrerer KI-Modelle untereinander sowie mit Endgeräten sicherzustellen – in Echtzeit und mit geringstmöglicher Latenz. Die zugrunde liegende Plattform ist hochskalierbar und verteilt sich über mehrere Rechenzentrumsstandorte und Cloud-Anbieter innerhalb einer Metropolregion.

Um das volle Potenzial des „AI-IX“ auszuschöpfen, bedarf es einer hohen Dichte netzwerkrelevanter Akteure – darunter Anbieter von KI-Infrastrukturen, -Plattformen und -Modellen (AI-as-a-Service) sowie Content-Provider. Netzwerkelemente sind ein essenzieller Bestandteil dieses Konzepts: Netzbetreiber, insbesondere im Bereich Content Delivery Networks (CDN), sowie Infrastrukturanbieter wie GasLINE mit ihren leistungsfähigen Weitverkehrsnetzen, und natürlich Internet Service Provider (ISPs), die die notwendige Netzkapazität bereitstellen.

Viele Telekommunikationsunternehmen nutzen bereits die Infrastruktur der GasLINE, um DE-CIX-Standorte in ganz Deutschland zu erreichen – sei es in einzelnen Regionen oder standortübergreifend. GasLINE stellt damit das Rückgrat für die Verteilung von Datenströmen und AI-Workloads bis an die Netzgrenzen („Edge“) dar.

Der „AI-IX“ wird natürlich auch von global agierenden Unternehmen genutzt. Die Konnektivität und die Netzkapazitäten gehen geografisch innerhalb Deutschlands über mehrere Regionen hinweg und bis in andere Länder. Im Idealfall wird diese Infrastruktur durch einen globalen Backbone ergänzt,



Abbildung 3: Das Dreieck der KI-Inferenz verbindet die Übertragungstechnologien mit den KI-Modellen und den vernetzten Geräten

der die Interkonnektivität mehrerer IX-Standorte weltweit sicherstellt.

Die Zukunft der KI ist untrennbar mit der Zukunft der Vernetzung verbunden. Alle Stakeholder, die mit digitaler Infrastruktur zu tun haben, müssen sehr viel leisten und dazu beitragen. Die Inferenz, als Prozess der Extraktion von KI-generierten Ergebnissen aus KI-Modellen, wird immer mehr am Edge stattfinden, was zur Folge hat, dass die Rechenleistungen, KI-Anwendungen und die verwendeten Daten nah am Benutzer bleiben müssen. Ohne eine weit verteilte Netzinfrastruktur in der Fläche ist das nicht mit kurzen Latenzen möglich. Direkte hochperformante Netzverbindungen zwischen Rechenzentren, Cloud-Providern und den Computing-Einheiten bis zum Edge sind zu planen und zu betreiben.

Von „Glas“ hängt sehr viel ab, auch die Mobilfunkantennen-Standorte sind über Glasfaser erschlossen. Für Satelliten-Betreiber gilt das ebenfalls, beispielsweise sind die Bodenstationen von Starlink in Seattle über Glasfaser an den DE-CIX angeschlossen. Für den internationalen Einsatz von KI braucht es europäische und internationale Partnerschaften, damit ein einheitlicher, hochentwickelter Dienstraum entstehen kann. Eine

Ziellinie, die wir beim DE-CIX für das Zusammenwirken der Stakeholder sehen. Der Auftrag, den wir, DE-CIX, GasLINE und die anderen Infrastrukturbetreiber, haben und erfüllen, lautet digitale Infrastruktur zu entwickeln – für die Wirtschaft und im privatwirtschaftlichen Bereich. Das ist kein Selbstzweck! Die Aufgabe schafft Wirtschaftswachstum,

fördert Wohlstand und stellt der Bevölkerung gleichberechtigt einen Internetanschluss für Ausbildung und andere Zwecke zur Verfügung.

KI ist allerdings ein echter Beschleuniger. Die Geschwindigkeit, mit der die Anforderungen der KI an die digitalen Infrastrukturen wachsen und eine Anpassung erfordern, ist bisher beispiellos. ■



© DE-CIX

Ivo Ivanov,
CEO von DE-CIX
und Vorstandsvorsitzender der DE-CIX Group AG

www.de-cix.net



© GasLINE

Wolfram Rinner,
Geschäftsführer der GasLINE GmbH & Co. KG

www.gasline.de

Schön.

Aber ein Stück Heimat fehlt.

Retten Sie Geschichte. Spenden Sie Zukunft.
www.denkmalschutz.de



**DEUTSCHE STIFTUNG
DENKMALSCHUTZ**

Wir bauen auf Kultur.

Was wären unsere Städte und Dörfer ohne historische Gebäude? Ohne Bauwerke, die Geschichten erzählen, die typischen Eigenheiten einer Region verkörpern oder Wahrzeichen eines Ortes sind? Historische Bauwerke machen unsere Städte und Dörfer einmalig und unverwechselbar. Deshalb setzt sich die Deutsche Stiftung Denkmalschutz für den Erhalt einzigartiger Denkmale ein. Mit Ihrer Hilfe.

www.denkmalschutz.de



Spendenkonto

Commerzbank AG
BIC: COBA DE FF XXX
IBAN: DE71 500 400 500 400 500 400