BLUEPRINT

HOW TO USE DATA SPACES FOR AI

DIO

AUTOR:INNEN:

Günther Tschabuschnig

Cover: © designed by Freepik
Diese Broschüre wurde erstellt mit CanvaPro
Die Bilder wurden KI-generiert.

Data Intelligence Offensive
Hintere Zollamtstraße 17 / 3.OG
1030 Wien

ÜBER DATA INTELLIGENCE OFFENSIVE (DIO)

DIO fördert die österreichische Datenwirtschaft, um eine internationale Vorreiterrolle bei der intelligenten Nutzung von Daten einzunehmen. DIO schafft mit dezentralen Data Spaces ein sicheres Ökosystem, um datengestützte Innovationen, nachhaltige Wertschöpfung und Wohlstand zu fördern.

Österreich ist ein attraktiver Wirtschaftsstandort, der durch Data Sharing und Data Spaces die Wertschöpfung, die Innovation und den Wohlstand erhalten kann. Österreich verfügt über einen perfekt funktionierenden Datenmarkt, in dem alle Stakeholder:innen ohne Hürden miteinander lückenlos entlang der gesamten Wertschöpfungskette vernetzt sind, kommunizieren und vertrauen.

Österreich stärkt seine bestehenden Unternehmen und fördert neue Geschäftsmodelle durch ein verfügbares und funktionierendes Datensystem, das auf gesetzlichen Regulierungen und transparenten Verträgen basiert.

Data Spaces und Use Cases helfen dabei, Herausforderungen im Bereich Daten konkret und domänenspezifisch zu betrachten.

Data Spaces fokussieren sich auf übergeordnete Domänen (Wirtschaftsbereiche, Industriesektoren oder sonstige fachliche Anwendungsfelder), mit einer dezentralen Dateninfrastruktur, auf der Use Cases aufbauen können.

Das vorliegende Dokument ist eine Empfehlung zur Umsetzung des Data Acts.

Autor dieses Blueprints:
Günther Tschabuschnig, DIO

DIE RECHTLICHE EBENE DES DATENAUSTAUSCHS

EINLEITUNG ZUM DATA ACT

Der **Data Act** ist ein wichtiger Bestandteil der europäischen Gesetzgebung, der darauf abzielt, die Nutzung und den Zugang zu Daten in der Europäischen Union zu regulieren und eine **faire Datenwirtschaft** zu fördern. Das Gesetz, das im Jahr 2022 von der Europäischen Kommission vorgeschlagen wurde, ergänzt den bereits bestehenden Data Governance Act und ist Teil der umfassenden europäischen Datenstrategie, die den Umgang mit Daten in einer zunehmend digitalisierten Gesellschaft verbessern soll. Dabei verfolgt der Data Act mehrere Hauptziele: die Förderung von Innovation und neuen Geschäftsmodellen, die Stärkung der Rechte von Verbrauchern und Unternehmen in Bezug auf die Nutzung von Daten sowie die Gewährleistung eines fairen und transparenten Zugangs zu Daten.

FAIRER ZUGANG ZU DATEN

Ein zentrales Anliegen des Data Act ist die Schaffung eines **fairen Zugangs zu Daten**, die von vernetzten Geräten und Produkten generiert werden. Dazu gehören beispielsweise Daten aus intelligenten Haushaltsgeräten, Fahrzeugen, Maschinen und anderen Geräten des Internets der Dinge (IoT). Die Idee ist, dass sowohl Verbraucher:innen als auch Unternehmen **mehr Kontrolle** über die von ihnen erzeugten Daten erhalten und nicht ausschließlich von den Hersteller.innen der Geräte abhängig sind. Dadurch soll eine Innovationsförderung angeregt werden, indem Dritte auf Grundlage dieser Daten neue Dienstleistungen entwickeln können, wie zum Beispiel Überwachungs-, Wartungs- oder Optimierungsservices.

FÖRDERUNG DER DATENWIRTSCHAFT

Ein weiteres Ziel des Data Act ist die **Förderung der Datenwirtschaft**. Die Europäische Union erkennt die Bedeutung von Daten als wesentlichen Faktor für wirtschaftliche Innovation an. Durch die Schaffung eines klaren und einheitlichen rechtlichen Rahmens soll die Bereitschaft von Unternehmen gefördert werden, Daten miteinander zu teilen. Die Regeln des Data Act sollen dabei den Datenaustausch zwischen verschiedenen Akteuren erleichtern und gleichzeitig den Schutz der betroffenen Daten sicherstellen. Insbesondere kleine und mittelständische Unternehmen (KMU) sollen davon profitieren, indem ihnen der Zugang zu Daten erleichtert wird, wodurch sie **Wettbewerbsvorteile** erzielen und **innovative Geschäftsmodelle** entwickeln können.

Siehe dazu auch den Blueprint “How to Data Space” - abrufbar auf der DIO Website:
https://dataintelligence.at/downloads/

AI ACT -
WAS STEHT DRINNEN

AI ACT: GOALS & RISKS

Der EU AI Act ist das erste umfassende Gesetz zur Regulierung von Künstlicher Intelligenz (KI) weltweit. Er trat am 1. August 2024 in Kraft und zielt darauf ab, sowohl die Sicherheit als auch die Grundrechte der Bürgerinnen und Bürger zu schützen sowie Innovation im Bereich KI zu fördern. Der AI Act schafft einen einheitlichen Rechtsrahmen für die Entwicklung, den Einsatz und die Nutzung von KI-Systemen innerhalb der Europäischen Union und gilt auch für Anbieter außerhalb der EU, wenn deren Systeme innerhalb der EU verwendet werden.

Das Gesetz folgt einem risikobasierten Ansatz, bei dem KI-Systeme in vier Kategorien eingeteilt werden. Systeme mit „unvertretbarem Risiko" sind grundsätzlich verboten – hierzu zählen etwa staatliche Social-Scoring-Systeme oder KI zur manipulativen Verhaltensbeeinflussung. Hochrisiko-KI-Anwendungen, etwa in Bereichen wie Bildung, Beschäftigung, Strafverfolgung oder biometrischer Überwachung, unterliegen strengen Auflagen. Diese betreffen unter anderem die Transparenz, die Qualität der verwendeten Daten, die menschliche Aufsicht und die Durchführung von Konformitätsprüfungen. Systeme mit begrenztem Risiko müssen bestimmte Transparenzpflichten erfüllen, zum Beispiel klarstellen, dass es sich um KI-generierte Inhalte handelt. Anwendungen mit minimalem Risiko, wie KI-gestützte Spamfilter oder Videospiele, sind von speziellen Auflagen ausgenommen.

GENERATIVE AI & BUSINESS SUPPORT

Ein besonderer Fokus liegt auf generativer KI, etwa Sprachmodellen wie ChatGPT. Für solche Systeme gelten zusätzliche Transparenzpflichten, wie die Offenlegung von Trainingsdaten, die Einhaltung des Urheberrechts und die Kennzeichnung von KI-generierten Inhalten. Open-Source-Modelle werden dabei mit geringeren Anforderungen bedacht.

Der AI Act sieht eine gestaffelte Umsetzung vor: Erste Bestimmungen gelten ab Februar 2025, unter anderem für KI mit unvertretbarem Risiko. Ab August 2026 treten weitere Regeln, etwa Transparenzanforderungen für generative KI, in Kraft. Hochrisiko-KI-Systeme haben eine längere Übergangsfrist bis August 2027. Bei Verstößen drohen hohe Geldstrafen – bis zu 35 Millionen Euro oder 7 % des weltweiten Jahresumsatzes.

Zur Förderung kleiner und mittlerer Unternehmen sowie Start-ups sind vereinfachte Verfahren und Unterstützungsangebote geplant, unter anderem ein zentraler AI Act Service Desk. Ziel ist es, die Einhaltung des Gesetzes auch für kleinere Akteure praktikabel zu gestalten, ohne Innovation zu bremsen.

AI - WHAT DO WE NEED TO START?

COMPLIANCE & ETHIK

Einhaltung von Datenschutzgesetzen (z. B. GDPR) und Berück-sichtigung ethischer Fragen (Value based Engineering)

DATENSAMMLUNG

- Daten aus vernetzten Produkten und zugehörigen Dienstleistungen Produktdaten, die durch die Nutzung vernetzter Produkte generiert werden
- Dienstleistungsbezogene Daten, die während der Bereitstellung von Dienstleistungen entstehen
- beinhaltet sowohl Rohdaten als auch vorverarbeitete Daten
- umfasst sowohl personenbezogene als auch nicht-personenbezogene Daten
- schließt Inhalte aus, die durch Urheberrechte geschützt sind

UPDATES

Retraining des Modells und Integration neuer Funktionen.

Nutzt man sogenanntes **überwachtes Lernen**, ist ein manuelles oder automatisiertes Labeling der Daten notwendig.

FEEDBACKSCHLEIFEN

Sammlung von Nutzer:innen-rückmeldungen und Verbesserung des AI Projekts.

SKALIERUNG

Optimierung für größere Datenmengen und höhere Nutzerzahlen

DATENAUFBEREITUNG

Das Datengesetz muss sicherstellen, dass erhöhter Datenaustausch weder die Privatsphäre noch bestehende Datenschutzgesetze beeinträchtigt.

Mit zunehmendem Datenaustausch wird die Aufrechterhaltung der Datenqualität noch entscheidender für eine effektive Entwicklung.

Ohne Standardisierung könnten die Vorteile des Datenaustauschs eingeschränkt sein.

Interoperabilität ist entscheidend für den nahtlosen Datenaustausch zwischen verschiedenen Anbietern, Geräten und Systemen.

MODELLTRAINING

- ergänzt die bestehenden EU-Datenschutzgesetze (DSGVO und ePrivacy)
- ergänzt die bestehenden EU-Verbraucherschutzgesetze
- beeinträchtigt bestehende Schutzgesetze für geistiges Eigentum nicht
- gilt nicht für Strafverfolgung, nationale Sicherheit oder Bereiche außerhalb der Zuständigkeit

DATEN-INTER-OPERABILITÄT

DATA SPACES FÜR DEN AI ACT

Data Spaces – also föderierte, vertrauenswürdige Datenräume – können eine entscheidende Rolle bei der Umsetzung des EU AI Act spielen. Sie unterstützen nicht nur die Einhaltung der gesetzlichen Anforderungen, sondern fördern auch die Zielsetzung des Gesetzes: die Entwicklung sicherer, transparenter und grundrechtskonformer KI-Systeme in Europa.

ENSURING DATA QUALITY

Ein zentrales Element des AI Act ist die Qualität der Daten, insbesondere bei Hochrisiko-KI-Anwendungen. Data Spaces helfen hier, indem sie strukturierte, semantisch standardisierte Datenumgebungen schaffen, in denen Unternehmen auf hochwertige, aktuelle und vertrauenswürdige Daten zugreifen können. Das verbessert nicht nur die Datenbasis für KI-Modelle, sondern unterstützt auch die geforderte Nachvollziehbarkeit und Prüfbarkeit der Datennutzung. So lässt sich etwa die Herkunft von Daten – also deren Provenienz – in einem Dataspace eindeutig dokumentieren und auditieren, was die Transparenzanforderungen des AI Act erfüllt.

DATA SOVEREIGNTY & COMPLIANCE

Gleichzeitig legen Data Spaces großen Wert auf Datensouveränität. Teilnehmende Organisationen behalten die volle Kontrolle darüber, wer auf ihre Daten zugreift und zu welchem Zweck. Dies steht im Einklang mit den Anforderungen des AI Act und der Datenschutz-Grundverordnung (DSGVO), insbesondere in Bezug auf die Wahrung der Grundrechte. Dadurch können KI-Systeme datenschutzkonform und dennoch innovationsfreundlich entwickelt und eingesetzt werden.

DIO

DATA SPACES FÜR DEN AI ACT

Auch im Bereich generativer KI leisten Dataspaces einen wertvollen Beitrag. Der AI Act fordert bei solchen Systemen transparente Angaben über die verwendeten Trainingsdaten und die Einhaltung von Urheberrechten. Dataspaces ermöglichen den Einsatz gekennzeichneter, lizenzkonformer Datenquellen mit klaren Nutzungsrechten und nachvollziehbarer Herkunft – ein wichtiger Schritt, um rechtssichere generative KI-Modelle zu trainieren.

STANDARDS & IMPLEMENTATION

Darüber hinaus fördern Dataspaces Innovation im regulierten Rahmen. Sie bieten die Möglichkeit, sogenannte KI-Sandboxes einzurichten, in denen neue Technologien in geschützter Umgebung getestet werden können – ein Angebot, das insbesondere Start-ups und kleinen Unternehmen den Zugang zu hochwertiger Dateninfrastruktur erleichtert, ohne dass sie regulatorische Risiken eingehen müssen.

Schließlich tragen Data Spaces durch ihre Orientierung an gemeinsamen europäischen Standards und Interoperabilität dazu bei, ein vernetztes und regelkonformes KI-Ökosystem in Europa zu schaffen. Sie schaffen die technologische und organisatorische Grundlage, um den AI Act nicht nur umzusetzen, sondern aktiv mit Leben zu füllen.

Insgesamt leisten Data Spaces somit einen konkreten Beitrag zur Operationalisierung der KI-Regulierung und schaffen die Voraussetzungen für eine europäische KI, die vertrauenswürdig, rechtssicher und zugleich wettbewerbsfähig ist.

Info

https://digital-strategy.ec.europa.eu/de/policies/data-act

Vgl. Art. 14 ff. DA
Vgl. Art. 14 i.V.m. Art. 15 DA
Vgl. Art. 23 ff. DA
Vgl. Art. 25 DA
Vgl. Art. 29 DA
Vgl. Art. 30 DA
Vgl. Art. 30 & 35 DA
Vgl. Art. 35 DA
Vgl. Art. 32 DA

(1) Kapitel 2-4 des Data Act

DATA SPACES FÜR AI

Wenn man Data Spaces für Künstliche Intelligenz nutzen möchte – also nicht selbst einen Datenraum aufbaut, sondern auf bestehende Datenräume zugreift, um KI-Anwendungen zu entwickeln oder zu betreiben – sind verschiedene technische, rechtliche und organisatorische Schritte erforderlich. Ziel ist es, vertrauenswürdige Datenquellen regelkonform zu nutzen und gleichzeitig die Anforderungen des EU AI Act sowie der DSGVO einzuhalten.

1. Zunächst gilt es, einen geeigneten Data Space zu finden, der für den konkreten Anwendungsfall relevante Daten anbietet. Es existieren bereits branchenspezifische Datenräume wie Catena-X (Automobilindustrie), der Mobility Data Space (Mobilitätsdaten) oder Health-X (Gesundheitsdaten).

2. Um an einem Data Space teilnehmen zu können, wird die notwendige Software installiert und heruntergeladen. Welche Software nutzbar ist, findet man in unserem Blueprint „How to use Data Spaces“. Darüber werden Zugriffsrechte, Nutzungsrichtlinien und Datensicherheit umgesetzt. Die eigene Infrastruktur muss so gestaltet sein, dass sie sowohl mit den Standards des Datenraums als auch mit den Anforderungen des AI Act kompatibel ist.

3. Im nächsten Schritt registriert man sich als Datenkonsument:in oder Datenlieferant:in im jeweiligen Datenraum. Die Teilnahmebedingungen, Lizenzmodelle und Nutzungsregeln müssen geprüft und akzeptiert werden. Wichtig ist dabei, auch die vertraglichen Rahmenbedingungen zu verstehen – insbesondere im Hinblick auf Datenschutz, Weiterverwendung der Daten und mögliche Haftungsfragen.

4. Hat man Zugriff auf die gewünschten Daten, müssen diese verantwortungsvoll genutzt werden. Insbesondere bei Trainingsdaten für KI-Systeme ist auf Datenqualität, Relevanz und rechtliche Zulässigkeit zu achten. Bei generativer KI sind zusätzlich Transparenzpflichten zu erfüllen – etwa hinsichtlich der Herkunft und Nutzung der Daten sowie möglicher urheberrechtlicher Einschränkungen.

5. Mit den gewonnenen Daten können KI-Modelle trainiert, validiert oder betrieben werden. Dabei ist stets zu prüfen, ob es sich um eine Hochrisiko-Anwendung im Sinne des AI Act handelt. In solchen Fällen sind umfangreiche Anforderungen zu erfüllen – etwa in Bezug auf menschliche Aufsicht, Risikomanagement, Dokumentation und Konformitätsbewertung. Auch für generative KI gelten besondere Regeln, etwa zur Kennzeichnung KI-generierter Inhalte.

DATA SPACES FÜR AI

6

Ein zentrales Element ist die Einhaltung gesetzlicher Vorgaben. Dazu zählen nicht nur der AI Act, sondern auch die DSGVO und andere branchenspezifische Vorschriften. Es sollte daher ein internes Compliance-Management eingerichtet werden, das die Datennutzung dokumentiert, Datenflüsse nachvollziehbar macht und bei Bedarf gegenüber Aufsichtsbehörden Nachweise erbringen kann.

7

Data Spaces bieten nicht nur Zugang zu Daten, sondern auch die Möglichkeit zur Zusammenarbeit mit anderen Akteuren. So lassen sich beispielsweise Modelle gemeinsam entwickeln, Forschungspartnerschaften eingehen oder Innovationsprojekte gemeinsam gestalten. Der Data Space dient dabei als vertrauenswürdige Plattform für gemeinschaftliche Wertschöpfung.

Wer KI-Systeme innerhalb von Dataspaces einsetzen möchte, profitiert von einem strukturierten, sicheren und rechtlich abgesicherten Datenzugang. Gleichzeitig müssen die damit verbundenen Regeln bekannt sein und eingehalten werden. Mit der richtigen technischen Anbindung, klaren Prozessen sowie einem Bewusstsein für ethische und rechtliche Aspekte lässt sich KI im europäischen Kontext sicher, innovativ und im Einklang mit dem AI Act nutzen

FEDERATED LEARNING

Federated Learning ist ein dezentraler Ansatz des maschinellen Lernens, bei dem KI-Modelle direkt auf den Geräten oder Systemen trainiert werden, auf denen die Daten entstehen – etwa auf Smartphones, in Krankenhäusern oder Unternehmen. Dafür können dezentrale Dataspaces genutzt werden.

SO FUNKTIONIERT'S

Dabei verbleiben die Daten stets lokal und werden nicht an eine zentrale Stelle übertragen. Stattdessen wird ein gemeinsames Grundmodell an alle beteiligten Instanzen verteilt, dort lokal mit den jeweiligen Daten trainiert und anschließend in Form von Modell-Updates – beispielsweise veränderten Gewichtungen – an einen zentralen Server zurückgeschickt. Dieser aggregiert die Updates und erstellt daraus ein verbessertes globales Modell, das wiederum verteilt wird. Dieser Zyklus wird so lange wiederholt, bis das Modell eine ausreichende Genauigkeit erreicht.

SICHER & DATENSCHUTZKONFORM

Der große Vorteil dieses Verfahrens liegt darin, dass personenbezogene oder sensible Daten – etwa Gesundheits- oder Nutzungsdaten – nicht zentral gespeichert oder übermittelt werden müssen. Das macht Federated Learning besonders attraktiv für Anwendungsbereiche mit hohen Datenschutzanforderungen, etwa im Gesundheitswesen, in der Industrie oder bei mobilen Anwendungen. Gleichzeitig trägt dieser Ansatz dazu bei, regulatorische Anforderungen – wie jene der Datenschutz-Grundverordnung (DSGVO) oder des AI Act – zu erfüllen, da die Kontrolle über die Daten weitgehend bei den jeweiligen Datenhaltern verbleibt.

So funktioniert Federated Learning

- Ein globales Modell wird initialisiert und an mehrere Teilnehmer (z. B. Smartphones, Kliniken, Unternehmen) verteilt.
- Jeder Teilnehmer trainiert das Modell lokal mit seinen eigenen Daten.
- Die lokal trainierten Modelle oder deren Updates werden zurück an einen zentralen Koordinator geschickt.
- Der Koordinator aggregiert die Updates (z. B. via Federated Averaging) und aktualisiert das globale Modell.
- Der Zyklus beginnt erneut, bis das Modell konvergiert.

FEDERATED LEARNING

HÜRDEN & RISIKEN

Neben diesen Vorteilen bringt Federated Learning auch Herausforderungen mit sich. Die lokal vorliegenden Daten sind häufig sehr unterschiedlich verteilt, was das Training erschwert. Auch der Kommunikationsaufwand kann erheblich sein, da regelmäßig Modell-Updates ausgetauscht werden müssen. Zudem besteht theoretisch die Möglichkeit, dass über sogenannte Modellinversionstechniken Informationen aus den Updates rekonstruiert werden können – was zusätzliche Schutzmaßnahmen erforderlich macht.

CHANCEN FÜR ZUSAMMENARBEIT

Trotz dieser Herausforderungen bietet Federated Learning einen innovativen und vielversprechenden Weg, um kollaborative KI-Systeme zu entwickeln, ohne zentrale Datenpools aufzubauen. Es eignet sich insbesondere für Szenarien, in denen viele Akteure mit sensiblen oder proprietären Daten gemeinsam an der Verbesserung von KI-Modellen arbeiten möchten – und dabei Datenschutz, Datensouveränität und regulatorische Vorgaben von Anfang an mitdenken.

DATEN FÜR AI NUTZEN

Beispiel - Healthcare:

Ein anschauliches Beispiel für den Einsatz dezentraler Dataspaces beim KI-Training findet sich im Gesundheitswesen – etwa im Bereich der radiologischen Bilddiagnostik:

Stellen wir uns vor, mehrere Krankenhäuser in Europa möchten gemeinsam ein KI-Modell entwickeln, das auf Röntgen- und MRT-Aufnahmen Lungenerkrankungen wie Lungenkrebs oder COVID-19 erkennt. Jedes Krankenhaus verfügt über große Mengen hochwertiger Bilddaten – doch aus Datenschutzgründen, insbesondere aufgrund der DSGVO, dürfen diese Patientendaten nicht ohne Weiteres an eine zentrale Stelle übertragen oder gemeinsam gespeichert werden.

Hier kommen dezentrale Data Spaces in Kombination mit Federated Learning ins Spiel: Jedes Krankenhaus betreibt in seinem eigenen IT-System einen vernetzten, aber datensouveränen Knotenpunkt innerhalb eines gemeinsamen medizinischen Data Spaces. Dieser regelt standardisiert und rechtssicher, wie auf Daten zugegriffen werden darf, unter welchen Bedingungen ein Datensatz genutzt werden kann und wer welche Rolle im Netzwerk übernimmt.

Anstatt die Bilddaten an einen zentralen Server zu übermitteln, wird das gemeinsame KI-Modell direkt an die beteiligten Kliniken ausgespielt. Dort wird es lokal mit den vorhandenen Bilddaten trainiert. Die dabei entstehenden Modell-Updates – also die trainierten Parameter – werden an eine zentrale Instanz (z. B. den Koordinator des Konsortiums) zurückgeschickt, die diese zu einem verbesserten globalen Modell aggregiert. Die Patientendaten selbst verlassen dabei zu keinem Zeitpunkt die jeweilige Klinik.

Gleichzeitig dokumentieren die Mechanismen des Data Spaces, woher die Trainingsdaten stammen, unter welchen Bedingungen sie verwendet wurden und wie sie sich auf das Modell ausgewirkt haben – also genau jene Nachvollziehbarkeit und Transparenz, die auch der AI Act fordert.

Ein solcher Ansatz ermöglicht es, datenschutzkonform, ethisch verantwortungsvoll und zugleich hocheffizient ein leistungsfähiges KI-Modell zu entwickeln. Zudem lässt sich durch den Einsatz von Data Spaces sicherstellen, dass die Datenquellen vertrauenswürdig sind, Nutzungsrechte eingehalten werden und etwaige Auditierungen oder Zertifizierungen möglich sind – was gerade bei Hochrisiko-KI im medizinischen Bereich von zentraler Bedeutung ist.
Dieses Beispiel zeigt: Dezentrale Data Spaces können nicht nur technologische Innovation ermöglichen, sondern auch regulatorische Anforderungen aktiv unterstützen und Vertrauen in KI-Anwendungen stärken.