

OpenFLaaS

Open Federated Learning as a Service

Projektbeschreibung

OpenFLaaS erstellt einen Baukasten für KI-Dienstleistungen, mit denen Unternehmen Dokumente wie Ausschreibungen, Verträge und Produktblätter automatisiert analysieren und so das darin enthaltene Wissen in digitalisierter Form effizient nutzen können. Über die OpenFLaaS-Plattform werden die KI-Dienstleistungen dabei in geschützten Räumen auf Endgeräten (Edge-Geräten) im Unternehmen ausgeführt. Die Rohdaten der Dokumente verbleiben im Unternehmen. Der sichere, souveräne und vertrauenswürdige Austausch der Daten sowie die Ausführung der KI über die verschiedenen Geräte hinweg erfolgt in Datenräumen, die dem Gaia-X-Standard folgen. OpenFLaaS fördert so die sichere und effiziente Digitalisierung von Unternehmenswissen zur Produktivitätssteigerung unter Wahrung der Datenhoheit der Unternehmen. Zugleich legt OpenFLaaS die Grundlage für den Aufbau eines Gaia-X-kompatiblen Ökosystems für die intelligente Dokumentenanalyse (IDA) und die Entwicklung systemunabhängiger, innovativer Dokumentenanalysedienste.

Marktperspektive und Produktversprechen

Unternehmen sammeln ihr Wissen, und damit ihr zentrales Unternehmensgut, in unzähligen Dokumenten. Mithilfe von KI lässt sich dieses Wissen so analysieren, dass es zur Steigerung der Effizienz, zum Erkennen neuer Zusammenhänge und so zur Extraktion neuer Informationen, zur Entscheidungsunterstützung, der Erweiterung von Serviceangeboten und vielem mehr genutzt werden kann. Angesichts dieser Vorteile wundert es nicht, dass die KI-gestützte Dokumentenanalyse bereits heute ein Marktvolumen von weltweit über einer Milliarde Dollar erreicht, obwohl das Feld noch in seinen Anfängen steckt.

Ein großes Hemmnis für IDA besteht derzeit allerdings noch in dem Dilemma, dass die rechenintensiven Analysen auf Cloud-Rechnern erfolgen. Aus Datenschutzgründen und mangelndem Vertrauen in Cloudanbieter verzichten viele Unternehmen derzeit noch auf die Potenziale von IDA. Das Projekt openFLaaS löst dieses Dilemma, indem es Unternehmen IDA als Dienstleistungen (as a Service – aaS) ermöglicht, bei denen die Daten im Unternehmen verbleiben. Möglich wird das durch die Entwicklung von Edge-KI-Räumen, die einfach per Plug-and-Play mit Gaia-X-konformen Datenräumen verbunden werden. Den Unternehmen bietet OpenFLaaS damit ein Cloud-Edge-System für intelligente Dokumenten- und Textanalysedienste auf Edge-Geräten wie industriellen Anlagen und Maschinen, Computern, Smartphones oder Tablets. Die Unternehmen behalten bei Nutzung dieser Dienste stets die volle Kontrolle über ihre Daten.

Auf den Ergebnissen von OpenFLaaS können beliebige IDA-basierte Mehrwertprodukte und -dienstleistungen in Deutschland und Europa entwickelt werden. Diese wiederum lassen sich mit der Projektlösung kontinuierlich verbessern, und zwar sowohl hinsichtlich ihrer Generalisierbarkeit auf beliebige Wissensbereiche als auch der Anpassung, beziehungsweise Spezialisierung, auf bestimmte Domänen. Dank neuartiger Algorithmen wird der CO₂-Ausstoß der Analyseverfahren minimiert, da weniger Datentransaktionen zwischen Edge und Cloud notwendig sind. Die Rechtsdiskussion und die Rechtsprechung im Bereich der Datennutzung für Text- und Data-Mining (TDM) sind derzeit noch sehr dynamisch. Im Projekt werden die rechtlichen Risiken dazu abgeschätzt und Rahmenbedingungen definiert, unter denen eine Kommerzialisierung der Datenprodukte erfolgen kann.

Herausforderung und Innovation

OpenFLaaS setzt für die intelligente Dokumentenanalyse auf eine Dezentralisierung der KI. Die Analyse der Dokumente findet nicht in einer zentralen Cloud statt, sondern über Edge-Geräte hinweg. Den Ansatzpunkt dafür bietet das sogenannte föderierte Lernen (federated Learning). Dabei handelt es sich um ein verteiltes maschinelles Lernen, bei dem lediglich ein globales KI-Modell in der Cloud trainiert wird, die Trainingsdaten aber lokal gespeichert werden, was datenschutzkonforme Lösungen ermöglicht.

Die Innovation von OpenFLaaS besteht darin, die Prinzipien des föderierten Lernens auf Edge-Ebene in geschützten KI-Räumen mit den offenen Standards von Gaia-X zu kombinieren. Das ermöglicht das Training von KI-Modellen für die intelligente Dokumentenanalyse unter Wahrung der Datensouveränität. Im Projekt werden dafür sichere, föderierte Datenräume in einem Gaia-X-konformen Cloud-Edge-Kontinuum entwickelt und umgesetzt. Am Beispiel industrieller Dokumententypen wie Verträgen, technischen CAD-Zeichnungen und Ausschreibungsspezifikationen entwickelt OpenFLaaS föderierte intelligente Dienste, die in der Lage sind, unterschiedliche Texttypen und Layouts wie Tabellen, Fließtext, Überschriften, Formeln, Bilder, Kopf- und Fußzeilen etc. zu verstehen und zu analysieren. Die im Projekt entwickelte Software wird als Open-Source bereitgestellt.

Use Cases

Die Projektinnovation wird anhand eines Prototypen für die Planung und Ausschreibung industrieller Anlagen auf Basis realer Daten demonstriert. In einem zweiten Demonstrator werden dieselben Prinzipien für die KI-gestützte Analyse von klinischen Dokumenten zu Patienten auf Edge-Geräten übertragen und umgesetzt. Detaillierte Anforderungen von Anwendern und Nutzern bezüglich der technischen Realisierung, Datensouveränität und Benutzerfreundlichkeit werden gesammelt und die prototypischen Anwendungen mit Experten aus den beiden Anwendungsfeldern validiert.

OHNE OpenFLaaS

Zur industriellen Dokumentenanalyse sind Unternehmen in Deutschland und Europa weitgehend auf amerikanische und chinesische Cloud-Anbieter mit Monopol- bzw. Oligopolstellung angewiesen. Aus mangelndem Vertrauen und Datenschutzbedenken gegenüber diesen Anbietern wird auf IDA oft verzichtet, beziehungsweise werden die Potenziale von IDA nicht voll ausgeschöpft.

Da vertrauenswürdige, sichere Räume für den Austausch von Daten fehlen, bleibt die Entwicklung neuer digitaler Geschäftsmodelle und Dienstleistungen im Bereich IDA hinter den technologischen Möglichkeiten.

In der Industrie gibt es viele spezifische Dokumententypen mit spezifischem Branchenwissen, für deren Analyse KI speziell trainiert werden muss. Bislang gibt es keine Angebote für das Training und die Entwicklung derart hochspezialisierter IDA.

Etablierte Anbieter im Bereich IDA bieten Lösungen an, bei denen die Kunden nicht erkennen können, wie die Dienstleistungen zustande kommen („Blackboxing“).

KI-basierte Dokumentenanalyse ist mit hohen Datentransferaten in die Cloud verbunden und verursacht daher hohe CO₂-Emissionen.

MIT OpenFLaaS

OpenFLaaS ermöglicht die dezentrale Ausführung auf IDA auf Edge-Ebene, sodass Unternehmen keine Daten in die Cloud laden müssen. Sie haben daher auch keinen Wissensabfluss durch Daten zu fürchten und können die Einhaltung von Datenschutzbestimmungen auch gegenüber Dritten gewährleisten.

openFLaaS schafft sichere und vertrauenswürdige KI-Räume für den Austausch von Daten und ermöglicht damit neue digitale Geschäftsmodelle und Dienstleistungen für eine Edge-orientierte Datenwirtschaft.

openFLaaS ermöglicht die Realisierung eines Gaia-X-kompatiblen Knotens für die industrielle Dokumentenanalyse in Deutschland und fördert damit den Aufbau von Diensten zur KI-basierten industriellen Entscheidungsunterstützung zur Optimierung der Produktivität.

openFLaaS schafft mit einer Open-Source-Lösung Transparenz über IDA-basierte Mehrwertprodukte und -dienstleistungen und schafft so Vertrauen in die automatisierte Dokumentenanalyse.

openFLaaS ermöglicht eine nachhaltige IDA, da weniger Daten in die Cloud übertragen werden.

Laufzeit:

01.02.2023 – 31.01.2026

Konsortium

DRIMCO GmbH (Konsortialführer), Apheris GmbH, Siemens AG, Fraunhofer IAIS, Goethe-Universität Frankfurt, InfAI, Universität Wien (Österreich)

Ansprechpartner

Dr. Pankaj Gupta
pankaj.gupta@drimco.net

<https://infai.org/openflaas>