

AI Model Hub powered by IONOS

IT Business Live Cast

19. Mai 2025





Benjamin Schmidt

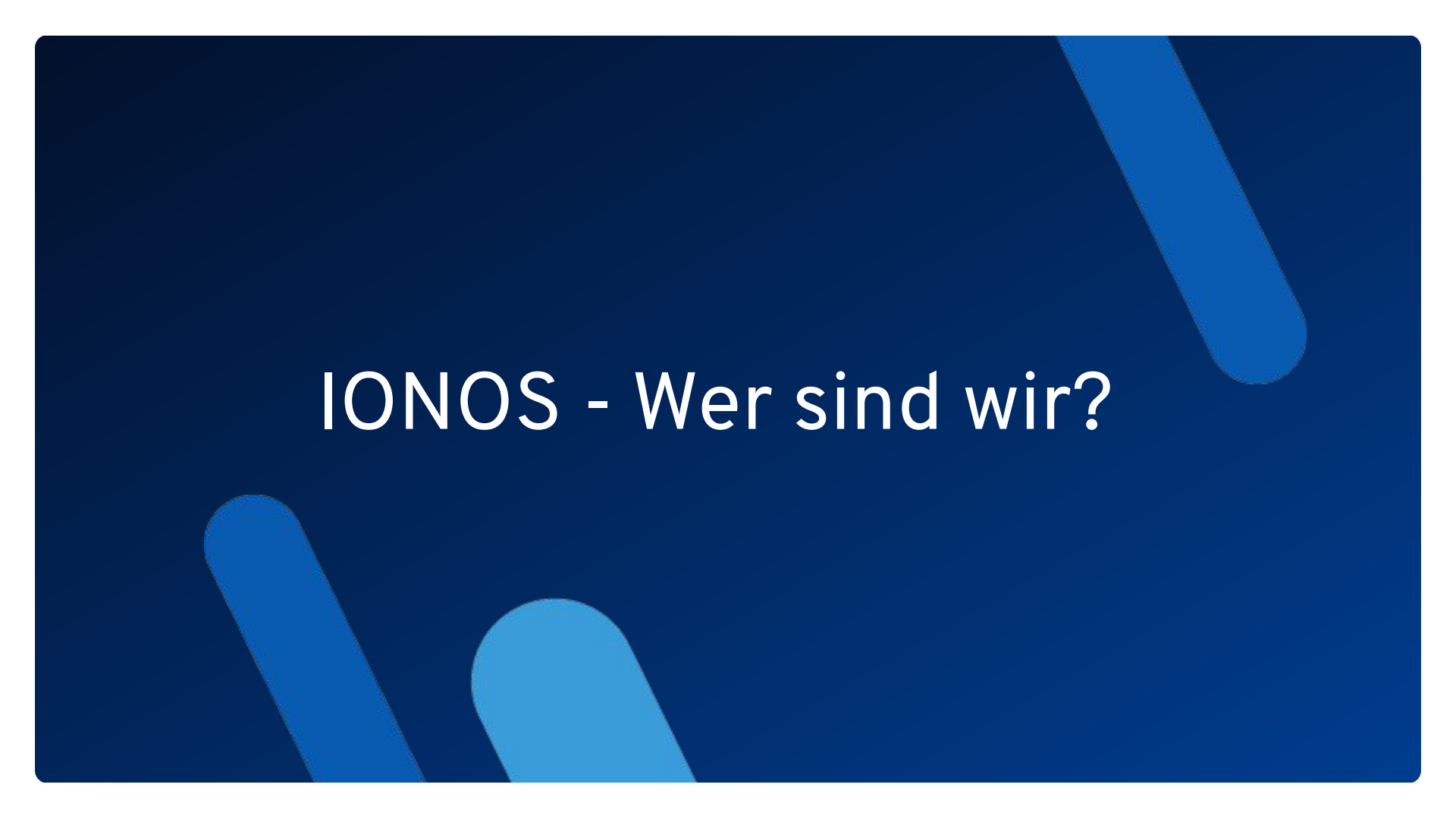
Senior Cloud Solutions Architect @IONOS



Peter Lobner

Senior Project Manager @eXXcellent solutions

- **IONOS Kurzvorstellung**
- **AI Model Hub**
 - Features
 - Use Case - Stadt Freiburg
- **Generative AI Pipelines**
 - Simple Chat
 - Retrieval Augmented Generation (RAG)
- **Partner Use Case - eXXcellent solutions**
 - Patent Engine
- **Q&A**



IONOS - Wer sind wir?

Up in the Cloud

Globale Plattform. Local Heroes. Universeller Erfolg.

IONOS

IONOS Markenfamilie

arsys

fasthosts

home.pl

InterNetX

IONOS

sedo
Buy. Park. Sell. Domains

STRATO

united domains

we22

WORLD4YOU®
INTERNET SERVICES GMBH

Unser **One-Stop-Shop-Portfolio** begleitet unsere Kundinnen und Kunden bei ihren Digitalisierungs-Prozessen.

IONOS GROUP SE

Europas führender
Digitalisierungs-Partner und
zuverlässiger Cloud Enabler
für den Mittelstand.

Datacenter: 31
Kunden: 6,28 Mio.
Umsatz: 1,4 Mrd. €
Mitarbeiter: 4.364

XETRA

ISIN: DE000A3E00M
WKN: A3E00M

Ein rundes Portfolio für Ihr Business

IaaS, PaaS & Managed Services



IONOS

AI & Data Analytics

AI Model Hub ■ Event Streams for Apache Kafka

Managed Stackable Data Platform

Compute Services, Containers & Bare Metal Services

Dedicated Core Server ■ vCPU Server ■ Cubes ■ Managed Kubernetes

Private Container Registry ■ Sovereign Private Cloud ■ GPU Server ■ ...

Databases & Storages

In-Memory DB ■ MariaDB ■ MongoDB ■ PostgreSQL

Backup Service ■ Block Storage ■ IONOS Object Storage ■ NFS ■ ...

Network Services

API Gateway ■ CDN ■ Cloud DNS ■ DDoS Protect ■ VPN Gateway

Managed Application Load Balancer ■ Managed NAT Gateway ■ ...

Security & Observability

Security Central Information ■ Container Scanning ■ Flow Logs

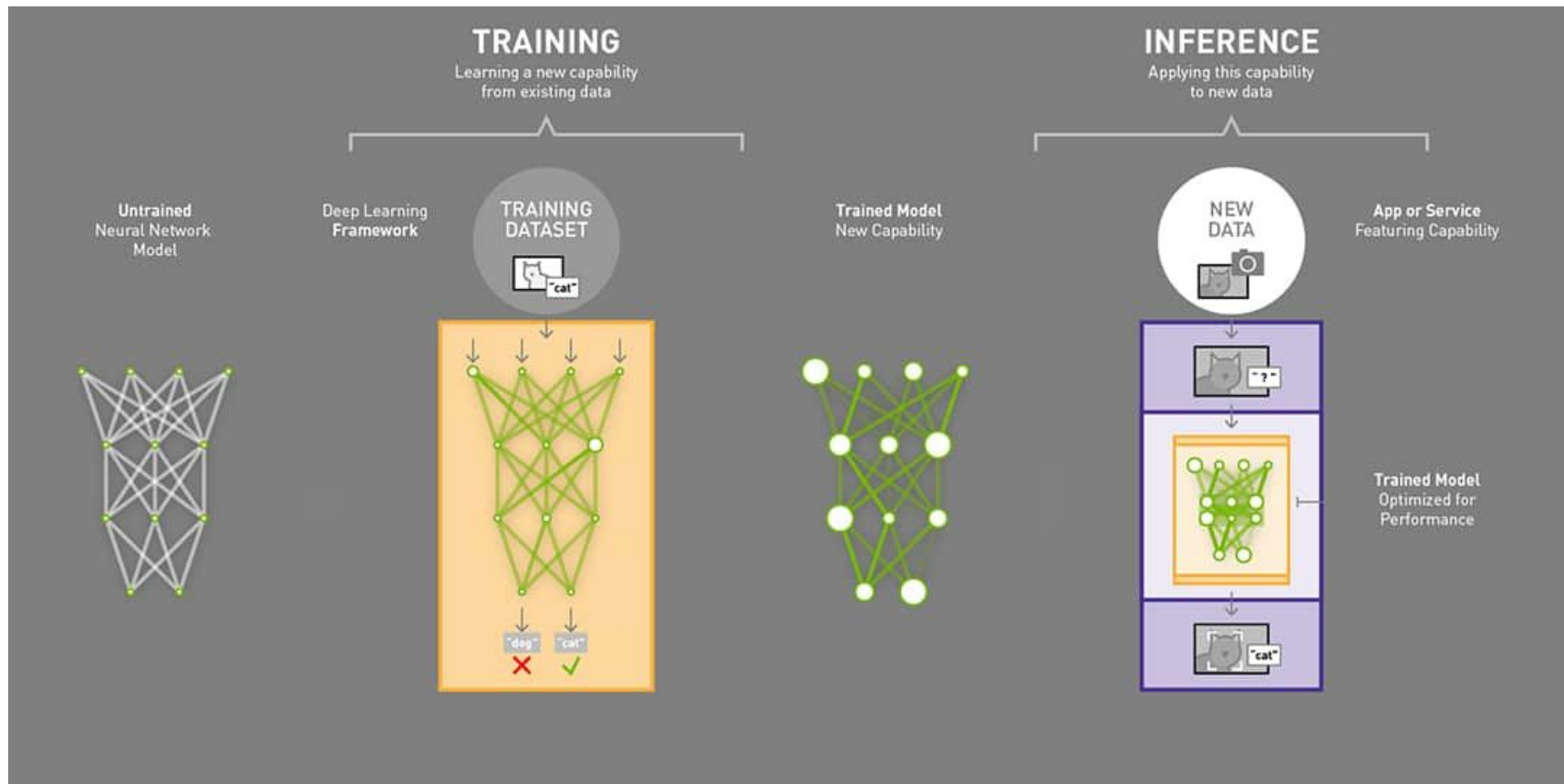
Activity Logs ■ Logging Service ■ Monitoring Service ■ ...

IONOS AI Model Hub

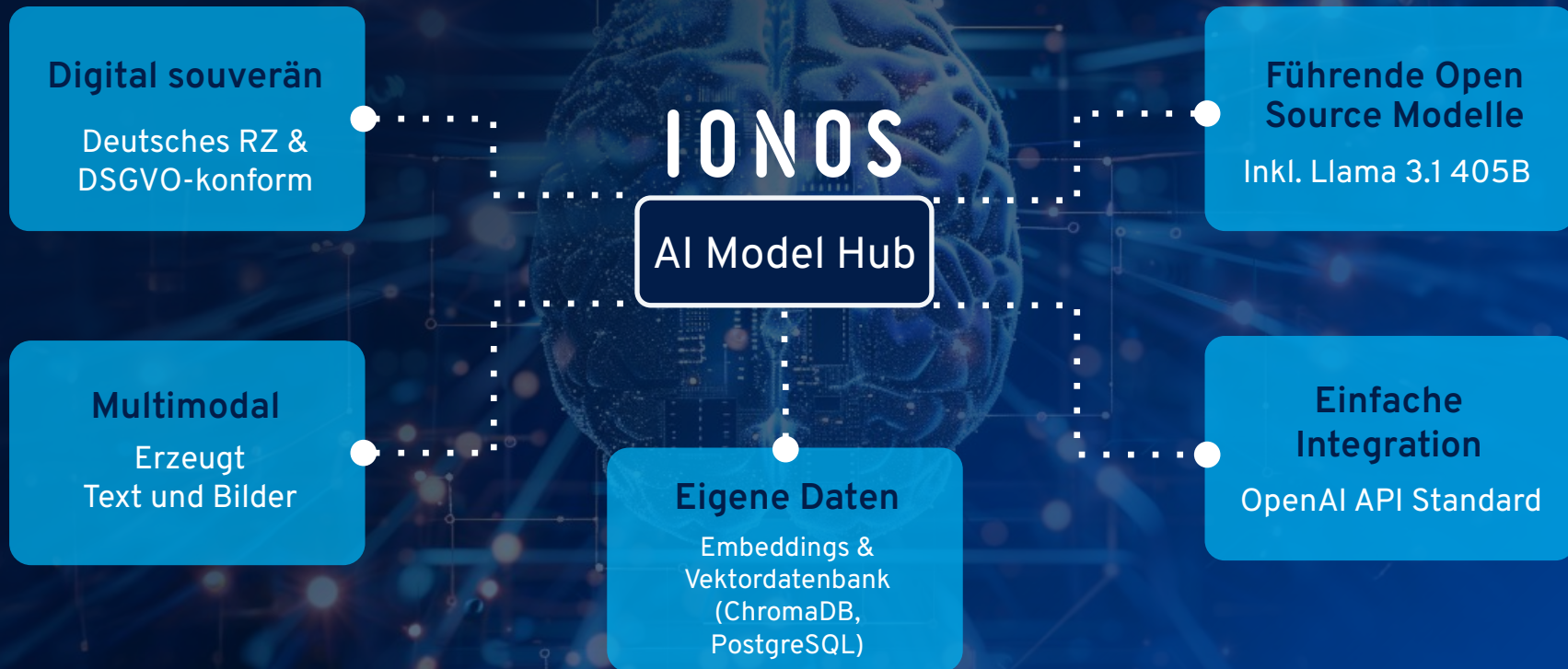
... ist ein in der Cloud gehosteter,

Inferenzdienst für künstliche Intelligenz.

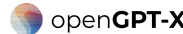




IONOS AI Model Hub: Erste multimodale souveräne KI



Verfügbare Modelle können aufgrund der schnellen Innovationen und der sich ändernden Marktpräferenzen in diesem Bereich jederzeit angepasst werden.

	Modelle	Use Cases
	- Llama 3.1 8B Instruct - Llama 3.3 70B Instruct - Llama 3.1 405B Instruct	Kundendienst, FAQ-Bots, interne Wissensabfrage.
	Code Llama 13B Instruct	Debugging, Code-Erklärung und -Generierung in mehreren Programmiersprachen.
	- Mistral 7B Instruct v0.2 - Mixtral 8x7B - Mistral Nemo	Erstellung juristischer Dokumente, mehrsprachige Textverarbeitung.
	Teuken 7B (Neu)	Schnelle mehrsprachige Textverarbeitung.
	Stable Diffusion XL	Kreative Werbekampagnen, Social-Media-Assets.
	Flux.1-schnell	Visuelle Inhalte für E-Commerce, Text-integrierte Bilder.

OpenAI kompatible APIs

- Chat Completion - jetzt mit Tools Calling
- Image Generation
- Embeddings

AI Model Hub native APIs

- Predictions (inkl. integriertem RAG)
- Document Collections
- Similarity Search

Document Collection Datenbanken

- ChromaDB
- PostgreSQL

Integriertes (optionales) Chunking

Data Center Designer (DCD) Playground

- Textgenerierung
- Bildgenerierung
- Embeddings

Generative AI Use Cases

Basierend auf den verschiedenen unterstützen Foundation Models

IT: **AI-powered Customer Support**

Implement AI bots to handle L1 and L2 support tickets, freeing up human agents for complex issues.

IT: **Software Development**

Copilot-like assistants are now widely used in dev workflows for code generation, review, and documentation

Marketing: **Brand Intelligence**

Sentiment analysis, trend identification, competitor brand analysis, website creation and updates

E-commerce: **Product Recommendation Engine**

Recommend products to customers based on their browsing history, purchase patterns, and preferences.

E-commerce: **Inventory Forecasting**

Use AI inferencing to predict inventory demand, avoiding overstock or stockouts

Manufacturing: **Predictive Maintenance**

Analyze IoT sensor data in real time to predict machine failures before they occur, minimizing downtime.

Healthcare: **Virtual Health Assistants**

Provide 24/7 virtual assistance for patients by answering common queries, booking appointments, and triaging symptoms.

Legal: **Legal and Compliance Review**

Provide 24/7 virtual assistance for patients by answering common queries, booking appointments, and triaging symptoms.

Finance: **Fraud Detection**

Detect fraudulent transactions in real-time by analyzing patterns in financial data.



Chat-Frontend



Wie hat sich die Anzahl...



Antwort des Chatbots



Prompt



Vektor Datenbank

75.000 Dokumente



LLM



Chatbot

Wie entwickelte sich die Anzahl der Kindergartenplätze in den letzten drei Jahren?



Die Anzahl der Kindergartenplätze in Freiburg hat sich in den vergangenen drei Jahren wie folgt entwickelt:

2021: 3.407 Kindergartenplätze

2022: 3.460 Kindergartenplätze

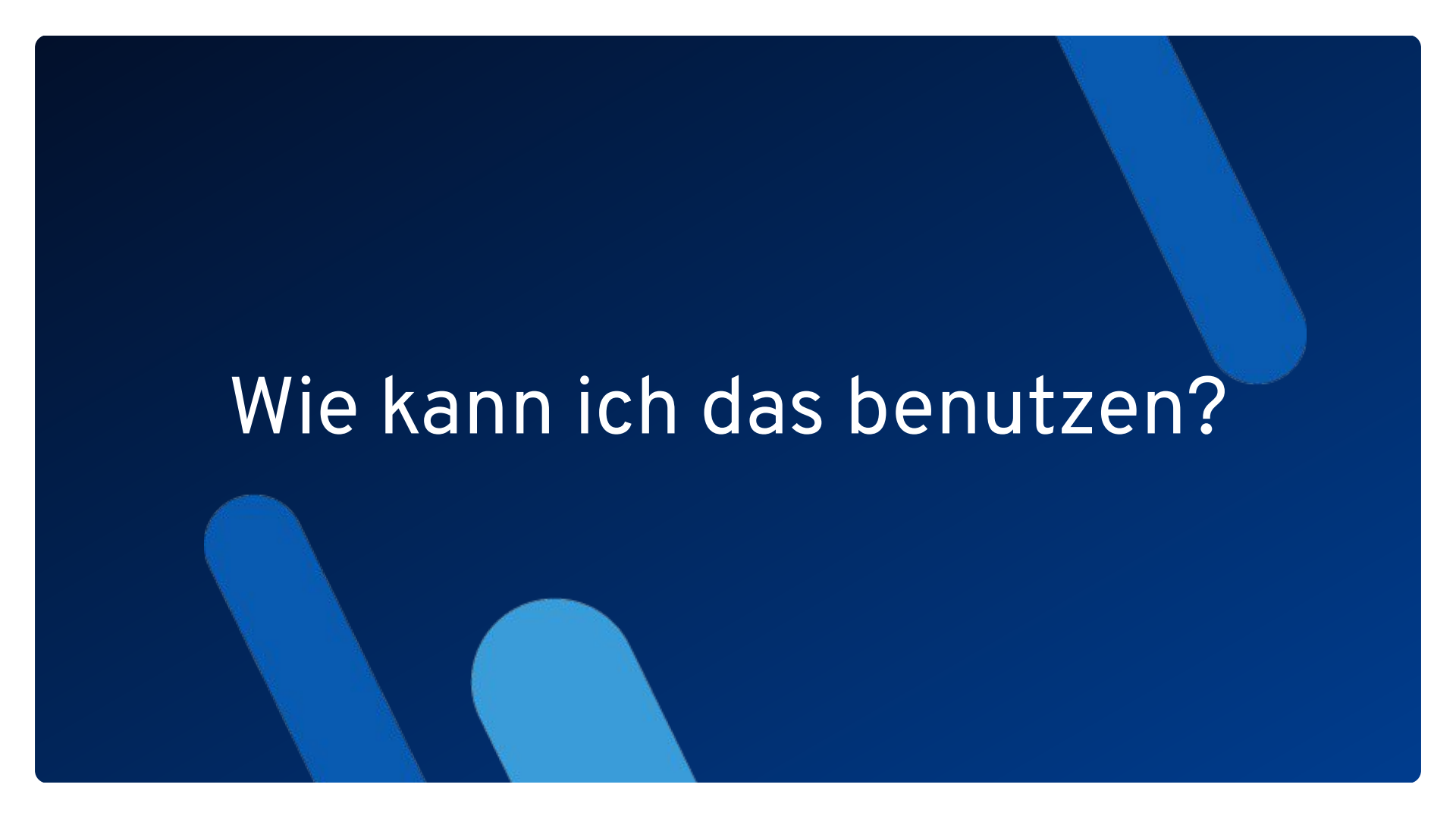
2023: 3.826 Kindergartenplätze



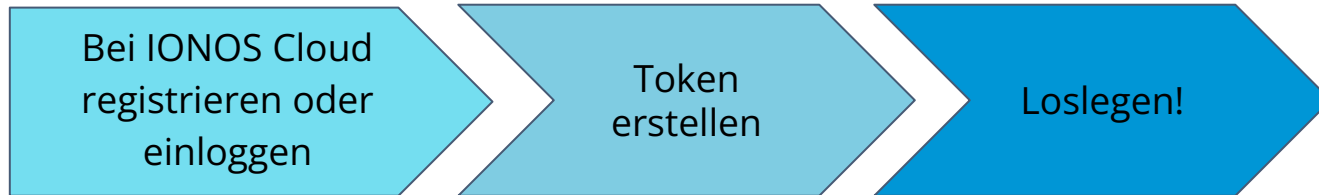
Die Anzahl ist um 419 Plätze innerhalb von drei Jahren gestiegen.

The background is a solid dark blue. There are three abstract, diagonal brush strokes in a lighter shade of blue. One stroke is in the top right corner, pointing towards the bottom left. Two other strokes are in the bottom left corner, pointing towards the top right.

Generative AI Pipelines



Wie kann ich das benutzen?





Partner Use Case: eXXcellent solutions

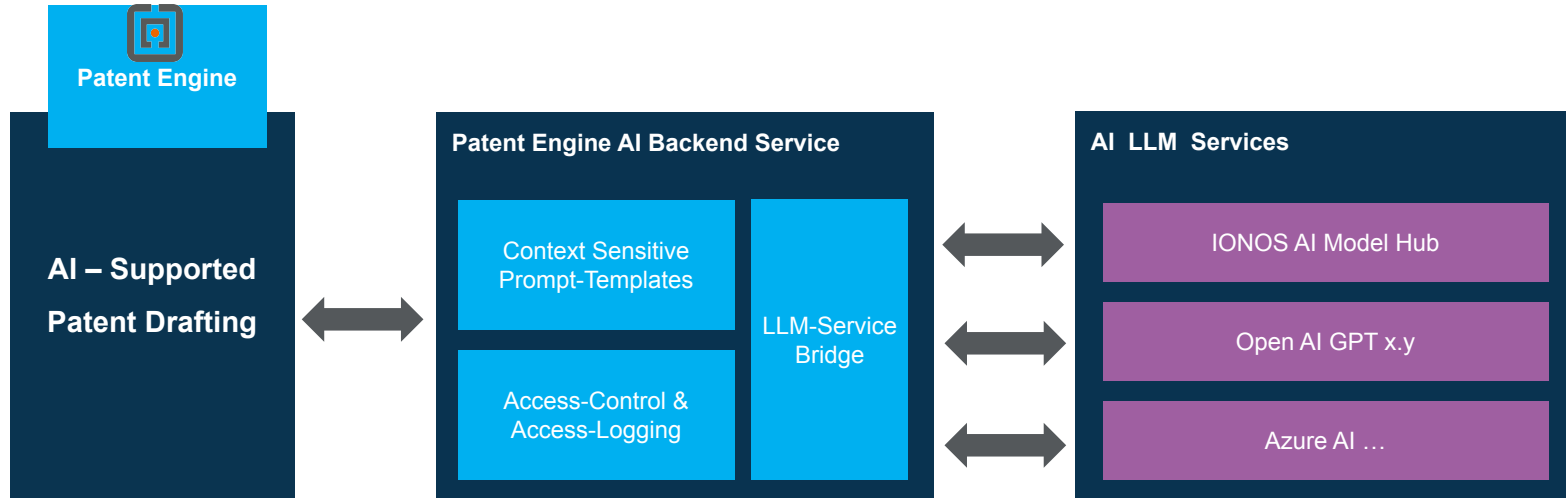
WE MAKE YOUR CLOUD VISION REALITY

From AI to Z.



EMPOWER YOUR BUSINESS BY INTEGRATING AI SUPPORT INTO YOUR BUSINESS APPLICATIONS

Example PatentEngine – AI Supported Patent Drafting





Empower your Business –

Integrate AI into your Applications.

The slide features a dark blue background with three abstract, rounded rectangular shapes in shades of blue. One shape is in the top right corner, and two are in the bottom left corner. The text "Q&A" is centered in a white, bold, sans-serif font.

Q&A

Danke für eure Aufmerksamkeit!

Mehr Informationen zum AI Model Hub unter:

- <https://cloud.ionos.de/managed/ai-model-hub>

Kontaktiert uns bei Fragen rund um das Partner Programm unter:

- channel-dach@ionos.com