SUSE

SUSE가 제안하는 Cloud Native AI 플랫폼

SUSE AI Suites

Extensible, Cloud Native Platform for AI Workloads

SUSE at a glance

1992
Founded in Nuremberg

HQ
Luxembourg

40
Offices

2,600
Employees

$700m+
Revenue

10,000+
Enterprise Customers

90%+
Of the world's leading companies rely on SUSE [6]

Common Criteria SLSA

최상의 SW 공급망 보안 인증 [3]

개발자 기여도 평가 →

공개적으로 세계 최대 기업들과 비교 평가

Consistently

Top 5 OSSRANK

Top 8 CLOUD NATIVE COMPUTING FOUNDATION

Top 12 OSCI

10/10
of the largest **Automotive** companies

13/15
of the largest **Pharmaceutical** companies

14/15
of the largest **Aerospace** companies

13/15
of the largest **FinServ** companies

컨테이너 관리와 가상화에서 리더로 인정 → [2]

Gartner IDC OMDIA

United Nations Global Compact

17% emissions reduction since 2022 [4]

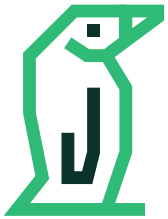
World-class ecosystem of partners
5 years in a row [5]

THE CHANNEL CO. CRN PARTNER PROGRAM GUIDE WINNER 2024

SUSE

[1] SUSE Company Data [2] Gartner Magic Quadrant + Omdia [3] Common Criteria, SLSA [4] The UN Global Compact [5] suse.com/partners [6] SUSE Customer Data

선택의 자유를 통해 고객의 차별화 지원



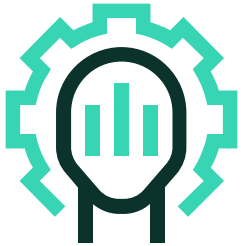
Linux



Cloud Native



Edge



AI

SUSE

GenAI 적용 상황

60%

데이터 보호가 가장 큰 걸림돌

25%

GenAI 도구를 성급하게 선택하면 명성에 해가될 수 있다고 생각

50%

프로젝트의 기대 성과 충족

6/6

고객 서비스가 6개 산업에서 비용 절감의 최우선 대상

!

PoC 의 범위가 넓어서 놀랍다.

>20%

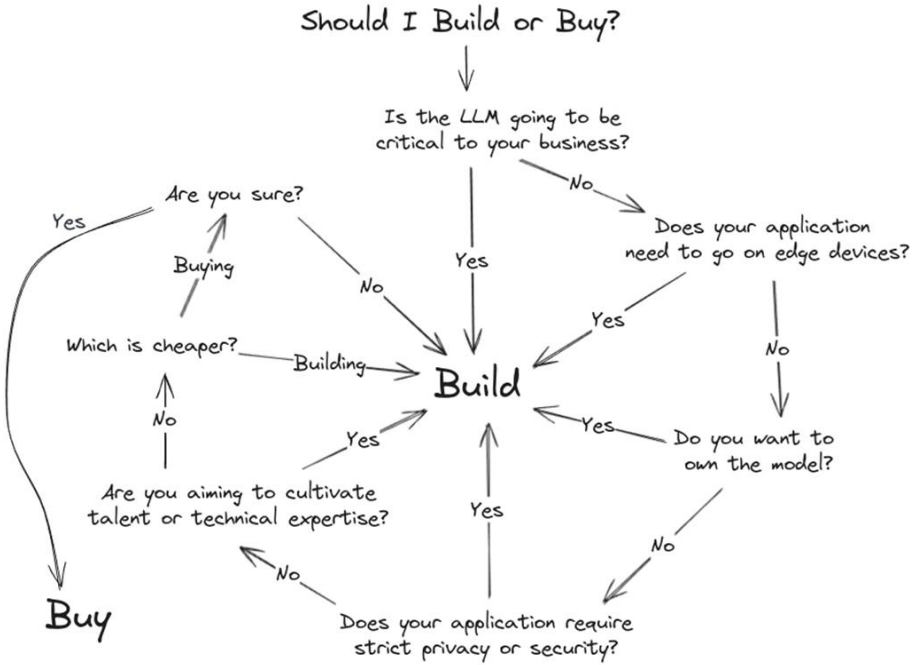
15조 이상 기업의 20% 이상은 수익 창출 보다는 비용 절감에 초점

(ref: JP Morgan)

AI - 만들 것인가 살 것인가??

LLM 및 규정된 "통합"에 액세스하는 방법은 다양합니다....

- 그들이 당신이 원하는 것을 해줄까요?
- 무엇이 더 저렴할까요?
- 비용, 품질 및 속도 (2개 선택)
- 피벗할 수 있나요?



대부분의 경우, 고객들은 "구축"을 해야하지만, AI 공급업체는 "지금 구매하세요!"라고 강조하며 지금 당장 구매하지 않으면 기회를 놓치고 경쟁우위가 감소한다고 위협합니다.

Key AI Challenges customers are facing

Security Threats	Regulatory Compliance
Rapidly Changing Tools	Bleeding edge Infrastructure

Challenge #1: 데이터 통합과 보안 위협

AI 워크로드는 새로운 보안 위협을 가져옴:

- → Data poisoning attacks
- → Personal and IP data leaks
- → Misuse of model outputs

SUSE AI platforms이 이런 문제를 해결



Challenge #2: 규제 준수

규제 환경은 진화하고 복잡하며 벌금은 비용이 많이 듭니다.

SUSE는 변화하는 요구 사항을 안전하고 예측 가능하게 충족할 수 있는 유연성과 보안 프레임워크를 제공합니다.

Fines: **€35million or 7%**

EU Artificial Intelligence Act

NIST NATIONAL INSTITUTE OF STANDARDS AND TECHNOLOGY U.S. DEPARTMENT OF COMMERCE

NIST AI 100-5: A Plan for Global Engagement on AI Standards

European Commission GDPR General Data Protection Regulation

US Executive Order on the Safe, Secure, and Trustworthy Development and Use of AI

Fines: **€20 million or 4%**

SUSELOGO SUSE AI Transparency. Accountability. Security. Safety.

Challenge #3: 급격하게 변하는 오픈소스 생태계

Models & LLMs

LLaMA by Meta · NVLM · MISTRAL AI_ · Grok · 文心一言 · NOMIC

Vector Databases

Milvus · Pinecone · Chroma · Qdrant · LanceDB

Synthetic training data

OpenSynthetics · gretel · datacebo Copulas · SYNTHEA · Sd

Configuration & tooling

Langflow · Tasking.AI · Open WebUI · composio · PrivateGPT · talkd.ai

Libraries & Frameworks

LangChain · LlamaIndex · Kubeflow · Ollama · LLaMA-Factory Easy and Efficient LLM Fine-Tuning · OpenCV · Keras · TensorFlow · PyTorch

Open Enterprise AI Infrastructure

SUSELOGO

SUSE AI

SUSE

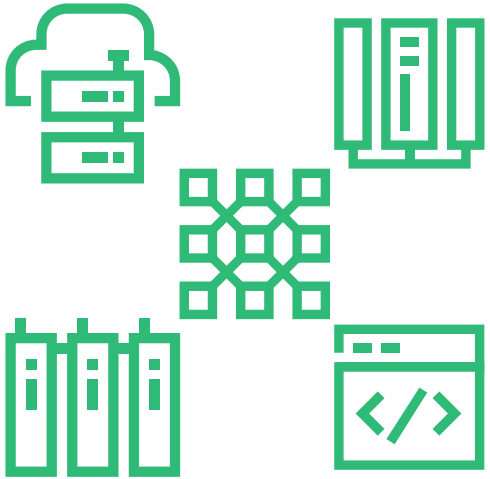
Challenge #4: 엣지 인프라와 결합

AI 배포는 복잡한 결합 동반:

→ 엣지 하드웨어와 드라이버 결합

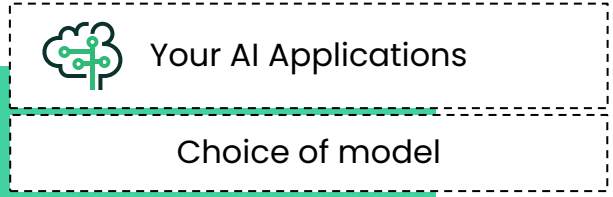
→ 다양한 소프트웨어 기술로 구성된 빠르게 움직이는 얽힌 인프라 스택

… 이를 작동하게 하고, 유지하는 것이 고객사의 부담이 되지 않도록 해야 합니다.



SUSE AI

GenAI 워크로드를 배포하고 운영하기 위한 클라우드 네이티브 플랫폼



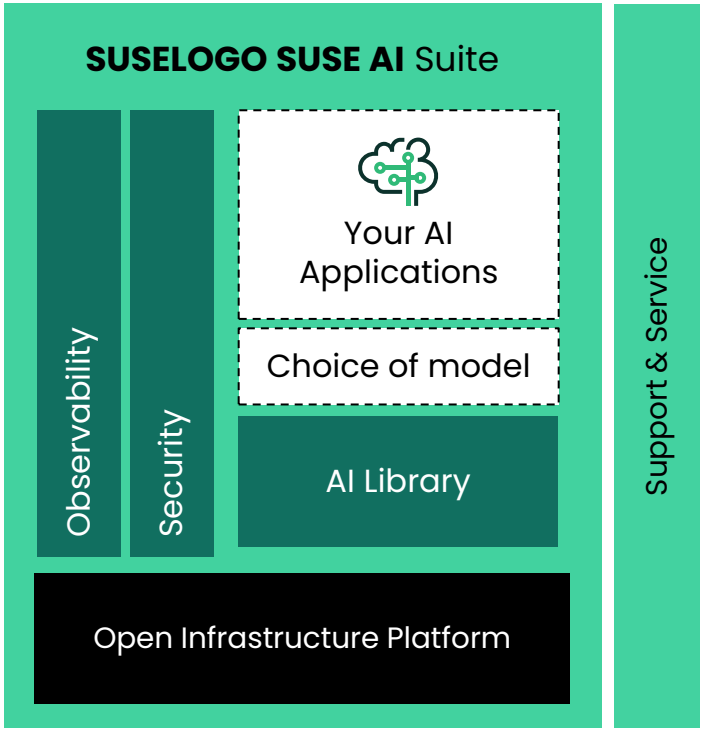
SUSELOGO SUSE AI

- Sanitized AI Components
- Observability Dashboard
  - GPU metrics
  - Workload insights
  - Application Health
- Security
  - AI Compliance/Governance
  - Zero-trust protection
  - Life-cycle coverage: development, deployment, to operations
- Open Infrastructure Platform

SUSE

SUSE AI: What's in the box?

- → 생성형 AI를 위한 확장 가능한 기반
  - → 검색 증강 생성(RAG) (RAG)
- → SUSE의 안전한 공급망에서 구축 및 배포
  - → Rancher Prime Application Collection을 통해 환경에 쉽게 제공
- → FIPS 및 Common Criteria 인증을 받은 SUSE의 Open Infrastructure Platform(SL Micro, SUSE Rancher)에서 실행
  - → AI 라이브러리 인증 진행 중
- → AI 사용 사례에 최적화된 통합 플랫폼
  - → Security, Governance, and Observability



Benefits of SUSE AI



Choice

모든 모델 실행

인프라와 통합하기 위한 모듈식 및

확장 가능

벤더 락인 없음

총 비용 제어 및 가시성

구축 옵션에는 클라우드, 하이브리드, 온프레미스가 포함



Security

보안 공급망으로 구축됨; 인증 상속

제로 트러스트 보안

완전한 관찰 가능성

데이터 주권

에어 갭 솔루션으로 배포 가능.



Trust

데이터는 엔터프라이즈 제어 범위 내에 존재

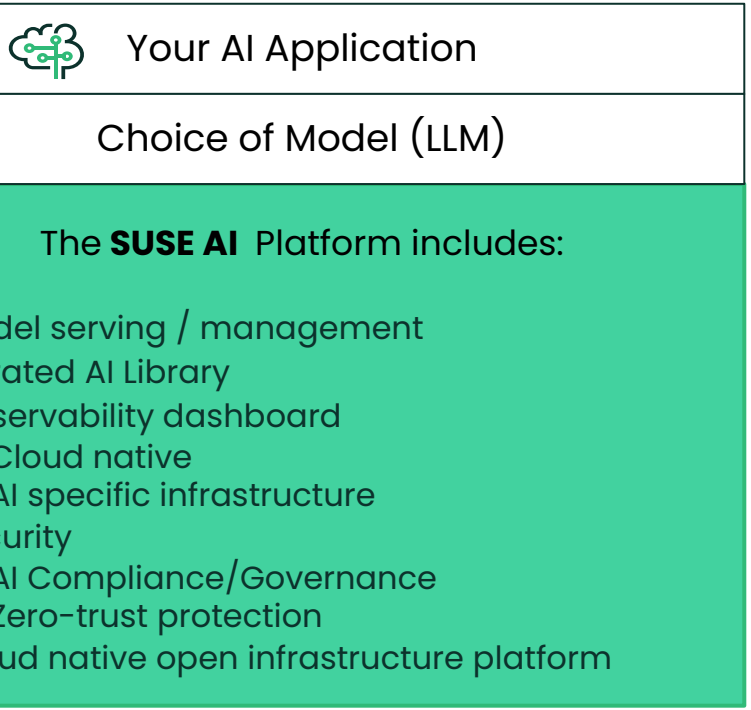
사용 및 활동 제어

생성된 출력에 대한 더 큰 확신

SUSE

SUSE AI - Platform Features

- 모든 AI 애플리케이션을 위한 안전하고 일관된 배포 및 런타임 플랫폼
- SUSE Security 및 SUSE Observability와 완벽하게 통합
- SUSE 보안 공급망을 사용하여 구축된 배포하기 쉬운 큐레이션된 AI 도구 및 구성 요소
- 투자를 미래에 대비할 수 있도록 확장 가능하고 모듈식
- LLM 및 배포 방법 선택 가능



SUSE

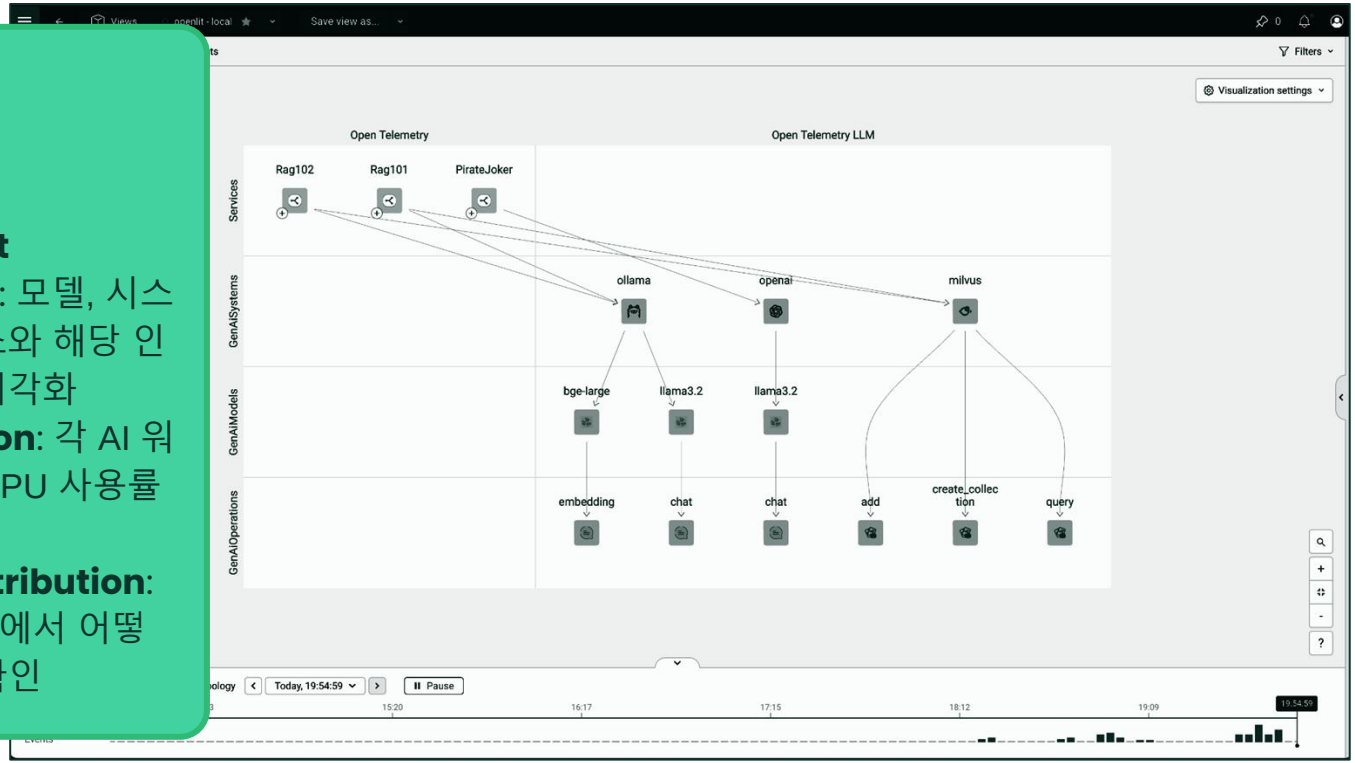
SUSE Observability Provides Insight Across Infrastructure

Understand AI Infrastructure

→ **AI Component dependencies**: 모델, 시스템, 데이터베이스와 해당 인프라 종속성을 시각화

→ **GPU Correlation**: 각 AI 워크로드에 대한 GPU 사용률을 추적

→ **Workload Distribution**: 워크로드가 어디에서 어떻게 실행되는지 확인

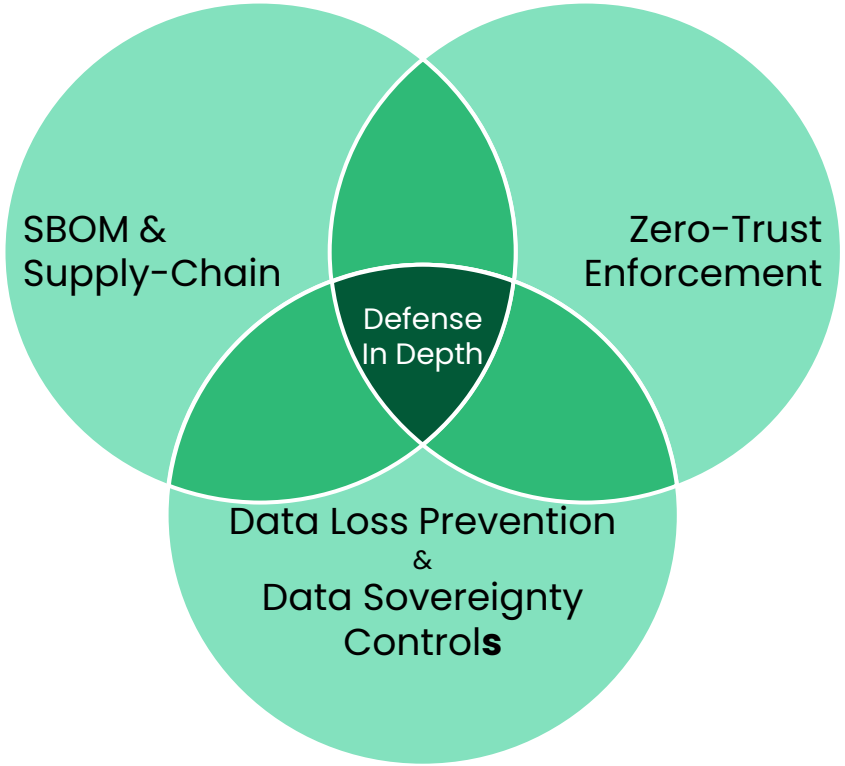


SUSE AI is Secure by Design

검증된 SUSE 솔루션과 사례를 사용하여 AI 워크로드를 위한 안정적이고 지속 가능한 프레임워크 구축.

- → AI Compliance/Governance
- → Zero-trust protection
- → Life-cycle coverage: development, deployment, to operations

Examples
EU AI 컴플라이언스에 매우 익숙한 통제구조가 적용되는 ollama.com의 모델들

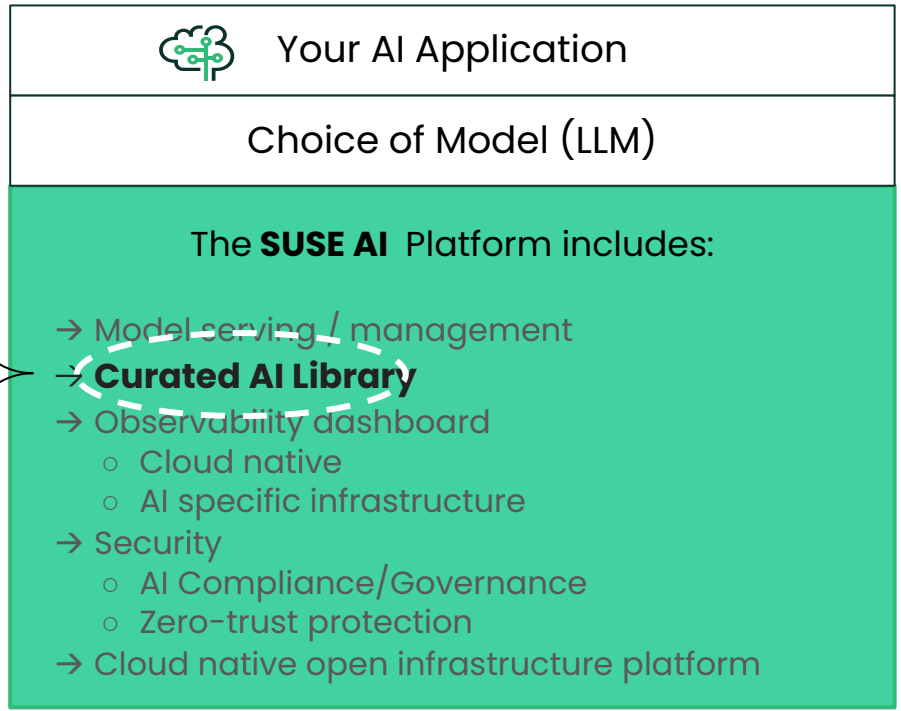


Day 0 Create → Day 1 Deploy → Day 2 Operate

SUSE

SUSE AI - The Curated AI Library

- **Ollama:** LLM 관리 및 배포
  - 로컬 설치 / 프라이빗 중심
  - 폭넓은 지원 / 사용 편의성
- **Milvus:** 방대한 데이터 세트에 대한 유사성 검색에 맞춰 제작된 강력한 벡터 데이터베이스
  - AI에서 RAG 사용에 이상적
  - 클라우드 네이티브 확장 지원
- **Open WebUI:** 확장 가능하고 기능이 풍부한 자체 호스팅 AI 플랫폼
  - RAG용 내장 추론 엔진
  - Ollama 및 OpenAI API와의 원활한 통합



SUSE

SUSE AI Suite SKU

Product	SUSE AI Suite
SUSE Rancher Prime Application Collection Advanced	✓
SUSE Linux Micro	✓
RKE/K3s	✓
Rancher Management Server	✓
SUSE Observability	✓
SUSE Rancher Prime Application Collection Base	✓
SUSE Observability Platform Optimization	✓
SUSE Storage	✓
SUSE Security	✓
SUSE Virtualization (HCI)	✓

SUSE AI에는 SUSE Rancher Suite의 모든 제품과 AI 소프트웨어 컨테이너와 헬름 차트에 대한 L1, L2, L3 지원을 제공하는 Rancher Prime Application Collection Advanced가 포함되어 있습니다.

Pricing Metric Options

Option #1
vCPU Metric

2 Cores or 4 vCPUs (hyperthreading)

워커 노드에 할당된 코어 /vCPU의 총 수를 계산

Option #2
Socket Metric

2 Sockets up to 64 Cores

각 베어 메탈 호스트에 적용

SUSE

Tata Consultancy Services

Customer Pain

TCS는 고객의 AI 애플리케이션을 위한 안전하고 유연한 기반이 필요했습니다.

Role of SUSE AI

SUSE AI는 다양한 AI 사용 사례를 호스팅하기 위한 클라우드 기반 인프라를 제공합니다.

Use Cases

TCS의 AI 솔루션 실무를 위한 플랫폼

> SUSE AI 플랫폼의 GenAI 기능은 오늘날의 규제 환경에서 필수적인 보안과 주권을 유지하는 데 도움이 됩니다.

Nidhi Srivastava
Global Head of AI.Cloud Offerings
TCS



FUJITSU

Unlocking the Power of Open Source: Co-Creating Innovative AI Solutions

Headstart on AI adoption

SUSE Rancher Prime 및 SUSE Virtualization을 기반으로 한 AI 테스트 드라이브 이니셔티브

Turnkey Private GPT solution

100% 데이터 주권 및 보호

85+% AI Use Cases Covered

이미 존재하는 솔루션 모듈을 사용하여 즉시 사용 가능

> 우리의 협력적 접근 방식은 고객이 AI에서 앞서 나갈 수 있는 솔루션을 만드는 데 도움이 되었습니다. 오픈 소스의 잠재력을 활용하여 고객의 요구를 진정으로 충족하는 AI 솔루션을 만들 수 있습니다.

Udo Würtz
Chief Data Officer

FUJITSU



SUSE



SUSE

Learn more,
Book a demo,
Work with our partners

Thank you!


SUSE