



**KI &  
Modern Work**

# CompanyAI Agent onPrem

Ihre KI auf Ihrer Hardware

**Der CompanyAI Agent onPrem von Medialine bietet eine skalierbare und zukunftssichere On-Premise-KI-Plattform, die nahtlose Integration von generativer KI in Ihre Unternehmensinfrastruktur ermöglicht. Durch seine flexible Architektur und umfassende Funktionen wird eine individuelle Anpassung an Ihre Anforderungen gewährleistet – alles auf Ihrer eigenen Hardware vor Ort.**

## Skalierbare und erweiterbare KI-Plattform

Der **CompanyAI Agent onPrem** ist darauf ausgelegt, die neueste KI-Technologie in Ihr Unternehmen zu bringen:

**Souveräne KI-Plattform:** Lokale Installation auf Ihrer Hardware, garantiert datenschutzkonforme KI-Nutzung nach deutschen und europäischen Standards.

**Flexible KI-Integration:** Anpassbar an vorhandene Technologien und KI-Modelle, sowohl Open-Source als auch Ihre eigens entwickelten.

**Zukunftssicher:** Die Plattform ist skalierbar und kann an wachsende Anforderungen und neue KI-Anwendungsfälle angepasst werden.

## Maßgeschneiderte KI-Lösungen

Sie erhalten Ihren eigenen, personalisierten CompanyAI-Agent, der...

**...natürliche Sprache versteht und generiert:** In Text, Sprache, Bild und optional auch Video.

**...Multi-Input und Multi-Output Optionen bietet:** Flexible Schnittstellen für unterschiedliche Datenquellen und -formate.

**...Individualisierung bereitstellt:** Passen Sie den KI-Agenten an Ihre Compliance-Richtlinien, ethischen Standards und spezifische Terminologien an.

**...effizientes Training ermöglicht:** Dank Multi-Tenancy können spezifische Anwendungsfälle separat trainiert werden, was Kosten und Zeit für das Training neuer Modelle reduziert.

## Datenschutz und Sicherheit

Mit dem **CompanyAI Agent onPrem** erfüllen Sie nicht nur die Anforderungen des **EU AI Acts**, sondern auch die deutschen und europäischen Datenschutzvorgaben. Ihre KI-Lösungen bleiben sicher in Ihrer eigenen IT-Infrastruktur, ohne Daten an externe Cloud-Dienstleister auszulagern.

## Zukunftssicher und wirtschaftlich

**Auto-Scaling und Kostenkontrolle:** Der präzise Einsatz von Ressourcen und eine skalierbare Architektur ermöglichen eine optimale Kostenkontrolle, ohne auf Leistung verzichten zu müssen.

**CI/CD-Integration:** Einfach automatisierbare Abläufe für die Integration und Bereitstellung Ihrer KI-Modelle über die Cloud-API.

## Leistungsstarke Erweiterungen und Monitoring

Durch die Integration in ein umfassendes Framework bietet die Plattform zahlreiche nützliche Features:

**Umfangreiche Überwachungs- und Monitoring-systeme:** Minimieren Sie Halluzinationen und maximieren Sie die Leistungsfähigkeit auf Basis Ihrer personalisierten Trainingsdaten.

**Erweiterbare Funktionen:** Mit der Unterstützung von Medialine können Sie die Plattform kontinuierlich um neue Modelle, Integrationen und Tools erweitern. Beispielsweise können Sie mit unserer AI-Agent befeuerten IPaaS-Integration komplexe KI-Workflow-Automatisierung inklusive Daten/Datenbanken-Abfragen unterschiedlicher Systeme kreieren.

## Highlight: Unser Compliance Wächter im CompanyAI Agent

Der CompanyAI Agent von Medialine integriert einen hochentwickelten Compliance Wächter, das Prompt-Anfragen überprüft und mit Daten aus der RAG-Pipeline oder Vektordatenbank anreichert. Es stellt sicher, dass nur berechtigte Nutzer auf bestimmte Dateien zugreifen können und prüft, ob die Datenverarbeitung in externen LLMs aus Azure EU mit OpenAI-, Google-, und Anthropic-Modellen den Datenschutzrichtlinien entspricht. Bei Verstößen gegen die Compliance-Richtlinien wird auf sichere, souveräne Open-Source-KI-Modelle wie neuste Medialine-NIM-Modellen oder Fujitsu Sauerkraut LLM verwiesen.

Der Compliance Wächter kann zudem die Prompt-Eingaben optimieren und berücksichtigt Unternehmensstandards, um eine professionelle und benutzerfreundliche Nutzung zu gewährleisten. Dies ist besonders wichtig für Open-Source-Modelle wie von Meta oder NVIDIA, die umfangreichere Prompts für eine optimale Verarbeitung benötigen. Dadurch werden die Effizienz und Usability für die Nutzer erheblich gesteigert.

	<b>Paket XXS</b> KI Plattform	<b>Paket XS</b> LLM 8x7B	<b>Paket S</b> KI Plattform + LLM 8x7B	<b>Paket M</b> KI Plattform + LLM 72B
	ab <b>85.000,-€</b> einmalig* <b>1.500,-€</b> /pro Monat	ab <b>165.000,-€</b> einmalig*	ab <b>170.000,-€</b> einmalig* <b>1.500,-€</b> /pro Monat	ab <b>400.000,-€</b> einmalig* <b>1.500,-€</b> /pro Monat
1. Setup	Standard	Standard	Standard	Standard
2. Support**	5 Jahre	5 Jahre	5 Jahre	5 Jahre
3. CompanyAI Agent	✓	✗	✓	✓
4. Anzahl User***	ab 128	ab 128	ab 128	ab 128
5. Private GPT	✗	✓	✓	✓
5.1 OnPrem Modelle	✗	Sauerkraut 8x7B	Sauerkraut 8x7B	NVLM-D 1.0 72B (FP16)
5.2 Cloud Modelle	Alle gängigen Modelle mit verfügbarer Schnittstelle	✗	Alle gängigen Modelle mit verfügbarer Schnittstelle	Alle gängigen Modelle mit verfügbarer Schnittstelle
5.3 Entwicklungsumgebung	✗	✗	✗	Nvidia NIM-Modelle inkludiert
6. Anzahl Server	3 (1 Master, 2 Worker)	1 LLM	3 CompanyAI 1 LLM Server	3 CompanyAI 1 LLM Server
7. GPU (Grafikkarte)	✗	1 Server mit 1x L40s	1 Server mit 1x L40s	1 Server mit 2x H100
8. Server-management****	✗	✗	✗	✗
9. Block Storage*****	✗	✗	✗	✗

\*Zzgl. Servermanagement, zzgl. Storage +50.000,- € je 128 weiterer User zeitgleich \*\*Support je nach LLM nach Rücksprache \*\*\*50 Shared Logins inkludiert, weitere Logins nach Rücksprache buchbar \*\*\*\*Servermanagement nach Rücksprache buchbar \*\*\*\*\*Storage nach Rücksprache buchbar

NVLM ist ein von Nvidia entwickeltes LLM, welches den Benchmarks zu Folge auf Augenhöhe mit GPT-4o latest und Llama 3.x liegt. Sauerkraut (SauerkrautLM) ist ein auf Mixtral-8x7B basierendes LLM welches von VAGO solutions entwickelt wurde und speziell auf den deutschen Markt zugeschnitten ist.

	<b>Paket L</b> KI Plattform HA + LLM 8x7B	<b>Paket XL</b> KI Plattform HA + LLM 72B	<b>Paket XXL</b> KI Plattform HA + LLM 72B HA	<b>Paket XXXL</b> KI Plattform HA + LLM 8x/B HA + LLM 72B
	ab <b>400.000,-€</b> einmalig* <b>1.500,-€</b> /pro Monat	ab <b>470.000,-€</b> einmalig* <b>1.500,-€</b> /pro Monat	ab <b>540.000,-€</b> einmalig* <b>1.500,-€</b> /pro Monat	ab <b>800.000,-€</b> einmalig* <b>5.000,-€</b> /pro Monat
<b>1. Setup</b>	Standard	Standard	Standard	Individuell
<b>2. Support**</b>	5 Jahre	5 Jahre	5 Jahre	5 Jahre
<b>3. CompanyAI Agent</b>	✓ HA	✓ HA	✓ HA	✓ HA
<b>4. Anzahl User***</b>	ab 128	ab 128	ab 128	ab 128
<b>5. Private GPT</b>	✓	✓	✓	✓
<b>5.1 OnPrem Modelle</b>	Sauerkraut 8x7B	NVLM-D 1.0 72B (FP16)	NVLM-D 1.0 72B (FP16, HA)	Sauerkraut 8x7B (HA) NVLM-D 1.0 72B (FP32)
<b>5.2 Cloud Modelle</b>	Alle gängigen Modelle mit verfügbarer Schnittstelle	Alle gängigen Modelle mit verfügbarer Schnittstelle	Alle gängigen Modelle mit verfügbarer Schnittstelle	Alle gängigen Modelle mit verfügbarer Schnittstelle
<b>5.3 Entwicklungsumgebung</b>	Nvidia NIM-Modelle inkludiert	Nvidia NIM-Modelle inkludiert	Nvidia NIM-Modelle inkludiert	Nvidia NIM-Modelle inkludiert
<b>6. Anzahl Server</b>	6 CompanyAI 1 LLM Server	6 CompanyAI 1 LLM Server	6 CompanyAI 2 LLM Server	6 CompanyAI 3 LLM Server
<b>7. GPU (Grafikkarte)</b>	1 Server mit 1x L40s	1 Server mit 2x H100	2 Server mit 2x H100	1 Server mit 4x H100 2 Server mit je 4x L40s
<b>8. Server-management****</b>	✗	✗	✗	✓
<b>9. Block Storage*****</b>	✗	✗	✗	✓

\*Zzgl. Servermanagement, zzgl. Storage +50.000,- € je 128 weiterer User zeitgleich \*\*Support je nach LLM nach Rücksprache \*\*\*50 Shared Logins inkludiert, weitere Logins nach Rücksprache buchbar \*\*\*\*Servermanagement nach Rücksprache buchbar \*\*\*\*\*Storage nach Rücksprache buchbar

NVLM ist ein von Nvidia entwickeltes LLM, welches den Benchmarks zu Folge auf Augenhöhe mit GPT-4o latest und Llama 3.x liegt. Sauerkraut (SauerkrautLM) ist ein auf Mixtral-8x7B basierendes LLM welches von VAGO solutions entwickelt wurde und speziell auf den deutschen Markt zugeschnitten ist.

## Ihr Trusted Advisor für On-Premises KI-Lösungen

Medialine bietet Ihnen mehr als nur Technologie – wir begleiten Sie als Partner bei der Implementierung und kontinuierlichen Weiterentwicklung Ihrer KI-Anwendungen:

**Individuelle Beratung:** Maßgeschneiderte Lösungen, die sich an Ihre spezifischen Anforderungen und Ziele anpassen.

**Langfristige Unterstützung:** Flexibilität und Erweiterbarkeit für künftige Projekte und Anpassungen.

**Sicherheit und Vertrauen:** Unsere Lösungen sind zukunftssicher, datenschutzkonform und bieten eine solide Grundlage für Ihre KI-Strategie.

Kontaktieren Sie uns unter [sales@medialine.ag](mailto:sales@medialine.ag) um mehr über den CompanyAI Agent zu erfahren. Vertrauen Sie auf unsere Expertise und lassen Sie uns gemeinsam Ihre KI-Initiative zum Erfolg bringen.

Änderungen und Irrtümer vorbehalten. Es gelten unsere allgemeinen Geschäftsbedingungen in der jeweils aktuellen Fassung. Die Produktbeschreibung stellt noch kein verbindliches Angebot dar und dient ausschließlich der Information. Vertragsdetails sind aus Angeboten und Leistungsverzeichnissen zu entnehmen, welche wir gerne für Sie erstellen. Stand: 01/2025