



AI &
Modern Work

CompanyAI Agent onPrem

Your AI on your hardware

Medialine's **CompanyAI Agent onPrem** provides a scalable and future-proof on-premises AI platform that enables seamless integration of generative AI into your enterprise infrastructure. Its flexible architecture and comprehensive features ensure customization to your requirements - all on your own on-premises hardware.

Scalable and extensible AI platform

The **CompanyAI Agent onPrem** is designed to bring the latest AI technology to your company:

Sovereign AI platform: Local installation on your hardware, guaranteed data protection-compliant AI use in accordance with German and European standards.

Flexible AI Integration: Adaptable to existing technologies and technologies and AI models, both open source and your own as well as your own in-house developments.

Future-proof: The platform is scalable and can be adapted to growing requirements and new AI use cases.

Customized AI solutions

You get your own personalized CompanyAI agent that...

...understands and generates natural language: In Text, speech, image and optional video.

...offers multi-input and multi-output options: Flexible interfaces for different data sources and formats.

...provides individualization: Customize the AI agent to your compliance guidelines, ethical standards and specific standards and specific terminologies.

...enables efficient training: Thanks to Multi-Tenancy, specific use cases can be trained separately use cases can be trained separately, reducing the cost and time for training new models.

Data protection and security

With the **CompanyAI Agent onPrem**, you not only meet the requirements of the **EU AI Act**, but also the German and European data protection regulations. Your AI solutions remain securely in your own IT infrastructure without outsourcing data to external cloud service providers.

Future-proof and economical

Auto-scaling and cost control: The precise use of use of resources and a scalable architecture enable optimal cost control without having to compromise on performance.

CI/CD integration: Easily automated processes for the integration and deployment of your AI models via the cloud API.

Powerful extensions and monitoring

Thanks to the integration into a comprehensive framework the platform offers numerous useful features:

Extensive monitoring and surveillance systems:

Minimize hallucinations and maximize performance based on your personalized training data.

Extensible functions: With support from Medialine, you can continuously expand the platform with new models, integrations and tools. For example, with our AI agent-powered IPaaS integration, you can create complex AI workflow automation including data/database queries from different systems.

Highlight: Our Compliance Guardian in the CompanyAI Agent

Medialine's CompanyAI Agent integrates a sophisticated compliance Guardian that checks prompt requests and enriches them with data from the RAG pipeline or vector database. It ensures that only authorized users can access certain files and checks whether data processing in external LLMs from Azure EU with OpenAI, Google and Anthropic models complies with data protection guidelines. If compliance guidelines are breached, it refers to secure, sovereign open source AI models such as the newest Medialine-NIM models or Fujitsu Sauerkraut LLM.

The Compliance Guardian can also optimize prompt inputs and takes corporate standards into account to ensure professional and user-friendly use. This is particularly important for open source models such as Meta or NVIDIA, which require more extensive prompts for optimal processing. This significantly increases efficiency and usability for users.

	Package XXS AI Platform	Package XS LLM 8x7B	Package S AI Platform + LLM 8x7B	Package M AI Platform + LLM 72B
	starting from €85.000 one time* €1.500 /per month	starting from €165.000 one time*	starting from €170.000 one time* €1.500 /per month	starting from €400.000 one time* €1.500 /per month
1. Setup	Standard	Standard	Standard	Standard
2. Support**	5 years	5 years	5 years	5 years
3. CompanyAI Agent	✓	✗	✓	✓
4. Number of users***	starting from 128	starting from 128	starting from 128	starting from 128
5. Private GPT	✗	✓	✓	✓
5.1 OnPrem Models	✗	Sauerkraut 8x7B	Sauerkraut 8x7B	NVLM-D 1.0 72B (FP16)
5.2 Cloud Models	All current models with available interface	✗	All current models with available interface	All current models with available interface
5.3 Development environment	✗	✗	✗	Nvidia NIM-Modelle included
6. Number of servers	3 (1 Master, 2 Worker)	1 LLM	3 CompanyAI 1 LLM Server	3 CompanyAI 1 LLM Server
7. GPU (graphic card)	✗	1 server with 1x L40s	1 server with 1x L40s	1 server with 2x H100
8. Server management****	✗	✗	✗	✗
9. Block storage*****	✗	✗	✗	✗

*Plus server management, plus storage + €50,000 per 128 additional simultaneous users **Support depending on LLM after consultation ***50 shared logins included, further logins bookable after consultation ****Server management bookable after consultation *****Storage bookable after consultation

NVLM is an LLM developed by Nvidia which, according to the benchmarks, is on a par with GPT-4o latest and Llama 3.x. Sauerkraut (SauerkrautLLM) is an LLM based on Mixtral-8x7B which was developed by VAGO solutions and is specially tailored to the German market.

	Package L AI Platform HA + LLM 8x7B	Package XL AI Platform HA + LLM 72B	Package XXL AI Platform HA + LLM 72B HA	Package XXXL AI Platform HA + LLM 8x7B HA + LLM 72B
	starting from €400.000 one time* €1.500 /per month	starting from €470.000 one time* €1.500 /per month	starting from €540.000 one time* €1.500 /per month	starting from €800.000 one time* €5.000 /per month
1. Setup	Standard	Standard	Standard	Individual
2. Support**	5 years	5 years	5 years	5 years
3. CompanyAI Agent	✔ HA	✔ HA	✔ HA	✔ HA
4. Number of users***	starting from 128	starting from 128	starting from 128	starting from 128
5. Private GPT	✔	✔	✔	✔
5.1 OnPrem Models	Sauerkraut 8x7B	NVLM-D 1.0 72B (FP16)	NVLM-D 1.0 72B (FP16, HA)	Sauerkraut 8x7B (HA) NVLM-D 1.0 72B (FP32)
5.2 Cloud Models	All current models with available interface	All current models with available interface	All current models with available interface	All current models with available interface
5.3 Development environment	Nvidia NIM-Modelle included	Nvidia NIM-Modelle included	Nvidia NIM-Modelle included	Nvidia NIM-Modelle included
6. Number of servers	6 CompanyAI 1 LLM Server	6 CompanyAI 1 LLM Server	6 CompanyAI 2 LLM Server	6 CompanyAI 3 LLM Server
7. GPU (graphic card)	1 server with 1x L40s	1 server with 2x H100	2 servers with 2x H100	1 server with 4x H100 2 servers with 4x L40s each
8. Server management****	✘	✘	✘	✔
9. Block storage*****	✘	✘	✘	✔

*Plus server management, plus storage + €50,000 per 128 additional simultaneous users **Support depending on LLM after consultation ***50 shared logins included, further logins bookable after consultation ****Server management bookable after consultation *****Storage bookable after consultation

NVLM is an LLM developed by Nvidia which, according to the benchmarks, is on a par with GPT-4o latest and Llama 3.x. Sauerkraut (SauerkrautLLM) is an LLM based on Mixtral-8x7B which was developed by VAGO solutions and is specially tailored to the German market.

Your trusted advisor for on-premises AI solutions

Medialine offers you more than just technology - we accompany you as a partner in the implementation and continuous development of your AI applications:

Individual consulting: Tailor-made solutions that adapt to your specific requirements and goals.

Long-term support: Flexibility and expandability for future projects and adaptations.

Security and trust: Our solutions are future-proof, data protection-compliant and provide a solid foundation for your AI strategy.

Contact us at sales@medialine.com to find out more about the CompanyAI Agent. Put your trust in our expertise and let us us together to make your AI initiative a success.

Subject to change and errors. Our general terms and conditions apply in the current version. The product description does not constitute a binding offer and is for informational purposes only. Contractual details can be found in our offers and service catalogs, which we would be happy to create for you. as of: 01/2025