XITASO



Shaping the future

AI-supported knowledge management

Optimizing access to knowledge with AI

Whitepaper – Artificial intelligence

Content

Executive Summary 04

1. Knowledge management in times of shortage of skilled labor 04
2. Current state of the industry 05
3. Access to organizational knowledge 06
   - 3.1 Retrieval Augmented Generation 06
   - 3.2 Challenges 07
     - 3.2.1 Best Practices 08
     - 3.2.2 Advanced RAG 09
     - 3.2.3 Users in focus 10
   - 3.3 Business impact 11
4. Current developments and new opportunities through Agentic AI 12
5. Customer references 13
   - 5.1 MULTIVAC 13
   - 5.2 AI- Chatbot telli 13
6. With XITASO to project success 14

© XITASO GmbH | info@xitaso.com | +49 821 885882-0

© XITASO GmbH | info@xitaso.com | +49 821 885882-0

Executive Summary

This white paper shows how important artificial intelligence (AI) is for knowledge management in industry and explains how it can increase productivity by up to 12.5%. We illustrate that knowledge should be used efficiently to retain it in a shrinking workforce and to transfer it to new employees. AI systems are introduced as helpful tools to accomplish these tasks, although they are not yet directly suited for industrial use. Technical improvements need to be implemented to increase their benefits. It is also important to develop solutions that are easily accessible and scalable to move beyond prototype status. A key finding is that the success of AI systems depends heavily on how well users are involved. This requires a cooperative attitude to overcome fears of job loss. Good user experience and accompanying training can greatly improve employees‘ skills in AI literacy and thus contribute to increasing productivity.

1. Knowledge management in times of shortage of skilled labor

Efficient knowledge management is becoming increasingly relevant for the following reasons:

1. The amount of data that surrounds us is exploding - within the last 9 years, it has grown almost exponentially from 20 zettabytes (ZB) to 225 ZB (Hassani & MacFeely, 2023) - to put this into perspective, one ZB is equivalent to one billion terabytes. Knowledge in society is alwso continuously increasing. This is indicated by the Knowledge Quantification Index on scientific publications, which suggests that it is increasing linearly (Kang et al., 2024).
2. Industry, especially the German one, is characterized by a persistent shortage of skilled workers - the skilled worker index, an indicator that measures the demand for skilled workers, is at a consistently high level (Rudnicka, 2025). Similarly, the proportion of over-60s in the workforce has been rising for years (Michelsen & Junker, 2024), which indicates that the situation will worsen significantly in the foreseeable future. Work preferences are also changing - the length of employment with the current employer is declining significantly (Destatis, 2025). As a result of these aspects, companies have fewer and fewer experienced experts who have acquired large amounts of implicit knowledge through many years of practical experience.

To summarize: We have to use knowledge more efficiently – more knowledge has to be available more quickly to allow employees and companies to continue to be as productive as possible in the future. Years of training and onboarding times, as are currently widespread, will no longer be competitive. Through proactive and efficient knowledge management, companies will be superior in terms of costs and quality to their competitors who are confronted with similar problems.

© XITASO GmbH | info@xitaso.com | +49 821 885882-0

2. Current state of the industry

But what is the current state in companies? Organizational knowledge, i.e. knowledge about products, processes, manufacturing processes, etc. ensures that a company can operate competitively. However, it is often distributed across storage locations and data formats or only available in the heads of employees. Over time, organizational knowledge also grows organically. These aspects mean that finding the required knowledge is anything but efficient: for example, 95% of service staff search for information; 86% of staff need more than 30 minutes per day to do so, with 12% needing more than 2 hours (Kothes GmbH, 2023).

The use of AI in knowledge management and information searches has been an obvious choice at the latest since the extensive spread of ChatGPT. It is therefore an ideal field in which the industry‘s desire to play a pioneering role in the use of AI (79%) can be realized without significant investment in infrastructure. Thus, the assessment that AI is crucial for competitiveness (82%) (Höbbel & Spohr, 2025) can be confirmed. This white paper analyzes the potential of AI in knowledge management - especially in accessing and searching for information. It reports on challenges and best practices from industrial applications and highlights current emerging developments.



3. Access to organizational knowledge

With the spread of ChatGPT at the end of 2022, the major advances of AI in dealing with natural language information, which were already foreseeable through advances in machine translation, became known to a wider audience. In particular, natural-language interaction using chatbots with the knowledge base implicitly present in neural networks represented a paradigm shift from previous search engines. In an industrial context, data protection-compliant hosting and assurances that the data entrusted to the AI will not be used for training and will therefore remain with the company. This was promptly addressed by corresponding services of the hyperscalers (AWS, Azure, Google Cloud), in theory enabling deployment. In practice, however, questions quickly arose such as:

- **Where does new knowledge come from that did not yet exist at the time the AI was trained?**
- **How does the AI gain access to organization-specific knowledge that was not part of the training data set?**
- **How to prevent frequent hallucination, i.e. the „invention“ of information unbacked by truthful content, likely caused by no corresponding knowledge in the training data set?**

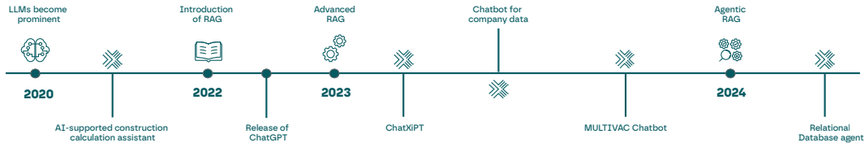


Figure 1: AI development & AI at XITASO

3.1 Retrieval Augmented Generation

RAG consists of 3 essential aspects: the ingest of knowledge in a central knowledge base, the retrieval of knowledge segments or documents relevant to the query and the generation of the desired answer based on the relevant knowledge segments.

During ingest, the various data sources in which organizational knowledge is available (sharepoint, ERP, wiki, etc.) are divided into manageable segments (chunking), then converted into semantically meaningful vectors (embedding) and finally stored in a vector database that functions as a knowledge base.

For example, Figure 2 shows the interactions taking place withing the system for the query „How can I maintain the forming tool?“. In order to find relevant documents, the query is also transferred to the vector space (embedded) and similar knowledge segments or documents (e.g. excerpts from operating manuals or process instructions) are identified and returned based on their similarity in the vector space. This and the original query are then fed to a suitable language model (e.g. GPT-4o) to generate a suitable response (e.g. „You have to dismantle the lower part and clean it by hand“).

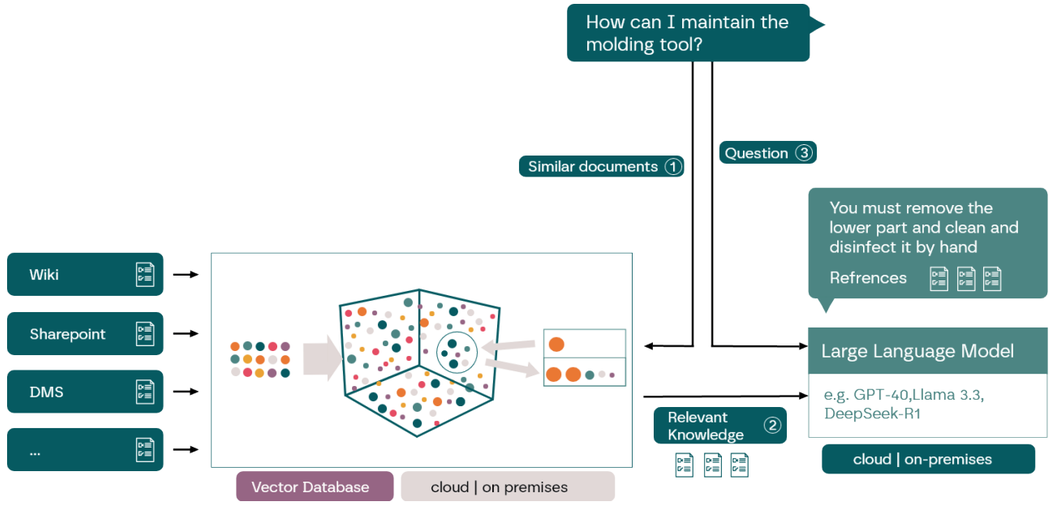


Figure 2: Chatbots for organizational knowledge

3.2 Challenges

One of the biggest challenges of AI in knowledge management and especially with RAG-supported chatbots is the initial inaccuracy of their answers. This response quality can be described more precisely using more granular metrics such as contextual relevancy or correctness of generated answer. With the „off-the-shelf“ approach, however, experience has shown that only 60-70% correctness can be expected according to the metrics mentioned. It is therefore unclear to the user whether the information presented is correct and complete and they have to undertake a time-consuming in-depth check to double check. For service applications, this is at the very least a capacity constraint. In the case of solutions that are intended for end users, incorrect information can lead to a loss of trust or even recourse claims, which in the worst case can also have a negative impact on the brand perception.

There is also the risk of using the chatbot for unforeseen purposes through so-called jailbreaks: Through tricks and special formulati-

ons, a chatbot can be manipulated into revealing things that it should not relay according to specifications. This can lead to sensitive information, such as internal company data, being disclosed.

The „off-the-shelf“ use of the technology is therefore not sufficient for industrial use, even though one often hears the contrary. Therefore, measures have to be taken on two levels when implementing AI solutions: On the technical level, the accuracy of the answers has to be ensured and misuse of the technology has to be prevented. At the user level, users have to be guided in their interaction with the system to ensure they sufficiently understand the limits and possibilities of AI systems. This allows the users to integrate AI easily into their daily workflows.

3.2.1 Best practices



Juan Pablo Bottaro

Principal Staff Software Engineer, LinkedIn
(Bottaro, 2024)

> “The team achieved 80% of the basic experience we were aiming to provide within the first month and then spent an additional four months attempting to surpass 95%”

Best practices can be used to control the challenges described above and enable the use of AI systems in an industrial context. First, we focus on the technical dimension and then detail considerations at the user level.

3.2.2 Advanced RAG



Figure 3: Naive RAG vs. extended RAG

Technical extensions to individual components of a RAG system are grouped under the umbrella term Advanced RAG (Gao et al., 2023), which can achieve accuracies of over 95% and thus raise the system to human level – with significantly less time required. Exemplary extensions are

- **the combination of vector & full-text search to give greater weight to important technical terms**
- **Smart chunking, which allows meta information or document hierarchy to be taken into account,**
- **Reranking of retrieved segments based on relevance defined by LLMs,**
- **Groundedness detection, which ensures all aspects of the answer are based on the referenced knowledge segments.**

Some others are listed in Figure 4 in red labels.

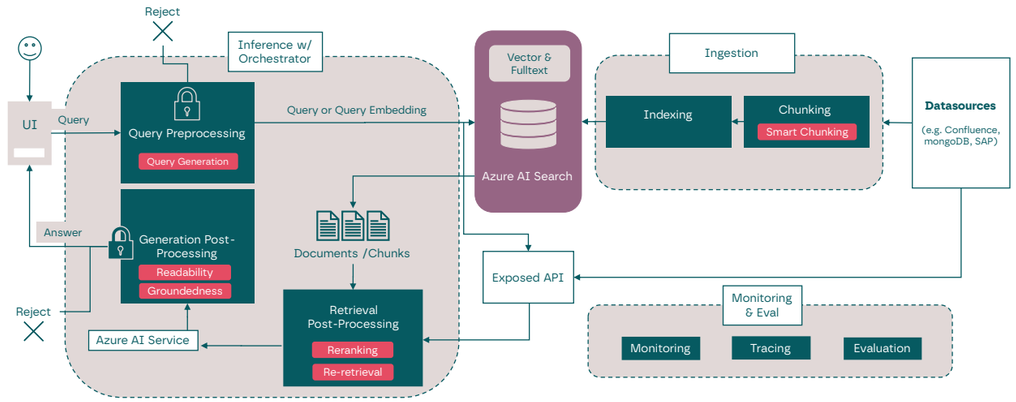


Figure 4: Advanced RAG Architekturskizze

Each of these methodological enhancements is a building block for increasing the accuracy of RAG-based AI systems. Their use should be made consciously, considering the respective limitations of the current state of development and the characteristics of the respective organizational knowledge. Thus, it can be ensured that only features are implemented which actually lead to a directly measurable increase in value.

In addition to increasing response quality, protection against misuse is also important. To ensure adequate use, guardrails can be used as a kind of safety net to prevent LLMs from generating or disclosing unwanted content. These can be implemented at different points in the architecture. Alternatively, a possible outflow of information can also be prevented by sophisticated authorization concepts up to multitenant systems with separate databases and large language models.

From the point of view of data sovereignty, it should be noted that different hosting variants can be selected depending on the respective infrastructure requirements or preferences of a company. These range from international hyperscalers and hosting in EU countries to on-premise solutions.
(see figure 5)

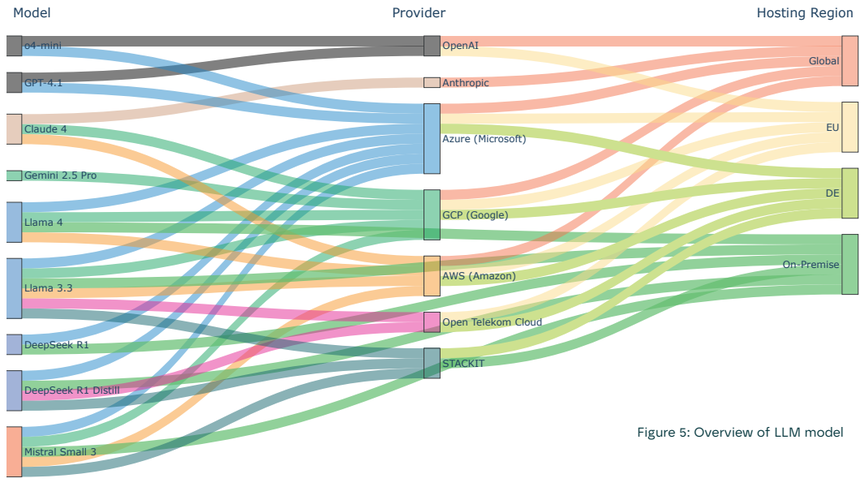


Figure 5: Overview of LLM model

It is also essential to consider the users, as without their acceptance, any solution is doomed to failure, especially if it involves rethinking existing workflows. The following chapter deals with this aspect in more detail.

3.2.3 Users in focus

The ability of users to categorize the responses of a chatbot or assistance system and to work with them efficiently and effectively can be significantly supported by user guidance and a focus on the context of use.

Through a deep understanding of the context of use, the suitable interaction modality can be identified. Combined with thoughtful user guidance, this enables the formation of an intuitive understanding of AI.

Does it always have to be a chatbot?

One example

In a company, knowledge distributed across various sources should be made more easily available to employees. For the usability engineer trying to identify a suitable interaction modality during the design phase of a project the following questions arise:

Do we expect users to conduct open, dialog-like research? That their questions will develop the more information they receive? Or do we expect that users want to find exactly the documents that are relevant to them so that they can continue working quickly?

Or to put it more generally: What is the context of use for users? Do they need individual and customized learning support in natural language? Or perhaps something more like a „Google my Company“ search service that sorts relevant content from internal wiki pages and SharePoint files according to relevance using AI technology? Perhaps a combination of both: a natural language answer plus a list of documents? Or how about if the user only receives a generated answer in natural language if they have actually asked a question, as Google currently does?

The right interaction modality facilitates user guidance.

A chatbot is often a universal solution for learning guidance, knowledge acquisition and material generation. The main problem for users is evaluating the answer for relevance and correctness. Although the input is decisive for the chatbot result, the user receives little feedback on the interpretation of their input. Often only a warning is given: „AI can make mistakes. Check important information“. Instead of a universal solution, specialized solutions should be developed that are better adapted to the context of use and clearly communicate their functionality to strengthen user skills in dealing with AI systems without having to rely solely on warnings.

A Google-like search function instead of a chat could be more effective in obtaining knowledge. Users are used to skimming through lists of results with headlines, short descriptions and relevant text excerpts to find the right element and understand how their search input has been interpreted.

If this list of results is now to be expanded to include natural language answers for the convenience of the user, a list of relevant documents can be enriched with brief information that helps to the classify and check the AI-generated answer. In contrast to a pure list, the visualization of the relevant passages makes it much easier for the user to check for correctness.

Finally, the focus on the context of use results in an actually helpful extension of user workflows with AI, while at the same time ensuring that users directly learn to interact with AI solutions in a responsible manner. Based on context-dependent questions and hints, users can be empowered to make queries (prompts) that lead to better results. This trains AI literacy through daily interaction.

3.3 Business Impact

In practice, Advanced RAG-supported chatbots often save more than 80% of the time required to find information. Combined with the results of the Kothes study that 86% of service employees spend over 30 minutes per day searching for information, with 12% needing over 2 hours (Kothes GmbH, 2023), the impact on employee productivity and thus on everyday working life quickly becomes clear. Looking at the weighted average, this results in an average time saving of around one hour per day, which corresponds to a productivity increase of 12.5% for an 8-hour day. This figure seems all the more significant when you consider the context that a productivity gain of 5% has been recorded for machinery and plant engineering over the last 14 years (VDMA, 2025).



Figure 6: Business impact of chatbots in knowledge management

4. Current developments and new opportunities through Agentic AI

The current trend is clearly towards agentic AI (Gartner, 2025), i.e. AI-systems that use autonomous agents to react dynamically to circumstances and interact with their environment. These agents divide complex requests into several subtasks, plan independently and orchestrate specialized agents. They can also call up separate tools to enrich or validate information. Integrated into RAG-systems, they allow the AI system to adaptively retrieve information directly from program interfaces (APIs) without having to add it to the vector database, validate it and contextualize it. These capabilities lead to a higher quality of the generated answers through agent-based, quantifiable quality assessments and, if necessary, refinements through automated adaptation of the queries. However, this approach leads to longer response times due to the increasing complexity. Therefore, a combination with classic advanced RAG techniques is ideal. By involving agents (for e.g. additional information collection, decomposition or quality inspection), depending on the complexity of the query, AI solutions can excel in an industrial context.

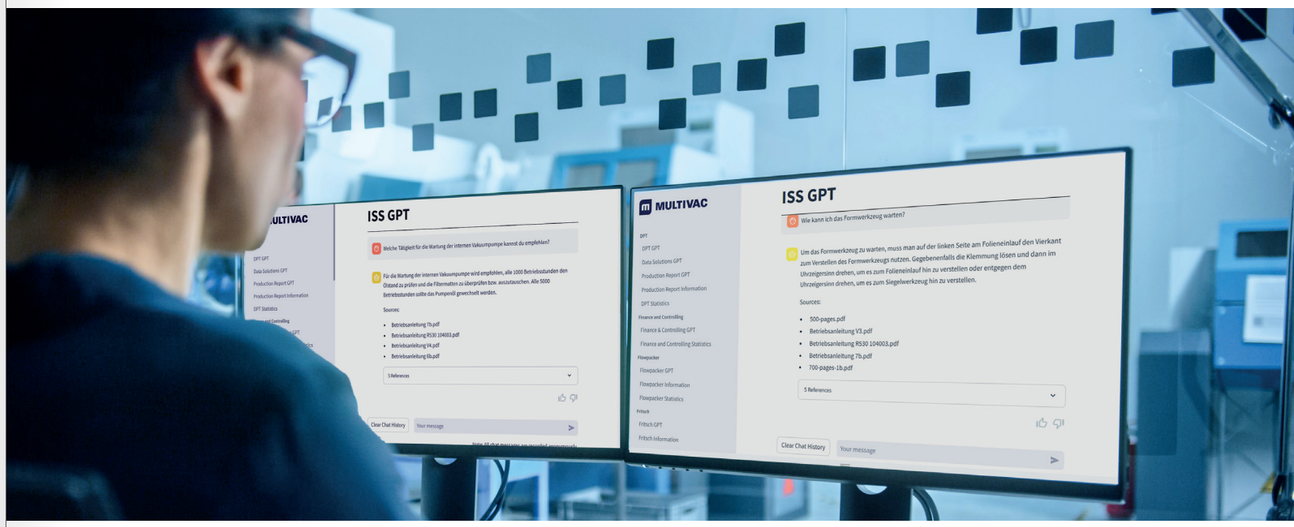
In addition, agents also open up new application possibilities that cannot be addressed by classic LLMs and RAG in a satisfactory manner. One example of this is natural-language access to databases. An agent-based approach implemented by us comprises different agents to manage orchestration, query, code and response generation. This ensures that initially incorrect database queries are automatically corrected and queries that require more complex, nested database queries are answered in a targeted manner. The integrated code generation means that not only data can be returned but also prompts that normally require separate evaluation logic can be answered correctly, e.g. prompts that require a correlation analysis such as „Is the speed of the extruder related to deviations in diameter“. Previously, this was only possible with knowledge of database query languages and data science. Now, even skilled workers without this specialist knowledge can interact directly with process data and understand correlations based on data. This enables more objectively sound decisions to be made.

5. Customer references

5.1 MULTIVAC

MULTIVAC offers complete solutions for packing and processing food, medical and pharmaceutical products as well as industrial goods. The portfolio includes various packaging technologies, automation solutions, labeling and inspection systems and packaging materials.

Due to the large number of company divisions, an immense knowledge base was stored in different databases at MULTIVAC. This made the search for relevant information very confusing and time-consuming. Also, the constant growth of both structured and unstructured data negatively affected availability. In order to make the growing amount of knowledge efficiently accessible to all employees, we strengthened MULTIVAC's Data Solutions team with our expertise to optimize internal knowledge management with the help of a chatbot. The chatbot uses data from various areas of the company to generate precise answers to employees' queries and refers to the underlying sources. This significantly speeds up the search for information.



5.2 AI-Chatbot telli

The Institute for Film and Image in Science and Education (FWU) supports the states of the Federal Republic of Germany in developing modern media for education.

XITASO and the FWU are developing the open-source chatbot telli for use in the classroom, which supports both teachers in their teaching and students in their learning process.

Teachers are supported by AI-supported assistants in preparing lessons, e.g. in creating evaluation forms, and students have the opportunity to interact with topic-related chatbots to learn complex contexts in an interactive environment.

6. With XITASO to project success

As a digitalization partner and expert in high-end software engineering, XITASO has been advising B2B customers since 2011, identifying digitalization potential, optimizing business processes and creating digital strategies and solutions. The company is characterized by an agile way of thinking and works with proven expertise in the areas of Industry 4.0, Internet of Things (IoT), data science, augmented reality and artificial intelligence. The focus is on making complex systems simple and intuitive to use.

With a focus on excellent solutions for the challenges of digital change, XITASO supports customers from various industries throughout Germany, including the focus sector of mechanical and plant engineering. With more than 250 employees, XITASO has a unique wealth of experience and knowledge to lead digitalization and AI projects to success together with its customers. Based on the countless successful projects since its foundation, XITASO has developed process models along the various competencies required to minimize project risks and ensure that schedules and budgets are adhered to. The following model has proven itself for the successful introduction of AI projects.

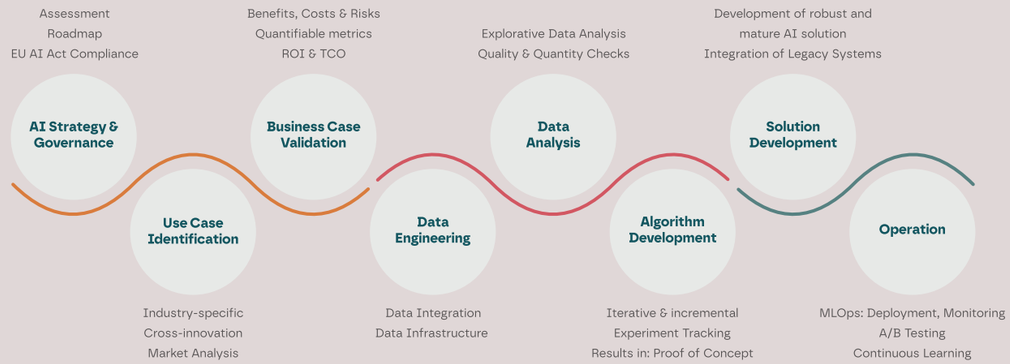


Figure 7: The XITASO process model

Within the framework of this process model, the following six experiences and guiding principles have proven to be central to the success of the project:

1. Before the actual start of the project, the company should use XITASO to define the framework for the introduction and development of software based on AI technologies. Within this framework, companies can discuss and define the relevant aspects of data protection and security, ethical and ecological goals, etc. for their AI strategy and governance. The answer to the fundamental question of whether AI software should tend to be purchased or developed in-house also has an impact on further organizational and procedural measures and should therefore be given before the project starts.
2. „AI is hard“: Even if companies do not develop their AI systems themselves, but „only“ operate them, it is essential that they have the necessary expertise. XITASO has years of experience in this area and invests in researching new technologies and methods to provide you with state-of-the-art support.
3. From the first day of the project, the domain experts should work together with the data scientists and AI experts and focus on the application scenario. Only once they have jointly worked out the benefit of the AI solution should they take a closer look at the quantity and quality of the data required for it.
4. Think big, start small: this principle applies to few other projects as much as it does to AI projects. Before implementing a comprehensive and high-performance infrastructure or developing a smart and intuitive user interface, a proof of concept is always required. Only if the prototype meets expectations is the effort involved in setting up the productive environment worthwhile. If, on the other hand, it turns out that the AI cannot answer the defined questions or cannot answer them with sufficient accuracy, at least a fraction of the project costs have only been invested up to this point. This is risk and financial management in action.
5. XITASO follows this principle with its AI experts in all projects. Only once the results of the pilot phase have been evaluated together with the customer and are satisfactory, further decisions on implementation and the production environment are to be made and implemented. After all, beginning and end of every AI project is the business case.
6. With the help of its comprehensive security expertise, XITASO supports customers in providing long-term security updates for security-critical components. In addition, XITASO‘s security team monitors all dependencies in the IT environment for security vulnerabilities for its customers and helps to close them as quickly and permanently as possible.

A strong team for joint success

With XITASO as a partner, companies can be confident that they will benefit from comprehensive expertise that will enable them to successfully master the challenges of knowledge management in the digital age. Please do not hesitate to contact us if you have any questions or suggestions. We are always here for you.

Your Contact

We want to work with you to make your AI project a success. Get in touch!



Dr. Richard Nordsieck

Head of AI & Data

Phone +49 821 885882-89
Email richard.nordsieck@xitaso.com



Matthias Widmann

Sales & Business Development

Phone +49 821 885882-228
Email matthias.widmann@xitaso.com

About XITASO

Your partner for high-end software engineering in the age of AI

As a digitalization partner and expert in high-end software engineering, XITASO works with its customers on demanding digitalization projects to create customized solutions that scale and function securely from the hardware to intuitive user interfaces and into the cloud.

The company supports B2B customers in identifying digitalization potential and optimizing business processes. With industry experience, enthusiasm for innovative technologies such as artificial intelligence, and a passion for excellent collaboration, XITASO helps its customers master the challenges of digital transformation and secure their market advantage.

2011	13	240+	20,5
Founded	Locations	Employees (12/2023)	Revenue 2023 in mio. €

Bibliography

- Bottaro, J. P. (2024). Musings on building a Generative AI product. https://www.linkedin.com/blog/engineering/generative-ai/musings-on-building-a-generative-ai-product
- Destatis. (2025). Qualität der Arbeit—Dauer der Beschäftigung beim aktuellen Arbeitgeber. https://www.destatis.de/DE/Themen/Arbeit/Arbeitsmarkt/Qualitaet-Arbeit/Dimension-4/dauer-beschaeftigung-aktuell-Arbeitgeber.html
- Gao, Y., Xiong, Y., Gao, X., Jia, K., Pan, J., Bi, Y., Dai, Y., Sun, J., Wang, H., & Wang, H. (2023). Retrieval-augmented generation for large language models: A survey. arXiv preprint arXiv:2312.10997, 2(1).
- Gartner. (2025). Hype Cycle for Artificial Intelligence. https://www.gartner.com/en/documents/6579402
- Hassani, H., & MacFeely, S. (2023). Driving excellence in official statistics: Unleashing the potential of comprehensive digital data governance. Big Data and Cognitive Computing, 7(3), 134.
- Höbbel, K., & Spohr, L. (2025). Presseinformation—Industrie 4.0. https://www.bitkom.org/Presse/Presseinformation/Industrie-4.0-Unternehmen-KI-Produktion
- Kang, H., Fu, L., Funk, R. J., Wang, X., Ding, J., Liang, S., Wang, J., Zhou, L., & Zhou, C. (2024). Scientific and technological knowledge grows linearly over time. arXiv preprint arXiv:2409.08349.
- Kothes GmbH. (2023). Insight-Report Service.
- Michelsen, C., & Junker, S. (2024). Fachkräftemangel: Stille Reserven am Arbeitsmarkt heben. MacroScopePharma 09/24. https://www.vfa.de/de/wirtschaft-standort/macroscope/macroscope-fachkraeftemangel
- Ovadia, O., Brief, M., Mishaeli, M., & Elisha, O. (2023). Fine-tuning or retrieval? Comparing knowledge injection in llms. arXiv preprint arXiv:2312.05934.
- Rudnicka, J. (2025). Nachfrage nach Fachkräften: Entwicklung des Fachkräfteindex in Deutschland vom 1. Quartal 2015 bis zum 4. Quartal 2023. https://de.statista.com/statistik/daten/studie/378248/umfrage/entwicklung-des-fachkraefteindex-in-deutschland/
- VDMA. (2025). GenAI im Maschinen- und Anlagenbau.

List of figures

- **Figure 1:** Development of AI & AI at XITASO 06
- **Figure 2:** Chatbots for organizational knowledge 07
- **Figure 3:** Naive RAG vs. Advanced RAG 09
- **Figure 4:** Advanced RAG architecture sketch 09
- **Figure 5:** Overview of LLM model 10
- **Figure 6:** Business impact of chatbots in knowledge management 11
- **Figure 7:** The XITASO process model 15

A big thank you to our authors:

Dr. Richard Nordsieck,
Manish Bhandari,
Jessica Friedline,
Bernd Schächterle &
Anton Winschel

XITASO

enjoy digital excellence

Learn more about
AI at XITASO

XITASO GmbH
IT & Software Solutions
Austraße 35 · D-86153 Augsburg

Phone +49 821 885882-0
Email info@xitaso.com
Web www.xitaso.com