RAG Chatbots in Customer Service for Business and the Public Sector

Executive Summary

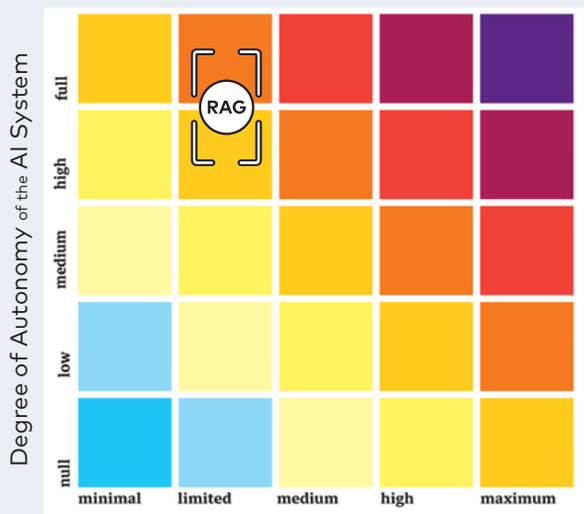
Our lives are increasingly shifting into the digital sphere. The increase in digital communication with customers and citizens presents challenges for businesses and public administrations, as processing enquiries and requests efficiently and effectively also incurs high costs. Chatbots powered by artificial intelligence (AI), particularly Retrieval-Augmented Generation (RAG) chatbots, are therefore becoming increasingly important as they can process a large number of enquiries quickly and cost-effectively.

Practical experience in various countries to date shows that RAG systems can deliver significant benefits for businesses and users alike.[2,6] RAG chatbots retrieve information from knowledge sources not included in large language models' (LLMs) training datasets. This means that users receive more accurate answers. Using RAG chatbots can help organisations improve cost efficiency and gain users' trust. However, there are also significant challenges that counterbalance these benefits. Due to the LLMs used as their underlying technology, RAGchatbots are susceptible to "hallucinations", orfalse statements, when the quality of the data is poor and the queries are highly complex. Such misinformation is particularly problematic when it originates from public authorities, as citizenstend to regard information from these sources as reliable. Overall, misinformation can have serious consequences, including unlawful behaviour, loss of legal rights and erosion of trust in public institutions.

Overall, according to the ALAIT Risk Radar, the use of RAG chatbots is classified as "high risk" (dark orange; see chart). The ***EU AI Act*** classifies the risk associated with the use of RAG systems in customer service by businesses and the public sector as „limited" (Level 2). This is because it can be assumed that – unlike, for example, automated decisions regarding government transfer payments – they do not have any immediate far-reaching consequences for individuals seeking information. However, in terms of the ***degree of autonomy***, the RAG systems in use are generally fully autonomous, as they operate independently and require minimal or no human intervention. Human involvement is thus limited to long-term oversight through audits (Levels 4 and 5).

Organisations and system developers can reduce the risk by implementing risk mitigation measures. Compliance with the General Data Protection Regulation (GDPR) and the ***EU AI Act*** is essential, as these include ***transparency*** guidelines. Furthermore, ***data protection*** issues can be avoided by using precise, high-quality data that does not

ALAIT Risk Radar for RAG Chatbots in Customer Service for Business and the Public Sector



Retrieval-Augmented-Generation (RAG) Chatbots

The ALAIT Risk Radar is a scientifically developed risk analysis tool for artificial intelligence (AI) that classifies AI applications based on context and taking into account their degree of technical autonomy, thereby making the risk landscape visible to users at a glance. The following principle applies: The higher the operational risk arising from the application context and the greater the degree of autonomy of the AI system with regard to decision-making, the riskier the application is classified. An expanded bracket indicates a range within the risk classification. A low degree of autonomy in an AI system does not mean one can sit back and relax. It requires a strong role for the people who use it. (For details on the tiered model, see p. 8ff.)

**ALAIT is a research project initiated by the Federal Ministry of Innovation, Mobility, and Infrastructure (BMIMI) to develop a novel form of technology assessment for AI technologies using scientific methods.*

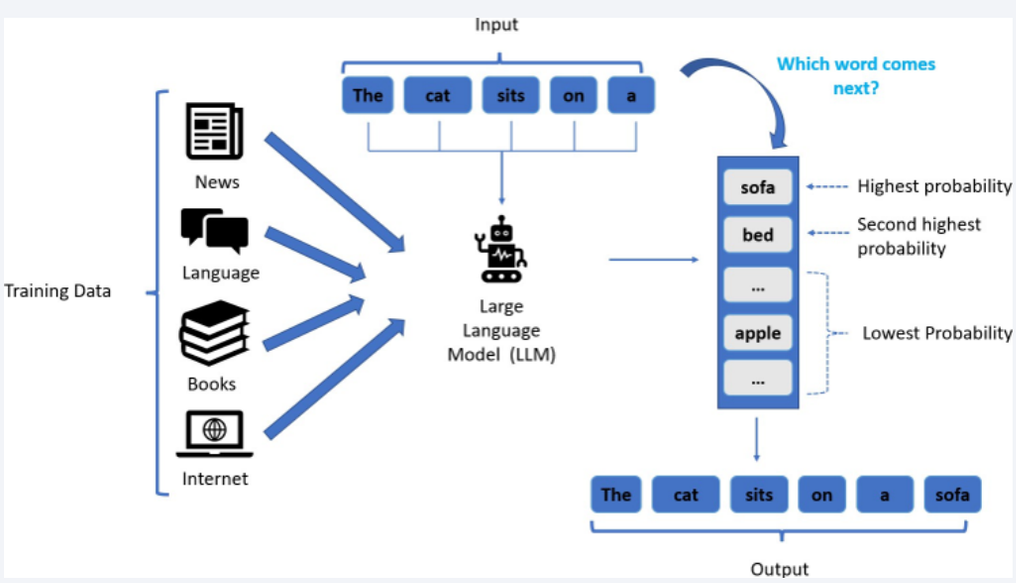
contain sensitive personal information. Human oversight of RAG systems is also crucial, in the form of checks on results and security breaches. In any use case involving RAG chatbots, users must be informed about how their data is used.

Introduction and Technology Description

In 2024, the Technical University of Munich (TUM) introduced 'OneTutor', a RAG system developed by one of its start-ups, to provide its students with up-to-date information on lectures and courses at the university. This successful system is already in use by around 9,700 students at 22 universities.[9]

Like TUM, an increasing number of companies and public sector organisations are turning to RAG chatbots to manage the high volume of enquiries from customers and members of the public.

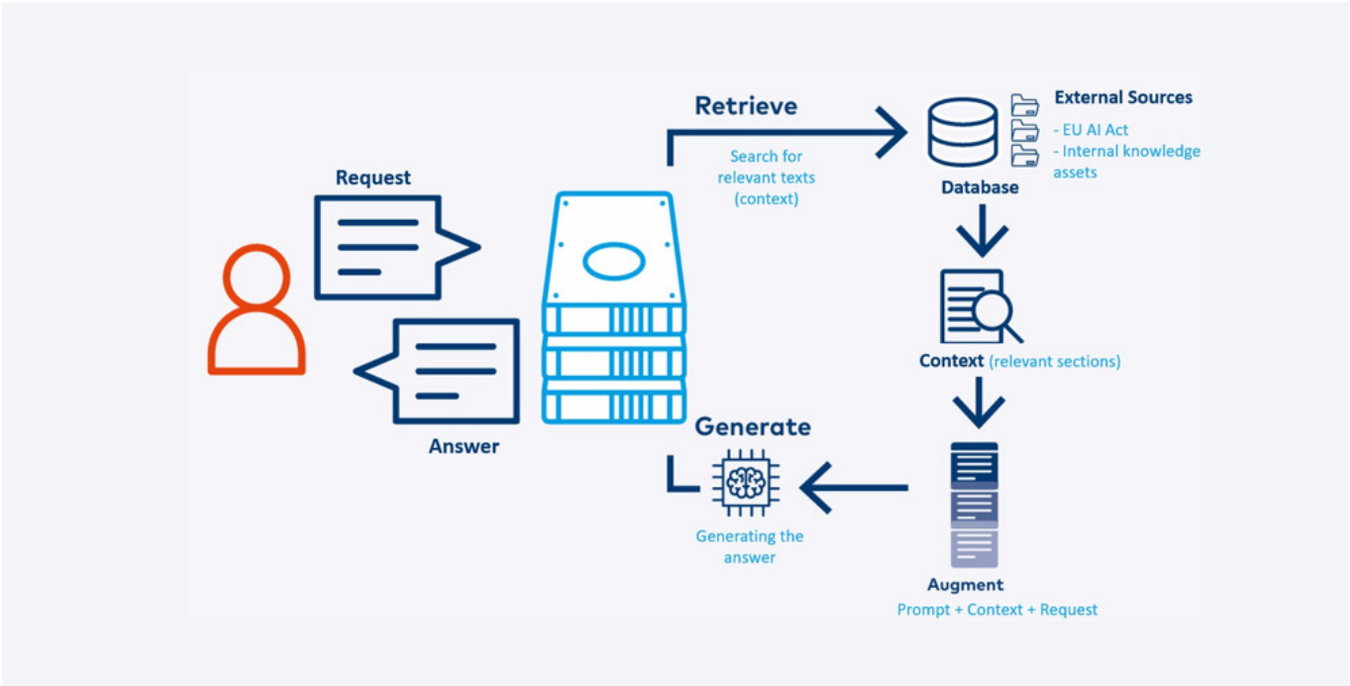
RAG chatbots are based on Large Language Models (LLMs). These models form the core AI component of all RAG systems. LLMs process user input and generate a response based on the information they have been trained on. The following diagram illustrates the basic principle of LLMs:



A language model (LLM) learns by recognizing patterns in how words and sentences typically relate to one another across a wide variety of text types, such as news articles, books, conversations, and web pages. If you give the model the beginning of a sentence, such as "The cat is sitting on the...," it calculates the word most likely to follow. In doing so, it considers numerous possibilities—such as "sofa," "bed," or "spoon"—and ultimately selects the most probable word. This results in the complete sentence "The cat is sitting on the sofa." In short, an LLM is a programme that uses learned linguistic knowledge to predict the next best word.

RAG chatbots retrieve information from knowledge sources that are not included in the LLM's training dataset, either internal or external. When a user asks a question, the system first searches these sources for relevant information, before using an LLM to generate a response.[2]

The following diagram illustrates the basic principle of RAG chatbots:[13,23]



The RAG model operates in the following three steps:[3,4,5]

1. **Convert the question:** The entered question is converted into a specific numerical format (vector). The system then performs a semantic search.

2. **Retrieve information:** The system searches for relevant information within its own data or predefined sources, such as specific databases or documents. The found texts are also converted into vectors and stored in a vector database.

3. **Generate a response:** A language model (LLM) generates a response based on the provided information. In more advanced versions, the system can compare the response with the original question again to improve or refine it.

This results in a response that is based on more up-to-date and relevant information than that provided by conventional chatbots, such as ChatGPT or Gemini. Consequently, RAG systems can answer questions on topics about which the language models have no direct knowledge gained from training.

Despite there being successful examples of its application, the application of RAG chatbots in practice presents a number of challenges. For instance, the city of New York's generative AI chatbot (MyCity Chatbot) provided responses that either encouraged users to break the law or contained inaccurate legal information.[11] For example, the chatbot stated that it was legal to dismiss individuals who had complained about sexual harassment.[11,24]

Even in business context, incorrect information provided by chatbots can have significant consequences. For instance, Air Canada's chatbot provided a customer with inaccurate information about the airline's bereavement fare policy. The customer relied on this information and subsequently successfully sued for a refund. This resulted in legal consequences for the company, as well as significant reputational damage and a loss of trust.[11,25] Such examples demonstrate that RAG systems that are inadequately tested or not regularly updated pose significant risks in terms of liability, consumer protection and institutional trust for both the public and private sectors.

The key opportunities and challenges posed by RAG chatbots are set out systematically below.

Opportunities

Compared to traditional chatbots, the use of RAG chatbots in public administration offers a number of advantages for end users and the organisations that deploy them:

- **Speed and *Accuracy:*** Although information is generated through several processing stages, RAG chatbots are faster and significantly more accurate than conventional chatbots.[1,3,6,31] As a study by Deloitte in Canada has shown, automating repetitive tasks using AI tools such as RAG chatbots can reduce processing errors by case handlers by up to 40%.[7] However, these improvements should be viewed with caution as they depend heavily on the dataset, scope of application, and evaluation method.

- **Simplified access to public information:** RAG chatbots are available around the clock and can retrieve information from complex legal or administrative sources, such as laws, forms and admission criteria, and translate bureaucratic language into simple, understandable terms. This reduces barriers for citizens who struggle with complex public documents.[11,21] Ein Beispiel dafür ist die Rundfunk und Telekom Regulierungs-GmbH (RTR), One example is Austria's Rundfunk und Telekom-Regulierungs-GmbH (RTR), which has developed an RAG system that can effectively and transparently answer questions about the ***EU AI ACT*** and the regulatory framework for AI, as the exact source of the information can be displayed.[23]

- **User trust:** As RAG chatbots are connected to real-time databases, they can provide context-relevant, up-to-date information, thereby enhancing the user experience[1,8,10] und somit die allgemeine Zufriedenheit mit dem Service im Vergleich zu herkömmlichen Chatbots verbessern.[2,5]

- **Workload reduction:** RAG chatbots can automatically respond to standardised and non-standardised enquiries that recur frequently and retrieve relevant information from verified documents. This significantly reduces staff workload when it comes to routine communication. The freed-up time can then be used for complex cases, providing personalised advice, or making decisions that require human judgement.[31]

- **Up-to-date information:** RAG systems can connect the LLM not only to local databases. They can also connect it directly to news sites or other frequently updated sources of information. This ensures that both staff and users are provided with the most relevant information.[2,5]

- **Customer satisfaction:** By integrating RAG chatbots, customers receive context-specific support. They benefit from immediate, personalised, multilingual and barrier-free communication, which is not always possible with human customer service.[10]

- **Cost-effectiveness:** Companies such as Vodafone, Alibaba and Klarna have reported achieving cost savings of millions since integrating AI-powered chatbots.[8] Building company-specific RAG bots on top of existing company-specific LLM ***foundation models***, can also significantly reduce development time.[5,8]

Challenges and Risks

1. **Hallucinations:** Although RAG chatbots generally perform better than conventional chatbots, they are prone to "hallucinating", i.e. providing inaccurate or false information that appears authentic but is factually incorrect. This issue affects all LLM-based systems and is exacerbated by the inclusion of irrelevant documents in the responses.[3,6,11] An empirical study of RAG-based tools for legal research found that between 17% and 33% of these systems' responses were inaccurate or made up. They often provided incorrect legal explanations and fabricated quotations.[27]

2. **Data Quality:** The effectiveness and reliability of RAG systems depends heavily on the quality of the database accessed during a query. This directly impacts the chatbot's performance.[3] The external knowledge sources used for training and querying the RAG system should be structured in a way that makes them easily identifiable to the system. This is particularly apparent with complex queries: if the RAG system cannot find the correct document, or if the data is poorly indexed, the chatbot may provide an irrelevant response or revert to the base LLM model. This increases the likelihood of errors being made by the system, whilst the ***accuracy*** of the response decreases.[3]

3. ***Data Protection*:** As RAG systems access databases that may contain sensitive information, disseminating private data has become a major problem. Studies show that misleading text and instructions may be embedded in hidden website content. This causes LLMs to follow these instructions when conducting searches via such "contaminated" websites. This process is known as "indirect prompt injection". Consequently, RAG systems may also retrieve personal data inadvertently included in model training and "over-share" it.[6,12] In one instance, for example, security researchers discovered a vulnerability in the Microsoft 365 Copilot RAG assistant. This vulnerability automatically made confidential files (emails, documents, etc.) from users' accounts available.[28]

4. **Excessive Trust in Technology:** Several studies have shown that people tend to trust AI systems more than other humans. This phenomenon is known as "automation bias".[17] There is a risk that people will no longer scrutinise AI-generated information sufficiently, even when they have conflicting information about a specific case.[18]

5. ***Transparency*:** From a technical perspective, the problem is that LLMs are not transparent or comprehensible due to their complexity – not even for experts. This phenomenon is also referred to as the "***black box***".[33,34] Furthermore, the use of AI tools requires appropriate expertise and training, which must be provided by public authorities and companies wishing to use such tools (see Article 4 of the ***EU AI Act***).[32]

6. **Copyrights:** The LLM components of the RAG chatbot are trained using vast amounts of data, much of which is protected by copyright. The extent to which using this protected content for AI training is permissible is currently the subject of heated debate, and is likely to occupy the courts in the coming years.[22] In 2025, for instance, a group of publishers (including Forbes and The Guardian) sued an AI start-up whose RAG chatbot's LLM component had copied and reproduced their copyright-protected articles. The chatbot displayed verbatim or summarised passages from protected news content in its responses, sometimes bypassing paywalls in the process.[29]

7. **Liability:** From a legal perspective, organisations that deploy RAG systems are ultimately liable for their use and cannot shift this liability onto users. According to § 43 of the German Act on Limited Liability Companies (GmbHG), directors are liable for breaches of their duty of care.[26] 26 Unlike with established technologies, where responsibility can be delegated to qualified individuals, there are no state-recognised certifications orregulated responsible parties for LLMs. Management therefore bears direct responsibility and cannot delegate it in a legally binding manner. An example of this is the Air Canada case described above.[11,25]

Recommendations for Practical Use

The following strategies are recommended to address the challenges and risks outlined below for public authorities, businesses and users:

For Public Authorities and Business:

- **Compliance with Guidelines:** All organisations, both public and private, must comply with the guidelines and must not deviate from them in their RAG systems.[20] In particular, these systems must comply with the General Data Protection Regulation (GDPR) to protect users' privacy. The processing of personal data must comply with the core ***data protection*** requirements of necessity and proportionality.6 Furthermore, the implementation and application of the ***EU AI Act*** is monitored by authorities such as the European Office for Artificial Intelligence ("AI Office") and national market surveillance authorities. Article 50 of the ***EU AI Act*** stipulates that users interacting with an AI system (e.g. a chatbot) must be informed that they are communicating with an AI system, unless this is obvious.[19]

- **Precise, High-Quality Data:** The reliability of RAG systems hinges largely on the quality of the datasets used to train LLMs, as well as those used to feed them during operation. To improve ***accuracy***, external knowledge sources should also be labelled clearly and in a structured manner. The following principles must be observed when using data for RAG systems:[14,30]

  - **Data minimisation:** Only the minimum amount of personal data necessary to fulfil the task should be used.
  - **Purpose limitation:** Data collected for the provision of public services should not be used arbitrarily for other purposes.

- **Option for Human Intervention:** Users of RAG chatbots should always have the option of contacting human staff, especially if their query is complex or requires personal judgement, or if they are unsure about the ***accuracy*** of the response. This prevents unintended automation and supports the "human-in-control" oversight principle, which should reduce the overall risk associated with RAG systems.

- **Technical *Transparency:*** A transparent RAG system should clearly indicate the source of the information provided in its responses. Explainable AI (XAI) allows users to view the sources used to generate responses. This is intended to increase transparency and encourage critical evaluation of the responses.[31]

- **Use-Case-Specific Design:** RAG systems must be precisely tailored to their intended use. The first step is to define the permitted topics, user groups and document types, and the questions that the system should not answer. The next step is to select a suitable language model, ideally one that is highly customisable (e.g. open source). Only verified, trustworthy content should be incorporated into the system to avoid errors and hallucinations. The quality of the responses should be regularly verified, for example by checking whether the information is relevant and accurate. Furthermore, safeguards are required to ensure that sensitive data is not inadvertently disclosed or misused.[30,31]

For Users (Citizens and Customers):

- **Right to *Transparency:*** Users should be able to recognise when AI is being used, and this should be made clear through a notice in the form of a disclaimer or disclosure. This promotes transparency and strengthens trust in the system, helping users to critically evaluate responses. This is especially important for RAG chatbots, whose responses are based on a combination of a language model and external documents.

- ***Data Protection Rights:*** Furthermore, users have the right to know how and when their personal data is used, and to which systems it is transferred (Art. 15 GDPR).[15] The Austrian Data Protection Authority (Datenschutzbehörde - DSB) is the point of contact in the event of a suspected breach of data protection rights.[16]

- **Building and Strengthening AI Literacy:** It is important to learn about AI technologies, develop the right skills, and understand one's rights.[18] In Austria, the Rundfunk- und Telekom Regulierungs-GmbH (RTR) serves as an information and contact point for the general public. The RTR AI Service Centre assists citizens with queries relating to transparency rights when interacting with AI systems and the review of AI decisions.[13]

Important Terms

AI accuracy: Refers to an AI system's ability to make proper predictions or decisions. It is an important measure of its performance and crucial for determining its effectiveness and reliability.

AI autonomy: The ability of an AI system to achieve a set of goals under a range of uncertainties in its environment, independently and without external intervention.

AI bias: Bias refers to the systematic unequal treatment of certain objects, individuals or groups compared to others. Treatment refers to any kind of action, including perception, observation, representation, prediction or decision-making.

Black box systems: Refers to all systems whose internal decision-making processes are not transparent or traceable to humans. This means that only inputs and outputs can be observed, without understanding exactly how the processing in between takes place.

Data protection in AI: The full range of practices and concerns relating to the ethical collection, storage and use of personal data by artificial intelligence systems.

Deepfakes: Realistic-looking media files (photos, audio files or videos) that have been altered, created or falsified using AI techniques.

European Union Law on Artificial Intelligence (EU AI-Act): A European regulation on artificial intelligence (AI) – the first comprehensive regulation on AI issued by a major regulatory authority. It focuses in particular on high-risk AI systems.

Explainable AI (XAI): Refers to methods and techniques that make it possible to better understand and account for the decisions and outcomes of AI systems.

Foundation Models: Large, pre-trained AI models (e.g., GPT or BERT) that are based on large datasets and can be fine-tuned for a wide range of tasks.

Generative AI: Artificial intelligence that can generate new content such as text, images, music or code, based on patterns learnt from training data.

Transparency: This means that the functionality, decision-making processes and areas of application of an AI system are comprehensible, explainable and openly accessible – to developers, users and other stakeholders.

Explanation of the ALAIT Risk Radar's tiered model

The ALAIT AI Risk Radar illustrates the relationship between application risk and the autonomy of an AI system. The risk levels are based on the EU AI Act, in particular Article 6 and Annex III, which deal with high-risk areas of AI application. Lower levels of system autonomy and application risks are represented by cooler colours (blue), whilst higher levels of system autonomy and application risks are represented by warmer colours (red).

The change of color conveys the increased risk of AI applications. This color scale can be used to identify the overall risk: violet and dark red – very high, red and dark orange – high, light orange and yellow tones – medium, blue tones – low overall risk. Ideally, high application risks and system autonomy should be avoided or only be used after very careful consideration.

Level of Autonomy of the AI System

Level 1: No Autonomy

AI is a passive tool; it is people who make all the decisions and take action.

Example: Diagnostic systems that display raw medical data or analyse the data (without providing recommendations!).

Recommended use cases: Scenarios involving high risks or in which ethical decisions are of crucial importance (e.g. medical diagnostics, the justice system).

Level 2: Low Degree of Autonomy (Human-in-the-Loop)

The AI provides recommendations or options, but users remain responsible for selecting and approving actions.

Example: AI suggests optimal routes for logistics, or recommendation systems in e-commerce.

Recommended use cases: Tasks of medium complexity involving moderate risks (e.g. supply chain optimisation).

Level 3: Medium Degree of Autonomy (Human-on-the-Loop)

The AI carries out certain tasks autonomously, with humans intervening in exceptional cases.

Example: AI-supported manufacturing processes where the system controls the machines, but users intervene in the event of anomalies.

Recommended use cases: Scenarios in which continuous human involvement is not required, but critical risks necessitate human oversight (e.g. industrial automation, monitoring of financial transactions).

Level 4: High Degree of Autonomy (Human in Control)

The AI system operates largely autonomously, but allows users to override it themselves in order to avoid undesirable results.

Example: Autonomous vehicles.

Recommended use cases: Low- to medium-risk environments (e.g. logistics, basic traffic management).

Level 5: Full Autonomy with Minimal Supervision

The AI system operates independently and requires minimal or no human intervention.
Human involvement is limited to long-term audits.

Example: Autonomous agricultural machinery, AI for electricity distribution, underground railways, airport rail links.

Recommended use cases: Environments with low safety or ethical risks and high reliability of the AI system (e.g. repetitive tasks in controlled environments).

Degrees of Application Risk

Level 1: Minimal Risk

The AI system has no impact on users or decision-making.

Examples: Filters, NPCs, recommendation algorithms without serious consequences (DeepL, other translation systems).

Criteria: No direct impact on the high-risk areas covered by the ***EU AI-Acts***.

Stufe 2: Limited Risk

AI systems that interact with users but do not make high-risk decisions. The risk increases when there is a lack of transparency regarding the involvement of AI.

Examples: Chatbots and AI-generated content without disclosure, simple automation tasks.

Criteria: Areas not included in the 'high-risk' list under the EU AI Act.

Level 3: Medium Risk

AI systems do not have any particular impact on individuals, but they do have an effect at a collective or societal level.

Examples: Generative AI such as ChatGPT and other systems that can indirectly influence the environment in which they are used.

Criteria: AI systems that are available for public use and have the potential to influence and, in the long term, change existing practices.

Level 4: High Risk

Any algorithm used in what the EU AI Act defines as 'high-risk areas' or which has a direct impact on individuals.

Examples: Medicine, biometrics, critical infrastructure, education and vocational training, employment, access to public sector services, law enforcement, migration.

Criteria: Classification as a 'high-risk sector' under the EU AI Act is only applicable if the rules on transparency and data quality are complied with.

Level 5: Extreme Risk

Any algorithm used in what the EU AI Act defines as 'high-risk areas'.

Examples: Medicine, biometrics, critical infrastructure, education and vocational training, employment, access to public sector services, law enforcement, migration.

Criteria: Classification as a 'high-risk area' if the rules on transparency and data quality are NOT complied with.

Sources

1 Singh, A. (2025). Agentic RAG Systems for Improving Adaptability and Performance in AI-Driven Information Retrieval. SSRN. https://doi.org/10.2139/ssrn.5188363

2 McKinsey. (2024, October 30). What is RAG (retrieval augmented generation) | McKinsey. https://www.mckinsey.com/featured-insights/mckinsey-explainers/what-is-retrieval-augmented-generation-rag

3 Jeong, C. (2024). A Study on the Implementation Method of an Agent-Based Advanced RAG System Using Graph. Knowledge Management Research, 25(3), 99–119. https://doi.org/10.15813/kmr.2024.25.3.005

4 Sutter, M. (2025, August 22). Native RAG vs. Agentic RAG: Which Approach Advances Enterprise AI Decision-Making? MarkTechPost. https://www.marktechpost.com/2025/08/22/native-rag-vs-agentic-rag-which-approach-advances-enterprise-ai-decision-making/

5 What is RAG? - Retrieval-Augmented Generation AI Explained - AWS. (n.d.). Amazon Web Services, Inc. Retrieved 19 August 2025, from https://aws.amazon.com/what-is/retrieval-augmented-generation/

6 Retrieval-augmented generation (RAG) | European Data Protection Supervisor. (2025, September 18). https://www.edps.europa.eu/data-protection/technology-monitoring/techsonar/retrieval-augmented-generation-rag

7 Deloitte. (2021). The robots are here: Are you ready?

8 Vohra, D. K. (2024, October 1). How AI and RAG Chatbots Cut Customer Service Costs by Millions. How AI and RAG Chatbots Cut Customer Service Costs by Millions. https://www.nexgencloud.com/blog/case-studies/how-ai-and-rag-chatbots-cut-customer-service-costs-by-millions

9 OneTutor. (n.d.). OneTutor—KI Tutor für Universitäten und Hochschulen. OneTutor. Retrieved 2 October 2025, from https://onetutor.ai

10 Benita, J., Tej, K. V. C., Kumar, E. V., Subbarao, G. V., & Venkatesh, CH. (2024). Implementation of Retrieval-Augmented Generation (RAG) in Chatbot Systems for Enhanced Real-Time Customer Support in E-Commerce. 2024 3rd International Conference on Automation, Computing and Renewable Systems (ICACRS), 1381–1388. https://doi.org/10.1109/ICACRS62842.2024.10841586

11 Nickodem, K. (2024, May 1). Unpacking the Potential Risks of Generative AI Chatbots on Local Government Websites. Coates' Canons NC Local Government Law. https://canons.sog.unc.edu/2024/05/unpacking-the-potential-risks-of-generative-ai-chatbots-on-local-government-websites/

12 Security Risks with RAG Architectures. (n.d.). IronCore Labs. Retrieved 30 September 2025, from https://ironcorelabs.com/security-risks-rag/

13 KI-Servicestelle der RTR. (n.d.). RTR. Retrieved 16 September 2025, from https://www.rtr.at/rtr/service/ki-servicestelle/ki-servicestelle.de.html

14 Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the Protection of Natural Persons with Regard to the Processing of Personal Data and on the Free Movement of Such Data, and Repealing Directive 95/46/EC (General Data Protection Regulation) (Text with EEA Relevance), 119 OJ L (2016). http://data.europa.eu/eli/reg/2016/679/oj/eng

15 Art. 15 DSGVO – Auskunftsrecht der betroffenen Person. (n.d.). Datenschutz-Grundverordnung (DSGVO). Retrieved 7 September 2025, from https://dsgvo-gesetz.de/art-15-dsgvo/

16 Datenschutzbehörde, Ö. (n.d.). Österreichische Datenschutzbehörde. Österreichische Datenschutzbehörde. Retrieved 5 September 2025, from https://dsb.gv.at/

17 Wihlborg, E., Larsson, H., & Hedström, K. (2016). 'The Computer Says No!' – A Case Study on Automated Decision-Making in Public Authorities. 2016 49th Hawaii International Conference on System Sciences (HICSS), 2903–2912. https://doi.org/10.1109/HICSS.2016.364

18 Thapa, B. (2019). Predictive Analytics and AI in Governance: Data-driven government in a free society. The European Liberal Forum. https://liberalforum.eu/publication/predictive-analytics-and-ai-in-governance-data-driven-government-in-a-free-society/

19 Article 50: Transparency Obligations for Providers and Deployers of Certain AI Systems | EU Artificial Intelligence Act. (n.d.). Retrieved 14 October 2025, from https://artificialintelligenceact.eu/article/50/

20 Richtlinie—2024/2853—DE - EUR-Lex. (n.d.). Retrieved 15 October 2025, from https://eur-lex.europa.eu/eli/dir/2024/2853/oj/deu

21 Digital government. (n.d.). OECD. Retrieved 15 October 2025, from https://www.oecd.org/en/topics/digital-government.html

22 Künstliche Intelligenz: Eine Herausforderung für das Urheberrecht. (n.d.). Retrieved 15 October 2025, from https://www.wu.ac.at/forschung/forschungsportal/news/archiv-news/detail/kuenstliche-intelligenz-eine-herausforderung-fuer-das-urheberrecht

23 RAG-Chatbot: Technische Dokumentation. (n.d.). RTR. Retrieved 17 October 2025, from https://www.rtr.at/rtr/service/ki-servicestelle/chat/technik.de.html

24 NYC's AI chatbot was caught telling businesses to break the law. The city isn't taking it down. (2024, April 3). AP News. Retrieved 21 October 2025, from https://apnews.com/article/new-york-city-chatbot-misinformation-6ebc71db5b770b9969c906a7ee4fae21

25 Cecco, L. (2024, February 16). Air Canada ordered to pay customer who was misled by airline's chatbot. The Guardian. https://www.theguardian.com/world/2024/feb/16/air-canada-chatbot-lawsuit

26 §43 GmbHG – Einzelnorm. (n.d.). Retrieved 31 October 2025, from https://www.gesetze-im-internet.de/gmbhg/__43.htm

27 Magesh, V., Surani, F., Dahl, M., Suzgun, M., Manning, C. D., & Ho, D. E. (2025). Hallucination-Free? Assessing the Reliability of Leading AI Legal Research Tools. Journal of Empirical Legal Studies, 22(2), 216–242. https://doi.org/10.1111/jels.12413

28 NewsBites Volume XXVII – Issue 45, June 13, 2025. (2025, June 13). SANS Institute. https://www.sans.org/newsletters/newsbites/xxvii-45

29 Tsai, L., & Tsai, P.-J. (2025, May 29). Generative AI Copyright Lawsuit: RAG Technology Once Again in Focus as News Publishers Sue Cohere. Lexology. https://www.lexology.com/library/detail.aspx?g=7a1f5a5b-0274-4fdc-8135-7bb5a80671f3

30 Kelbert, T. H., Dr Julien Siebert, Patricia. (2024, May 13). Retrieval Augmented Generation (RAG): Chat mit eigenen Daten. Fraunhofer IESE. Retrieved 5 November 2025, from https://www.iese.fraunhofer.de/blog/retrieval-augmented-generation-rag/

31 Alrabie, L., & Andolina, S. (2025). Towards Human-Centered RAG: A Study of AI-Supported Testing Practices in Italian Public Administration. Proceedings of the 16th Biannual Conference of the Italian SIGCHI Chapter, 1–6. https://doi.org/10.1145/3750069.3750103

32 Article 4: AI literacy | EU Artificial Intelligence Act. (n.d.). Retrieved 5 September 2025, from https://artificialintelligenceact.eu/article/4/

33 Thapa, B. (2019). Predictive Analytics and AI in Governance: Data-driven government in a free society. The European Liberal Forum. https://liberalforum.eu/publication/predictive-analytics-and-ai-in-governance-datadriven-government-in-a-free-society/

34 Santiso, C. (n.d.). Public Governance in the Age of Artificial Intelligence. Retrieved 1 July 2025, from https://www.chandlerinstitute.org/governancematters/public-governance-in-the-age-of-artificial-intelligence

About the ALAIT Project

The Austrian Lab for AI Trust (ALAIT) is a research and development project funded by the Austrian Federal Ministry for Innovation, Mobility and Infrastructure (BMIMI) aimed at building trust through knowledge in the field of artificial intelligence (AI). The ALAIT project aims to empower interested parties and key societal groups to use AI technologies responsibly and to establish ethical and high-quality standards for the use of AI.

The project is led by **winnovation** (Gertraud Leimüller and Lena Müller-Kress) and is being carried out in collaboration with **leiwand.ai** (Rania Wazir and Silvia Wasserbacher-Schwarzer), **TU Wien** (Sabine Köszegi and Ilya Faynleyb) and A**ustria Presse Agentur – APA** (Verena Krawarik and Sophia Marecek).

The ALAIT dossiers are available on the project website: https://science.apa.at/project/alait/

The contents of this dossier reflect the current state of the art and have been carefully compiled in accordance with scientific criteria. However, they do not constitute legally binding information or advice.

Imprint

Media owner and publisher:
winnovation consulting gmbh
Linke Wienzeile 124/5
1060 Vienna

Published in 2026, funded by the Austrian Federal Ministry for Innovation, Mobility and Infratructure.

Bundesministerium Innovation, Mobilität und Infrastruktur

winnovation

leiwand.ai

TU WIEN

APA

Acknowledgements:

We would like to thank the following experts for their helpful feedback on earlier versions of this dossier: Thomas Schreiber, Brigitte Krenn, and Martin Berninger.