EPiC Series in Computing

Volume 107, 2025, Pages 237–246

Proceedings of EUNIS 2025 annual congress in Belfast

Implementation of an AI support chatbot based on Microsoft Azure OpenAI with special consideration of quality

Bernd Decker[1*], Sarah Grzemski[1†] and Ingo Hengstebeck[1‡]

[1] RWTH Aachen University, Germany

decker@itc.rwth-aachen.de grzemski@itc.rwth-aachen.de, hengstebeck@itc.rwth-aachen.de

Abstract

This article describes the development and evaluation of an AI powered chatbot that was developed specifically to improve IT support at the IT-ServiceDesk (IT-SD) within the IT Center of RWTH Aachen University. The implementation is carried out using Microsoft Azure OpenAI and a Retrieval-Augmented Generation (RAG) approach. Given the varying complexity of support requests among different customer groups—namely, students, employees, and IT administrators—the AI chatbot will be a valuable supplement to established support channels such as email, telephone, chat, ticket portals, and personal interaction.

The article describes the key requirements for the chatbot and explains how the quality of the response is ensured by a structured feedback system. It further addresses the challenges that arise from the fact that there are two different scenarios with different requirements. On one hand, standardized and publicly available RWTH support content is provided for RWTH members as well as interested parties of the IT services and, on the other, specific internal support content for the supporters at the IT-SD. This necessitates that specific measurable quality requirements and criteria must be identified implemented and adapted.

* https://orcid.org/0000-0002-9627-5695
† https://orcid.org/0000-0002-3946-4780
‡ https://orcid.org/0009-0006-0592-4925

L. Desnos, R. Vogl, L. Merakos, C. Diaz, J. Mincer-Daszkiewicz and S. Mclellan (eds.), EUNIS 2025 (EPiC Series in Computing, vol. 107), pp. 237–246

1 Introduction

Since 2010, the IT-SD has been the single point of contact for the services of the IT Center at RWTH Aachen University (Bischof et. al. 2011) for students, employees and partners handling all inquiries and fault reports, hereinafter referred to as inquiries. The IT-SD is part of the Service & Communication department (SeKo). Its portfolio ranges from all services relating to identity management and the student life cycle management to high-performance computing, research data management and generative AI.[§] Due to this comprehensive and continually expanding service spectrum, the demand for user support has grown notably. The number of provided services has risen steadily over the years (2010: 7, 2016: 50, 2025: 109).

Consequently, the number of inquiries has risen steadily in recent years. In 2024 alone, the department processed a total of 64,359 inquiries from users of all IT Center services. This significant growth of service requests can be attributed to several circumstances. In addition to the steady increase in student enrollment (2010: 32,240; 2024: 44,892[**]) and the growing number of employees (2010: 7,992; 2024: 10,306), the IT landscape itself also grew. Not least due to the COVID-19 pandemic from spring 2020, the demand for support for services such as video conferencing systems and collaborative systems also increased steadily. The strong growth in the international student body from 1,259 in 2010 to 15,270 students in 2024 contributed to an overall higher support volume as well as the fact that support staff increasingly received requests in English.

In addition to the already established communication channels, telephone, email, in person and ticket portal, chat support was established as a further channel in 2015. The chat support can be accessed via, the website (www.itc.rwth-aachen.de), the IT Center's public documentation (help.itc.rwth-aachen.de) and through several applications. (Hengstebeck et. al. 2016)[††]

The public documentation for users consists of instructions and information on the IT Center's entire portfolio of services.[‡‡] There are also offers from various specialist departments that provide additional information in their own wikis or landing pages.[§§]

Furthermore in 2016, a certified (ISO 9001:15) Quality Management System (QMS) for 1st level support was established to ensure high quality user support for the ever-growing range of services (Pieters 2017). Today, the QMS is based on two core processes (support and marketing) with corresponding management and assistance processes. The handling of generative AI is currently based on the support core process.

Despite the wide range of support channels, inquiries via e-mail or telephone frequently reach the IT-SD outside regular opening hours (i.e. working days between 5 p.m. and 8 a.m. or at the weekend or on public holidays). Analysis of the IT-SD reports have shown that in 2024 a total of 1,271 telephone inquiries (compared to 652 in 2023) were made outside of regular opening hours. In the same period, the number of e-mail inquiries received beyond standard opening hours amounted to 9,153 (compared to 7.386 in 2023).

If a customer sends an inquiry by email on Friday evening at 8 p.m., this inquiry will not be answered until Monday morning. Although most customers (RWTH members & others) seem to understand these operating constraints, a faster response - even outside opening hours - can significantly increase customer satisfaction (Sheth, A., et. al. 2020). Moreover, while the most of these inquiries can be resolved using predefined standard replies, this approach entails two primary disadvantages. First, customers must wait a relatively long time for what is usually a simple answer. Second the time

[§] https://www.itc.rwth-aachen.de/cms/it-center/~rcyys/Services/.

[**] https://www.rwth-aachen.de/cms/root/die-rwth/profil/~enw/daten-fakten/

†† In this support chat, two people communicate with each other in writing, with no additional technical systems (e.g., chatbots) involved.

‡‡ https://help.itc.rwth-aachen.de/.

§§ Examples include https://about.coscine.de/ and https://www.hpc.itc.rwth-aachen.de/go/id/bcenat/.

allocated by employees to address these routine inquiries could be more effectively redirected toward managing more complex or pressing issues.

In order to systematically enhance the quality of its services and support channels, the IT Center conducts annual surveys. The findings from these surveys underline users' demand for continuous, 24/7 support availability.

In light of these circumstances, two overarching research questions emerged:

1. How can waiting times for user inquiries be minimized and service accessibility be extended beyond standard operating hours to further enhance customer satisfaction?
2. In what ways can the burden of addressing frequently recurring standard queries on departmental staff be systematically alleviated?

Consequently, these key questions formed the basis for the early evaluation of potential chatbot solutions aimed at optimizing the IT-SD's support processes.

2 Design and implementation of SeKoGPT

2.1 Preliminary initiatives and key findings

In order to deal with the large number of inquiries and increase customer satisfaction, a project group from SeKo explored the opportunities offered by chatbots in customer support back in 2020. To this end, discussions were held with other universities and the market was analyzed, e.g. through visits to the CallCenterWorld. Initial tests were also carried out with Microsoft Power Virtual Agents. However, it turned out that implementing a chatbot is very time-consuming and requires regular manual maintenance. Due to a lack of resources, work on the project to implement a chatbot with Power Virtual Agents was greatly reduced. An intensive conceptual phase followed.

In parallel the following key functional requirements for the AI-driven chatbot were identified. The staff's prior experience with chat-based support proved highly beneficial in this regard.

- The chatbot must integrate both internal and external documentation to establish a comprehensive knowledge base.
- It must be capable of interpreting incomplete or inaccurately formulated user inquiries and, if necessary, request clarifications.
- It must operate with high availability: 24 hours a day, 7 days a week
- The user experience should be intuitive and pleasant
- The chatbot must deliver fast and reliable responses
- Answers must comply with certain ethical principles and should not be discriminatory or offensive

Together, these requirements are intended to address the identified questions by reducing response times and optimizing staff resource allocation.

2.2 Opportunities arising from ChatGPT

The public presentation of OpenAI's Large Language Model (LLM) GPT on November 30, 2022, gave new impetus to the discussion on the use of AI-based chatbots in customer support. Many of the requirements set out above are already inherently addressed by the GPT model.

- The model exhibits a robust capacity to handle incomplete or imprecisely formulated user requests. It can prompt users for necessary clarifications, thereby ensuring that it accurately interprets and addresses their inquiries.

- By leveraging a scalable and resilient API infrastructure the GPT model offers high availability (24/7 operation) and rapid response times. This directly addresses the requirement for continuous service access and high-speed performance.
- The conversational design of the GPT model facilitates intuitive interactions. The natural language processing capabilities enable it to deliver responses that are both user-friendly and contextually appropriate, enhancing overall customer satisfaction.

Collectively, these attributes demonstrate that the GPT model effectively fulfills many of the predetermined requirements for the IT Service Desk chatbot. Consequently, adapting the GPT model to establish a custom SeKoGPT is a viable approach, as it directly addresses the essential functional and operational demands identified during the conceptual phase.

Consequently, the initial inquiry was refined to address the following question: By what means can a GPT model be adapted to function as the foundational framework for our SeKoGPT?

2.3 Decision for Microsoft Azure OpenAI

During the subsequent evaluation of potential AI service providers, Microsoft Azure OpenAI emerged as the most suitable platform for several reasons. First, Azure's stringent data sovereignty measures allow hosting and processing strictly within data centers located in Europe, thereby assuring compliance with GDPR regulations. This feature proved pivotal for an institution that handles sensitive user information and adheres to strict privacy mandates. Second, the platform's terms guarantee that prompts and user-generated data are not employed for additional model training, aligning with internal policies and alleviating concerns about data exploitation. Third, the institution had already secured the necessary licenses and obtained formal approval from its Data Protection Officer, ensuring a swift and compliant implementation process. Further bolstering the decision, Azure provides a flexible environment where Retrieval-Augmented Generation (RAG) can be configured with relative ease through its integrated tooling and comprehensive developer resources. This combination of technical convenience, regulatory alignment, and existing administrative clearance ultimately expedited the evaluation phase and laid the groundwork for a rapid proof of concept.

2.4 Implementation and RAG approach for a customized SeKoGPT

In August 2023, access to an early version of GPT within Microsoft Azure OpenAI led to an intensive examination and evaluation of the functionality and usability: an internal team of developers, supporters and testers intensively explored the possibilities. Based on the insights gained, the first prototype of a chatbot utilizing Azure OpenAI was developed in December 2023. A specially tailored export of the complete internal and external documentation of IT-SD served as the central knowledge base, containing all relevant domain-specific information. To efficiently leverage this rich data source, a Retrieval-Augmented Generation (RAG) approach was implemented. This method integrates the generative capabilities of the GPT model with targeted retrieval of relevant documents, thereby enhancing the precision and contextual relevance of the responses generated.

The AI was configured to respond exclusively to queries related to the content of the stored documents. These - and only these - then form the basis on which the AI generates answers.

It is possible to make the settings for the GPT model via Azure AI. The parameter temperature was particularly important for our purposes. Temperature controls the "creativity" of the AI when answering. As we want the answers to relate exclusively to our stored documents, we have set a value of "0" here. This ensures that the AI responds to the same/similar queries and does not hallucinate.

In addition to the notable benefits- such as improved efficiency, speed and accessibility, -, it is essential to keep an eye on the associated risks: Answers from AI must always be critically scrutinized, as even if they are generally reliable, they can be incorrect or misleading due to the fact that the AI's

pattern-based nature prevents a genuine understanding of factual contexts and real-world complexities. However, by integrating a feedback system for users, the behavior of the AI can be observed and improved.

The findings from the evaluation phase led to concrete considerations in 2024 as to how an AI chatbot based on Azure OpenAI can be integrated into the existing infrastructure of the IT-SD. From February 19, 2024, two prototypes were therefore evaluated by the project group:

- a prototype for the employees of the SeKo department: SeKoGPT for SeKo supporters
- a prototype for users of IT Center's services: SeKoGPT for RWTH members &others

2.5 SeKoGPT for SeKo supporters and RWTH members & others

A distinction is now made between two use cases, to which a separate AI instance with different settings is assigned: SeKo supporters and RWTH members & others. What both cases have in common is an increase in quality when answering support requests and an improvement in support documentation.

Figure 1: View for SeKo supporters

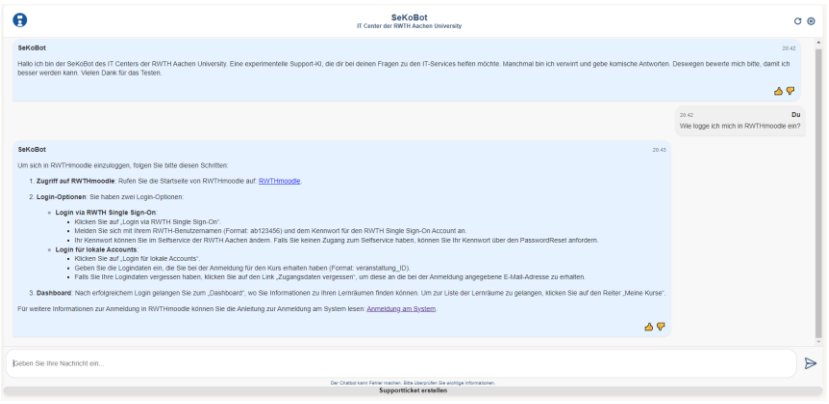

Figure 2: View for RWTH members & others

The instance for SeKo supporters can only be used by employees from the SeKo department. Here, SeKo supporters can be assisted by the AI in answering support requests. As additional content from the internal knowledge database and public documentation is made available to the AI in this instance, recommendations for action for users and quick guides can be created on the fly. It is also possible to summarize complex support requests and facilitate formulate multilingual requests. This should increase the efficiency, consistency and quality of support and reduce the overall workload for SeKo supporters.

The instance for RWTH members and other users is accessible to anyone with inquiries regarding the IT Center's services or who requires assistance. SeKoGPT provides users with an alternative to the existing support chat and is available outside regular opening hours. Frequently recurring standard questions are answered automatically. If the help provided by SeKoGPT is not sufficient, future plans include an automatically email to open a ticket in the IT-SD ticket system. This ticket will then be further processed by a SeKo supporter. Alternatively, users will have the option to be redirected from SeKoGPT to the regular support chat to discuss their concerns directly with a personal SeKo supporter. Due to the significant costs and technical complexities involved, there are currently no plans to integrate the chatbot with the RWTH Aachen University's telephone system. However, customers will have the possibility to leave a callback request.

Both SeKoGPTs run - analogous to the RWTHgpt[***] launched on July 16, 2024 - in their own encapsulated GPT instance, which is provided via Microsoft Azure OpenAI. The schematic structure and interaction of the two instances can be seen in the following figure:

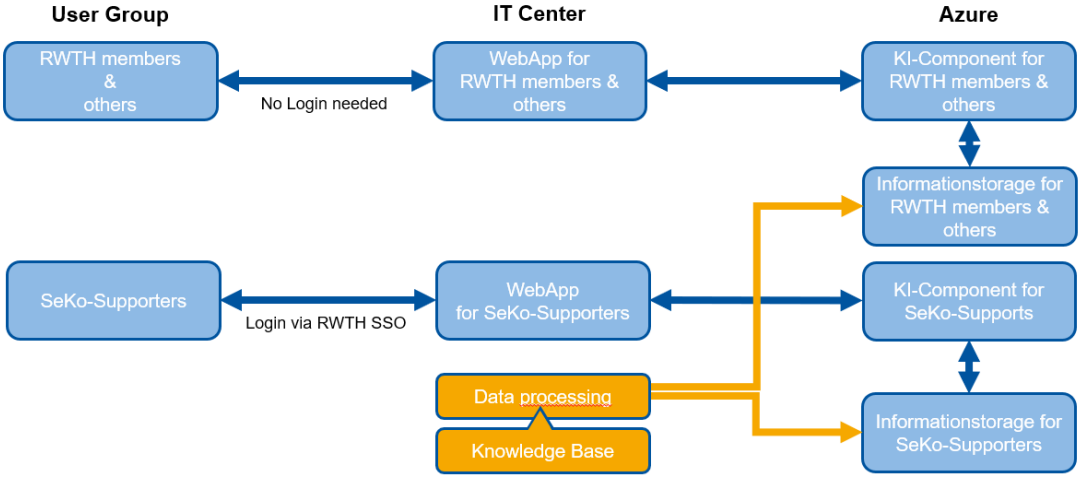

Figure 3: Structure of SeKoGPT

The corresponding servers are located within the EU. Supporters access the WebApp using the RWTH single sign-on (SSO), which provides them with a personal chat history to retrieve information from previous interactions. In contrast, other customers do not require a separate login because they do not utilize a chat history. This approach lowers the hurdle for customers to initiate a chat.

As already described a central point, SeKoGPT is based on the RAG functionality. The AI uses defined content from the knowledge database to answer questions. Only the stored data is taken into account for questions and only questions about IT Center services are answered.

[***] https://cls.rwth-aachen.de cms/cls/services-und-projekte/~bilrqx/rwthgpt

This has the advantage that the chatbot does not mix services from other providers with offers from other service providers. On the other hand, the response behavior can be controlled based on the stored documents and their structure.

To ensure that SeKoGPT provides targeted, precise and polite answers, this is defined accordingly in the configuration. In particular, the so-called "system message" plays a pivotal role in guiding the AI's behavior., SeKoGPT is informed via this directive that it possesses extensive expertise in IT support and must therefore articulate its responses with clarity and precision, ensuring they remain accessible and comprehensible. In essence, the system message delineates explicit parameters regarding response accuracy, subject-specific acumen, and appropriate stylistic nuances. It should be noted that a separate system message was created for each of the two instances in order to cover the different requirements.

SeKoGPT is told, among other things, that it is an AI with extensive knowledge of IT support and that it should formulate the answers precisely and in a way that is easy to understand.

What is not desired is for SeKoGPT to come up with answers, as Tyson (2023:3099) states. Or to combine aspects that have nothing to do with each other. The approach of "feeding" the AI with its own documents is a first protective measure alongside the "system message".

2.6 Testing and Evaluation

Hyeon and Park (2023) demonstrated that employees' perceived intelligence of ChatGPT is positively associated with their willingness to use AI for other purposes. This can certainly be transferred to other groups of people - such as the two target groups in our case. This suggests that better or more reliable answers (confidence score) can lead to greater adoption and increased use. To ensure optimal response accuracy and minimize fallback rates (inappropriate responses), it was necessary to test both applications very intensively. It is also necessary to analyze the extent to which the GPTs detect the intention of the question (intent scores) and the emotional tone (sentiment analysis) in the formulated question (Skodowski 2023).

At its core, a permanent group of several SeKo support employees is responsible for testing the two bots. They use potential and real RWTH members inquiries to check whether the answers given by the AI, are correct and complete. The other employees of SeKo are also asked to test the two bots at regular intervals. All feedback was recorded in the initial phase via a SharePoint list and was used to document incorrect answers to ultimately improve the response behavior by adjusting the stored information and the system message.

The results from these tests are fed back into the implementation and configuration processes, allowing SeKoGPT's configuration to be adapted step by step using the feedback from each round. In addition to instructions on the response style and wording, SeKoGPT also receives direct instructions on how to refer to the external documentation (at help.itc.rwthaachen.de) or how the AI should ask questions in the event of ambiguities. Just as important as what SeKoGPT should do is telling the AI what it should not do. For example, no information should be used that goes beyond the IT center services and their documentation. This results in a cycle of implementing/configuring, testing and providing feedback.

3 Quality assurance through feedback

During the test phase, the issue of ensuring the quality of the chatbot's responses during operation became apparent. Consequently, the decision was made to implement a simple feedback system in both SeKoGPT solutions.

SeKo supporters and RWTH members & others can provide feedback at any time, even during operation. This option is available for every response from the chatbot. A lean feedback system in the form of "thumbs up" and "thumbs down" has been implemented for this purpose. In the administration view, the submitted feedback is displayed in chronological order for review. For data protection reasons, the feedback and the associated chat histories are stored without reference to individuals. In concrete terms, this means that it is not possible to trace who asked the question and provided feedback

The administrators, quality management representatives (QMB) or other designated persons review the feedback. Based on this input and the associated chat histories, the underlying knowledge base or the system messages are adjusted accordingly

For example, it quickly became apparent that the SeKoGPT for SeKo supporters does not provide them with any standard information that is necessary for processing a request. As a result, an additional document was created that specifies which standard information is required for each service. Subsequent tests have shown that the revised approach enables the bot to answer corresponding questions correctly.

Another example concerns the use of terminology. Internally, the IT Center refers to its high-performance computer as “Cluster”, whereas the documentation uses the term “high-performance computer.” When users asked the chatbot how to use the “Cluster”, it generated inconsistent responses. By adjusting the system message to define "Cluster" as a synonym for "high-performance computer," the response behavior was improved.

A classic PDCA cycle (plan, do, check, act) has therefore been implemented to ensure continuous improvement of responses. For quality management, the feedback function is one of the core functions for enhancing the existing documentation and stored information. This means: (1) the feedback must be easy for the customer to provide, (2) it must be universally understandable, (3) it must be clearly visible.

Currently, the evaluation of submitted feedback is still done manually. However, in the future, artificial intelligence is planned to be used to enable automation.

For example, AI can be used to analyze negatively rated responses to determine whether certain question clusters consistently receive negative feedback. This information can then be used to make targeted adjustments to documentation, the system message, or supplementary materials.

Additionally, it can be determined which questions are asked most frequently. From this, it can be inferred which IT Center services may not be intuitive to use or for which services the documentation needs to be revised. This could also be achieved by analyzing requests in the ticketing system. Although similar insights could be gleaned from ticket system data, the feedback system provides more structured data that supports effective AI-based analysis.

It is also planned to automatically create FAQ documents from the positively rated questions and answers, which will also be stored in the documentation in the same way as the exports.

4 Outlook and future work

To determine whether the benefits of SeKoGPTs for the needs of the various target groups described in the article are being achieved, appropriate KPIs must be analyzed, among other things. Kakur et. al. (2024) points out the importance of defined KPIs, which should be presented in a structured and accessible analysis dashboard. A clear subdivision between “general KPIs,” “technological chatbot performance,” and “user behavior” is suggested. It is important to bear in mind that such a tool will be subject to constant expansion and adaptation. The extent to which such a tool can be provided for the SeKoGPTs will be evaluated shortly by the core project group.

However, to assess the acceptance of the two bots and the described benefits for customers and employees, the thumbs-up and thumbs-down functions have already been implemented. In addition,

starting in 2025, both applications will be included in the annual satisfaction surveys conducted by the IT Center and SeKo target groups.

Furthermore, future work will examine, among other things, how Azure OpenAI can be integrated into the IT Center's ticketing tool (helpLine) as well as into the existing chat support. This integration will address issues such as the automatic creation of support tickets from AI-based interactions when further clarification is required.

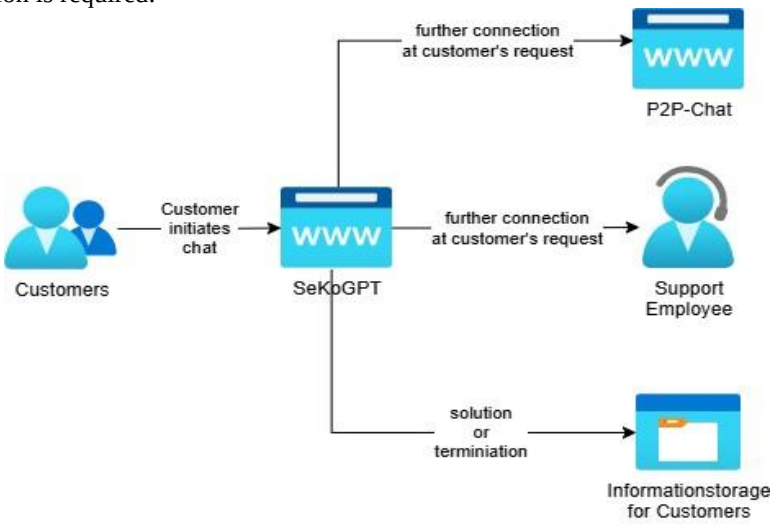

Figure 4: Further connection at customer support

In summary, the current implementation of the two SeKoGPT solutions is only a first step toward a more extensive use of AI in customer support (SeKo). As outlined above, numerous possibilities remain open regarding how AI can support and improve general customer support, benefiting both support staff and customers. However, to ensure the highest possible user acceptance, it is necessary to establish regular quality assurance measures to guarantee that the chatbot consistently provides correct and helpful answers. In the long term, ensuring defined and automated checks of the underlying documentation on which the chatbots are based, as well as automated monitoring of the documentation and the chatbots' response behavior using text mining and sentiment analysis, will be another challenging aspect for the future. Ultimately, only chatbots that generate friendly, accurate, and useful answers that genuinely assist users will be accepted by the user community and thus achieve the goal of relieving the burden on employees.

5 References / Citations

Bischof, C., Grzemski, S., & Hengstebeck, I. (2011). Introduction of a Service Desk at the Computing and Communication Center of RWTH Aachen University. A practical report. In A. Degkwitz et al. (Eds.), Process-oriented university [General aspects and practical examples] (pp. 181–198). Herchen Publishing House.

Grzemski, S., & Hengstebeck, I. (2017): Future challenges for quality-assured IT support though cooperative structures. In Shaping the Digital Future of Universities. Book of Proceedings Eunis 23rd Annual Congress, 6. https://doi.org/10.17879/21299722960

Hengstebeck, I., & Grzemski, S. (2016). New ways of customer support at the IT ServiceDesk of the IT Center of RWTH Aachen University. In H. C. Mayr et al. (Eds.), Lecture Notes in Informatics (LNI) (pp. 933–945). Society for Computer Science.

IT Center (RWTH Aachen University). (n.d.). IT Center Help. Retrieved from https://help.itc.rwth-aachen.de/

IT Center (RWTH Aachen University). (n.d.). Umfragen Retrieved from https://www.itc.rwth-aachen.de/cms/it-center/IT-Center/Engagement/~gpbx/Umfragen/

Jo, H., & Park, D.-H. (2023). AI in the Workplace: Examining the Effects of ChatGPT on Information Support and Knowledge Acquisition. International Journal of Human-Computer Interaction, 40(23), 8091–8106. https://doi.org/10.1080/10447318.2023.2278283

Kakur, R., Ritz, H., Hohmann, P. (2024). Analyse-Dashboard mit relevanten Kennzahlen für einen KI-basierten Chatbot im Hochschulbereich. In C. Müller et al. Anwendungen und Konzepte der Wirtschaftsinformatik (pp. 95-96). https://akwi.hswlu.ch/issue/view/610/688

Meyer von Wolff, R., Heuzeroth, T., Hobert, S., & Schumann, M. (2020). The Students' View on IT-Support Chatbots at Universities - A Case-based Pilot Study. AMCIS2020 Proceedings. 7.

Pieters, M., Hengstebeck, I., Grzemski, S. (2017). „Einführung eines zertifizierten Qualitätsmanagementsystems im IT-ServiceDesk des IT Centers der RWTH Aachen University“, In P. Müller, et al. Lecture Notes in Informatics (LNI) (pp. 77–87). Society for Computer Science. https://dl.gi.de/collections/08e1b8cd-ebb4-486f-9d84-557ee01bf50a

Sheth, A., Anantharam, P. and Thirunarayan (2020): “Challenges and Opportunities fot AI in Customer Support”, IEEEIntelligent Systems, Vol.35, No5, 16-22

Skodowski, M. (2023) 20 Chatbot KPI´s – Wie der Erfolg virtueller Assistenten gemessen wird. BOT frinds GmbH (Eds.). https://botfriends.de/blog/chatbot-kpi

Tyson, J. (2023). Shortcomings of ChatGPT. Journal of Chemical Education, 100(8), 3098–3101. https://doi.org/10.1021/acs.jchemed.3c00361

RWTH Aachen University. (n.d.). Facts and figures. Retrieved from https://www.rwth-aachen.de/cms/root/dierwth/profil/~enw/daten-fakten/

RWTH Aachen University. (2024, December 27). RWTHgpt and other AI systems. https://cls.rwth-aachen.de/cms/cls/services-und-projekte/~bilrqx/rwthgpt/.

6 Author biographies

Sarah Grzemski M.A. studied Economic Geography, Economics, and Geography, earning her Master's degree from RWTH Aachen University in 2002. Until 2007, she worked as a research assistant in the Department of Economic Geography of Services. Since then, she has been with the IT Center of RWTH Aachen University. In 2010, she became the division head of the IT-ServiceDesk, responsible for first-level IT support. Following organizational changes, her role expanded, and the department was renamed Service & Communication. She now oversees digital (web, social media) and analog (posters, flyers) external presentations, the RWTH printing service, user surveys on IT services, and the IT administration of the IT Center. (CRediT: Conceptualization, Investigation, Writing – original draft)

Dipl. Inform. Bernd Decker is deputy head of the Department “Process Management and Digitalization in Learning & Teaching” at the IT Center of RWTH Aachen University since 2011. From 2006 to 2009, he worked at the IT Center as a Software Developer, and since 2009 he has is leading the development group. His work focuses on IT solutions for processes in the fields of Learning Management Systems, E-Services, and Generative AI. (CRediT: Investigation, Writing – original draft)

Ingo Hengstebeck M.A. studied Technical Communication. He received his Master’s degree from RWTH Aachen University in 2009. Until 2009, he worked as an employee at the IT Center. Since 2014 he has been the deputy division head of the IT-ServiceDesk. His work is focused on quality management, process management, and communication in the field of user support. (CRediT: Conceptualization, Investigation, Project Administration, Supervision, Writing – original draft)