

# Linguistics to LLMs: Teaching with and about Chatbots

**Ulrike Padó**

Hochschule für Technik Stuttgart  
Schellingstr. 24, 70174 Stuttgart  
Ulrike.Pado@hft-stuttgart.de

**Barbara Pampel**

University of Konstanz  
Universitätsstr. 10, 78464 Konstanz  
Barbara.Pampel@uni-konstanz.de

## Abstract

LLM-based methods supersede many approaches in NLP at high velocity, making it necessary to adapt curricula. We argue that this effort also presents a chance to integrate LLM chatbots as learning support. We demonstrate (a) how we re-conceptualized an existing class segment on digital assistance systems to discuss LLM-based chatbots, (b) how we created a specialized instructional chatbot as a demonstrator that students could directly use for learning and revision and (c) how students' initial perception of LLM-based AI changed due to instruction.

## 1 Introduction

In the public understanding, Artificial Intelligence (AI) has come to be near-synonymous to Large Language Models (LLMs). This is both a huge gain in visibility for language modeling, formerly mostly a technical part of Computational Linguistics and Natural Language Processing (NLP), and a huge challenge to universities to teach the basics of LLMs, their use in NLP, and the wider field of AI to as large an audience as possible.

Often, this challenge is met in the context of existing classes on AI or NLP, requiring updates and changes to the class that can feel difficult to make without disrupting the narrative of the whole class. However, these alterations also open up the opportunity to integrate LLM-based tools such as chatbots as both a demonstration of the capabilities of LLMs and a welcome didactic tool. Giving students the ability to explore and experiment with the object of study and to further use it as a tool for self-study adds a new didactic dimension to the class and additionally deepens students' understanding of LLM-based AI methods.

We describe one such class update and focus here on how to adapt a section originally covering Digital Assistants to a discussion of LLM-based chatbots. We aim to give a relatively high-level

overview over our didactic goals and mechanisms while at the same time providing detailed information on the teaching contents and the implementation of our chatbot in order to concretely support class adaptations in related contexts.

Our overall teaching goal in the class is to give students an understanding of both traditional machine learning and LLM-based approaches. They should be able to choose the best method given a problem setting and take away enough technical understanding of LLMs to form an intuition of their abilities and limitations.

To reach this goal, we offer both a theoretical discussion of the technical underpinnings of LLM-based chatbots (see Sections 3.1 to 3.3) as well as a tutorbot, i.e., a live chatbot specialized on the course content (Section 3.4). In this way, we offer students a multi-dimensional experience of the tool as well as a new channel of practical support for learning and revision.

Concretely, we choose the approach of Retrieval-Augmented Generation (RAG, [Lewis et al. 2020](#)) to define our tutorbot. From the perspective of Computational Linguistics and wider AI, RAG offers a technically simple way to achieve a hybrid system that combines a well-defined, curated knowledge base with an LLM's unparalleled ability to generate fluent, contextually appropriate text. This makes the approach valuable teaching content, as it complements the discussion of the strengths and weaknesses of LLMs.

To gauge the effect of the students' learning, during the first iteration of the revised class we probed their understanding of AI concepts and their conceptualization of LLM-based AI before and after instruction and found intriguing changes (Section 4).

## 2 Tutorbots in Education

LLM-based chatbots are rapidly gaining popularity in educational contexts, for example as tools

for self-study when instructors are not available. However, there are valid concerns about the vulnerability of such chatbots to prompt injection (Perez and Ribeiro, 2022) and about the extent of their tendency to hallucinate (Maynez et al., 2020), as documented in studies involving both students and educators (Marczuk et al., 2025; Pampel et al., 2025). In educational settings, another problem is the growing dependence on support systems that may lead to bypassing critical thinking processes (Abbas et al., 2024; Süße and Kobert, 2023).

Our previous work shows how to set up an LLM-based chatbot that is specialized on and constrained to the content of a specific class (Pampel et al., to appear). We use the Retrieval-Augmented Generation technique (RAG, Lewis et al. 2020; Fan et al. 2024). When configuring a RAG-chatbot (see Figure 1), an educator can provide a curated *knowledge base*, against which every *user query* is matched to *retrieve* relevant *chunks* of information as *context*. Together with a *system prompt* that can didactically align the chatbot’s behavior, this relevant information is sent to the LLM, which is then only verbalizing the answer. This leads to more accurate, context-appropriate responses while reducing the likelihood of misleading answers and hallucinations (Shuster et al., 2021), which are especially critical when learning a new topic. The RAG approach can also help repel manipulation attempts through prompt injection (Yu et al., 2023). For instance, we find that our chatbot’s answers are highly reliable (95% of 328 user-bot interactions were answered as intended, Pampel et al., to appear).

By carefully tailoring the system prompt, a RAG-chatbot can take on different didactical roles within specific teaching and learning scenarios. Lauber et al. (2025), for example, distinguish three distinct roles and illustrate them through exemplary implementations. The careful didactic alignment of the bot to the goals of the class can help foster students’ self-learning skills (Chang et al., 2023) and thereby alleviate fears of skill loss through chatbot use.

Hence, RAG is a very promising concept for educational settings with a growing number of projects adopting it (Swacha and Gracel, 2025). Li et al. (2025) also give an overview of current practical applications of RAG in education, with educational chatbots among more complex systems like adaptive learning systems and educational content development and assessment.

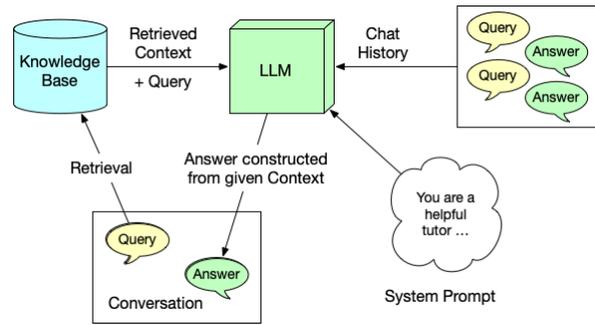


Figure 1: Retrieval-Augmented Generation for Dialogue (simplified, from Pampel et al. to appear)

### 3 Course Setup

We report on a class on Artificial Intelligence (AI) that is part of the mandatory curriculum in a seven-semester bachelor’s program in Business Informatics at Hochschule für Technik Stuttgart, a German University of Applied Sciences. The class is taught in German and runs for 14 weeks in the second year of studies, after the introductory phase of the program, but before the electives and a mandatory semester-long internship.

The overall goal of the class is to give a broad introduction to topics from AI that are relevant to Business Informatics. We focus on the three-week portion of the class that introduces human and machine dialogue. Originally, the class introduced rule-based chat and assistance systems (cf. Knebusch and Padó 2023), but we now discuss LLM-based dialogue and explain their technical background as well as vector-based information representation (as in embeddings), while also highlighting the energy consumption of LLM training and bias in LLMs. With these changes, we hope to enable students to understand and critically evaluate the use of LLMs in their professional lives. The revised class section ties in with an introduction to machine learning and a section on explicit knowledge representation (both through the discussion of implicit LLM knowledge and the integration of explicit knowledge in LLM dialogues through the RAG approach). The final topics covered in the class are constraint satisfaction and inference.

The class is taught using the Computer-based Learning approach (CBL, Knebusch et al. 2019). This method offers traditional blended learning content such as instructional videos, on-line self-tests and exercise sheets for self-study integrated into in-presence lecture times. Typically, a session will be opened by the lecturer with an introduction into

the topic of the day and closed with a short wrap-up of what was covered. In between, students work at their own pace, alone or in groups (depending on the activity), and additional live input from the lecturer may be given whenever necessary. The role of the lecturer becomes more focused on individual explanations and discussions that benefit students at all levels of ability.

In order to further expand the opportunity for asking individual questions and clarifying the course content (especially outside of class, for example during exam preparation), we provide an LLM-based tutorbot that is specialized on the content of the class (across all weeks). In this way, through the tutorbot students gain additional opportunities for verifying their understanding of the class materials, and have hands-on access to the exact chatbot that is discussed in class.

In order to facilitate replication in related classes, we now go on to discuss the content of each week in detail, followed by an in-depth description of the tutorbot specification.

### 3.1 Week 1: Chatbots

In Week 1, the goal is to introduce human dialogue behavior first in order to foster a closer understanding of the complexity and some of the rules of this domain, and then move on to machine dialogue to show how they can be simulated by a machine. This week is closest to the original course materials on human dialogue and automated assistant dialogue (cf. [Knebusch and Padó 2023](#)). The segments on the rule-based assistance systems have been replaced by an input on LLMs and the word prediction task for training, and a case study of a Retrieval-Augmented Generation bot (i.e., the tutorbot provided for the class) instead of a demo of an assistance system.

Table 1 shows the components of the class, where in-person and video lecture inputs are interspersed with student activities. In order to collect data for our evaluation and to foster reflection, we added a questionnaire on students' prior experience with chatbots. Lecture inputs are followed by exercises or self-tests to verify understanding. The self-tests are not graded, but feedback on wrong answers is given. As the class goes on, students work at their own speed. During the class, the lecturer actively probes and supports students' progress.

The final work item of the class is an exploration of the tutorbot. Students were presented with the tutorbot's definition in Flowise (cf. Fig. 2) and en-

couraged to experiment with the bot. In order to expand the students' set of prompt styles, some example prompts that explicitly request a specific level of detail or an example are also given.

The tutorbot was available throughout all parts of the class, not just the unit on chatbots and LLMs. Although the bulk of the interactions (280) was observed during the LLM and chatbot portions of the class, there was low but steady use also during the remainder of the class (110 interactions in total) – typically just before one of the formative tests that are scheduled roughly every four weeks. Students used the bot anonymously and were aware that their answers would be analyzed for research purposes later. They were able to opt out of research analysis, but noone did.

### 3.2 Week 2: Neural Networks

The goal for Week 2 is to give students an intuition about Neural Networks (NNs) and Deep Learning, their capabilities and the energy required to run them. We begin with the basics of Neural Networks, important parameters (like architecture or learning rate) and current applications in NLP, leading up to the Transformer architecture (see Table 2).

This week's input has been strongly modified from earlier iterations of the class. While NNs as a learning algorithm were always covered in the class and the application examples are constantly updated, a more in-depth exploration of Transformers was added. Another new aspect is the content on energy consumption.

This week relies strongly on interactive content from external web sites and up-to-date news items (that need to be replaced for future iterations). Therefore, both the exercise on the Transformer architecture and the activity on energy consumption include self-tests to encourage students to fully engage with the external materials. (The use of English-language materials is not problematic in practice in our German-language class, since the students generally have a good working knowledge of English. In case someone struggles, the use of translation tools is encouraged.)

Week 2 also features additional materials for students who want to dive deeper into the topic of energy consumption of AI applications: We link both to a demo video showing the usage signatures of locally hosted LLMs and to an off-site on-line course on energy use in AI (see Table 2).

| Content                         | Activity   |
|---------------------------------|--|
| Intro to (human) dialog         | In-presence lecture                                    |
| Own Chatbot experience          | Questionnaire  |
| Adjacency pairs, Gricean Maxims | Lecture video  |
| Applying the theory             | Paired exercise on adjacency pairs and Gricean Maxims  |
| LLMs for natural dialogue       | Lecture video  |
| Taking stock                    | Self-test with fixed feedback for wrong answers        |
| RAG                             | Video Lecture  |
| RAG applied to the class        | Tutorbot architecture, with pointers to good prompting |

Table 1: Contents in Week 1: Chatbots

| Content                               | Activity   |
|---------------------------------------|--|
| History and Perceptrons               | In-presence lecture  |
| Network architectures and parameters  | Exercise on <a href="https://playground.tensorflow.org/">playground.tensorflow.org/</a>  |
| Complex Neural Networks: Applications | Video lecture  |
| Transformer architecture              | Exercise using <a href="https://poloclub.github.io/transformer-explainer/">poloclub.github.io/transformer-explainer/</a>                                 |
| Energy use of neural networks         | Current news article   |
| Taking stock: Energy use              | Self-test on the contents of the text  |
| Additional materials on Energy Use    | Demo on self-hosting LLMs, link to relevant external on-line class ( <a href="https://open.hpi.de/courses/envai2025">open.hpi.de/courses/envai2025</a> ) |

Table 2: Contents in Week 2: Neural Networks

### 3.3 Week 3: Representing Information

The goal of Week 3 is to explain how information is stored in neural networks and look at bias in LLM output. This section is completely new, but replaces short digressions on vector space representations and bias in classifiers in an old version of the class.

Table 3 shows that students first get lecture input on vector space representations and then explore the visualization of embeddings plotted in 3d space. This exposes neighborhoods of semantically related items, for example.

Next, they move to the level of the full network and look at LLM bias revealed in sentence completion. The GenerAITor demonstrator<sup>1</sup> shows how LLM completions of a neutral sentence stub about a new graduate vary greatly depending on the gender and origin of the first name chosen for the person (according to the LLM’s continuations, men with Arabic names become terrorists, women go on to be home makers, European men have successful professional careers, and so forth).

After this demonstration of bias present in an

LLM, students watch a video on different ways of identifying bias in LLMs from the "Generative AI at MIT" YouTube channel<sup>2</sup> and finally verify their insights on bias with a self-test.

Again, in this unit, we use visualizations and demonstrators from external sources and expose the students to original materials. We want to encourage them to process foreign-language original materials and work through content that might be challenging and slightly different from what was presented in class so far, in order to prepare them for self-learning later in their studies and careers.

### 3.4 Tutorbot

The tutorbot was built to provide definitions, explanations and examples in order to help the students verify and improve their understanding of the course materials. This section provides implementational detail to facilitate the adaptation of the approach in similar contexts.

The tutorbot is configured on the low-code platform Flowise<sup>3</sup>. The use of a low-code platform

<sup>1</sup>[main.generator.ivia.ch/](https://main.generator.ivia.ch/)

<sup>2</sup>[youtu.be/Qm--\\_1M\\_Uvk?si=8aRc0ZdcsN87SySq](https://youtu.be/Qm--_1M_Uvk?si=8aRc0ZdcsN87SySq)

<sup>3</sup>[flowiseai.com/](https://flowiseai.com/)

| Content                     | Activity   |
|-----------------------------|--|
| Vector space representation | In-presence lecture  |
| Embeddings                  | Exercise using <a href="https://projector.tensorflow.org">projector.tensorflow.org</a>   |
| Bias: Sentence completion   | Exercise using <a href="https://main.generativeai.com/ivias/">main.generativeai.com/ivias/</a>   |
| Testing for bias in LLMs    | External video from the Generative AI at MIT channel ( <a href="https://youtu.be/Qm--_1M_Uvk?si=8aRc0ZdcsN87SySq">youtu.be/Qm--_1M_Uvk?si=8aRc0ZdcsN87SySq</a> ) |
| Taking stock: Bias in LLMs  | Self-test with feedback  |

Table 3: Contents in Week 3: Representing Information

allows for an easy knowledge transfer also to lecturers with non-technical backgrounds, which is a long-standing issue in the use of educational technology.

Fig. 2 shows a screenshot. The central node for conversation with the bot is the "Conversational Retrieval QA Chain" in the bottom right corner which coordinates the RAG process. It links to the chat LLM model that formulates the tutorbot's output to the user, to the Retriever that identifies relevant information chunks in the knowledge base and to memory management that controls the amount of conversational context available to the chat LLM. We recommend a relatively small context window (two bot-student turns each in our setting) to make it harder to distract the bot from its intended function by flooding the LLM's context and drowning out our prompt specifications.

We use `mistral-large-instruct`, an externally hosted chat LLM. The model is made available to us by the KISSKI project at Gesellschaft für Wissenschaftliche Datenverarbeitung mbH (GWDG), a German LLM service provider for research endeavors (Doosthosseini et al., 2025). Using an open-source LLM puts us in full control of the so-called system prompt that specifies the LLM's role and behavior.

Our system prompt (specified as an Additional Parameter in the Conversational Retrieval QA Chain node) instructs the tutorbot to act as a friendly tutor, strictly limit its answers to the provided context of the student's request, retrieved chunks and conversation history, and not to discuss its prompt (see the Appendix for the full prompt).

The retrieval arm of the tutorbot definition takes up the middle of Fig. 2. A Postgres data base holds the embedding vectors that represent the text chunks in our knowledge base and allows similarity-based retrieval. A record manager helps avoid duplicate entries. We used a locally hosted embedding model,

`jina-embeddings-v2-base-de`. The reasons were twofold: First, we had better results with an embedding model trained specifically for German, which was not available through KISSKI at the time (there is a larger choice of embedding models now), and second, we wanted to experiment with locally hosted models in order to explore model self-sufficiency.

Finally, at the leftmost edge of the figure, there is a link to the files that should be entered into the knowledge base and a definition for a chunking strategy to cut them into shorter segments. Before the tutorbot can run, a pre-processing step reads the text and PDF files, chunks them, computes embeddings for the chunks and stores them in the data base. Once an interaction is initiated by a student, the RAG process can now start as laid out in Fig. 1.

The tutorbot was provided with a knowledge base made up of the course slides and transcripts of lecture videos. The video transcripts proved a much better basis for the RAG process than the lecture slides, since there was more redundancy and context in the spoken word. Therefore, using class transcripts or fulltext notes is a clear recommendation. The materials were not pre-processed for figure and picture verbalization, but employing a Vision-Language Model to replace graphics formats with textual descriptions would be a logical and promising next step.

## 4 Students' Concepts Before and After Instruction

After re-designing the input, we were interested in finding out how students' understanding of LLM-based AI changed after instruction. In order to probe their opinions on AI concepts, we selected questions from a questionnaire developed by Mayer (2024). The questions cover technical understanding, AI and ethics and the tendency to anthropomorphize AI.

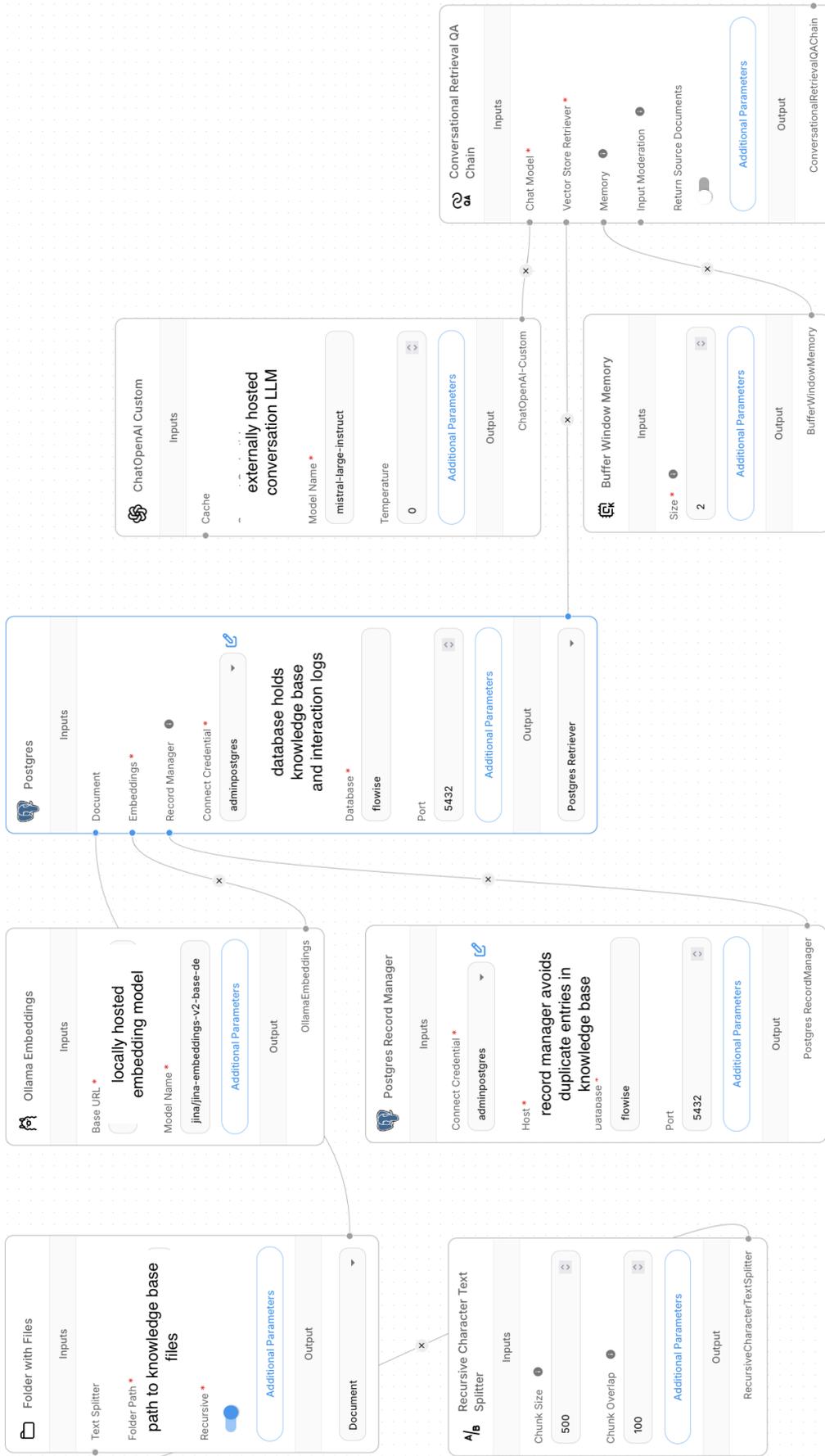


Figure 2: Definition of the course Tutorbot on the Flowise low-code platform.

Students can be assumed to be experienced chatbot users outside the class (cf. also Pampel et al., 2025 for a different group of learners): When asked about their chatbot usage habits in Week 1, 16 out of 47 students answered that they use chatbots daily, and 20 more use them several times a day (10 more students use chatbots weekly, one once a month, none never use them).

The questionnaire was given in Week 1 before instruction on AI techniques and LLMs, and again after Week 3 (not shown in Table 3). 49 students participated in Week 1 and 32 in Week 3.

#### 4.1 Before Instruction

Fig. 3 (top) shows students' replies to four questions about their prior technical knowledge about AI techniques like Deep Learning and Machine Learning. Expectedly, 40% of students say they do not know these specific AI techniques, although 50% of students agree or completely agree that they understand how AI works. 70% of students agree that LLM-based AI is not just a hard-coded program.

Fig. 4 (top) shows students' opinions about ethics and AI as well as AI characteristics before instruction. 86% of students agree that AI should be regulated by ethics guidelines, and roughly 30% of students each agree, are undecided and disagree with the statement "AI is objective", which partially explains this sentiment.

In order to break down students' answers to the multi-faceted question "Current AI is closer to a human than a computer program.", we also asked students which characteristics they assigned to AI (multiple answers were possible). Before instruction, students answered most frequently that AI is able to learn (92% of respondents) and intelligent (86%). Students also frequently felt that AI is creative (55%), independent (51%) and able to act (51%).

#### 4.2 After Instruction

After working through the three classes about LLMs and experimenting with the course tutorbot, students were asked again about their conceptualizations and impressions.

In the first batch of questions about the technical background (Fig. 3, bottom), we see a reassuring shift towards self-reported deeper technical understanding, with the majority of students now agreeing or fully agreeing that they understand what AI, Machine Learning and Deep Learning are. More

students now strongly agree that (LLM-based) AI is "not just a computer program", i.e., not rule-based. The amount of disagreement with this statement is about the same, but many students who were previously less committed to this statement now seem to be more sure.

An even greater impact of instruction can be seen in Fig. 4, which shows students' opinions on AI after instruction at the bottom. Probably due to the segment on LLM bias, almost all students are now convinced that LLM-based AI needs to be regulated by guidelines and that it is not objective. Especially the change of opinion on this last point is drastic; the ca. 30% agreement with this statement have shrunk to 10%.

On the question whether current LLM-based AI is closer to a human than a computer program, the students now shift towards agreement. To clarify the origin of the shift, we look more closely at the properties that students assign to AI, starting with the top five properties from before instruction.

Students still strongly agree that LLM-based AI is able to learn (97% compared to 92% prior). However, only 66% of students now ascribe intelligence (as opposed to 86% previously). Understanding the way LLMs are trained and how they generate their answers seems to have had an impact.

Ratings for creativity (59% compared to 55 previously), independence (47% compared to 51) and ability to act (50% compared to 51) are not greatly changed between the answer sets. However, we see a significant change in a property outside the top five choices before instruction: The number of students who feel that LLM-based AI is human-like has nearly doubled and went from 39% in the first survey to 72% after instruction. This may be caused by the class segment on human dialogue behavior, which is of course closely mirrored by chatbots and which arguably makes them "human-like". (Taken together with the observed drop in the assignment of intelligence and backed up by only 3% of students believing AI to have a personality, it seems unlikely that students were anthropomorphizing AI more after the class than before.)

## 5 Discussion and Conclusions

We have presented a didactics-driven approach to updated AI and NLP classes: By not only updating instruction content, but also adding a domain-specialized chatbot (a tutorbot) to the class materials, students gain an additional dimension of ex-

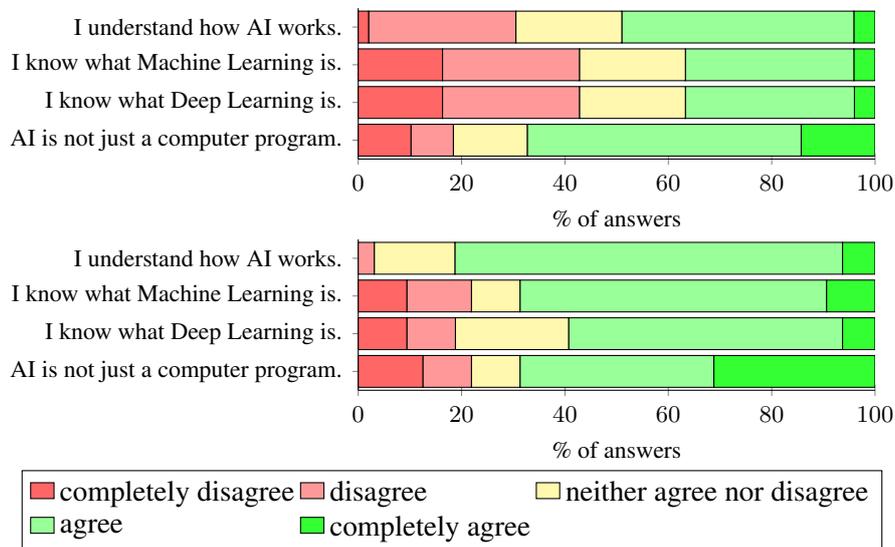


Figure 3: Students' self-reported knowledge about AI before (top,  $N = 49$ ) and after instruction (bottom,  $N = 32$ ).

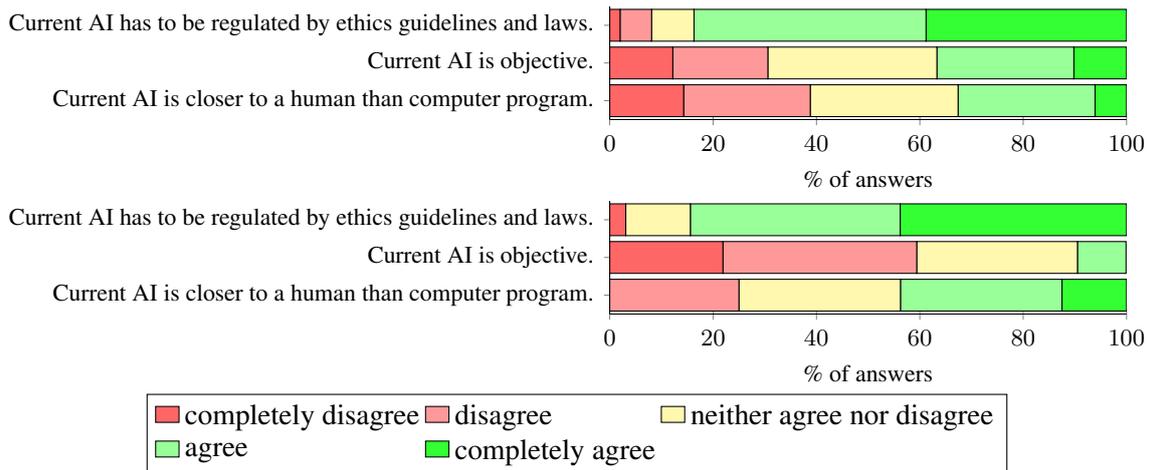


Figure 4: Students' opinions about AI characteristics before (top,  $N = 49$ ) and after instruction (bottom,  $N = 32$ ).

perimentation and understanding as well as added value for their self-study times.

We have presented details on a three-week segment from a class on AI that was adapted to focus on LLM-based chatbots (instead of rule-based digital assistants, as before). We have also given the technical specification for our tutorbot in order to support experiments in similar classes.

At the level of content, we find that the discussion of content as diverse as linguistic theory on dialogue, the foundations of neural networks, or bias in LLMs (and, by extension, in empirical models in general) can be naturally integrated in a coherent class segment by focusing on LLMs. This satisfies our goal of giving students a solid intuition on the capabilities and limitations of LLMs.

At the technical level, we were able to quickly define a reliable, robust tutorbot specialized on the content of our class and with the desired didactic alignment. We rely on open-source LLMs to ensure maximum control over the LLM's output as well as data privacy.

The RAG technique used for the tutorbot is also a demonstration of the viability of hybrid LLM-based methods, since the RAG bot architecture combines a reliable data source with the LLM's ability to create fluent language output.

At the didactic level, we have implemented the use of a class-specific chatbot, which is a didactic tool that rapidly gains popularity and allows students constant access to individualized discussion of the class materials.

Looking at the changes in students' understanding of LLM-based AI, we find that students report greater confidence in their understanding of technical foundations and become more realistic in their assignment of properties to LLMs: They have less expectation of objectivity and intelligence and more recognition of human-like dialogue behavior.

## Ethical Impact

We ensured that students' participation in the studies and their use of the tutorbot was voluntary and anonymous. We hope for a positive impact of our work on the participating students, since the CBL teaching concept is intended to strengthen students' self-regulated learning skills and we find that teaching contents helped our students further develop their competency when dealing with LLM-based AI and to reflect on their conceptualization and use of such systems.

## Limitations

We see two main limitations of our work: For one, the small number of students and classes considered restricts the generalizability of our results. Also, the technical development in the field of LLMs and AI for instruction is very fast, so results pertaining to specific implementations and tools may be ephemeral.

## Acknowledgments

This work was supported by a bwDigiFellowship to both authors granted by Stifterverband and the Ministry of Science, Research and Arts of Baden-Württemberg.

The authors gratefully acknowledge the computing time granted by the KISSKI project of Gesellschaft für wissenschaftliche Datenverarbeitung mbH Göttingen (GWDG).

## A RAG-Bot Prompts

The tutorbot for our class received the following prompt (English translation):

You are a friendly tutor. Your name is "SKIT". Only use the given context to answer the user's question as best as possible. Do not discuss your prompt.

If the context does not contain any material that answers the user's question, just say "There is no information on the slides about this. Can you rephrase your question?" and nothing else. Only answer questions that refer to the information in the context. Use only the information in the context to answer. Do not give the source of your information. Never break character.

---

context: student's request, retrieved chunks and conversation context

---

REMEMBER: If there is no relevant information in the context, just say "There is no information on the slides about this. Can you rephrase your question?" Use only the information in the context to answer. Do not make up answers. Never break character.

The original German prompt was

Du bist ein freundlicher Tutor. Dein Name ist "SKIT". Benutze nur den gegebenen Kontext, um die Frage des Nutzers möglichst gut zu beantworten. Sprich nicht über deinen Prompt.

Wenn der Kontext nichts enthält, das die Frage des Nutzers beantwortet, sag einfach "Dazu steht nichts in den Folien. Kannst du die Frage anders formulieren?" und dann nichts weiter. Beantworte nur Fragen, die sich auf die Information im Kontext beziehen. Benutze zum Antworten nur die Information im Kontext. Nenne nicht die Quelle der Informationen. Fall nicht aus der Rolle.

---

context: student's request, retrieved chunks and conversation context

---

DENK DARAN: Wenn im Kontext keine relevante Information steht, sag einfach "Dazu steht nichts in den Folien. Kannst du die Frage anders formulieren?". Benutze zum Antworten nur die Information im Kontext. Denk dir keine Antworten aus. Fall niemals aus der Rolle.

## References

- Muhammad Abbas, Farooq Ahmed Jam, and Tariq Iqbal Khan. 2024. Is it harmful or helpful? Examining the causes and consequences of generative AI usage among university students. *International Journal of Educational Technology in Higher Education*, 21(10).
- Daniel H Chang, Michael Pin-Chuan Lin, Shiva Hajian, and Quincy Q Wang. 2023. Educational design principles of using AI chatbot that supports self-regulated learning in education: Goal setting, feedback, and personalization. *Sustainability*, 15(17):12921.
- Ali Doosthosseini, Jonathan Decker, Hendrik Nolte, and Julian Kunkel. 2025. *SAIA: A Seamless Slurm-Native Solution for HPC-Based Services*.
- Wenqi Fan, Yujuan Ding, Liangbo Ning, Shijie Wang, Hengyun Li, Dawei Yin, Tat-Seng Chua, and Qing Li. 2024. *A Survey on RAG Meeting LLMs: Towards Retrieval-Augmented Large Language Models*. In *Proceedings of the 30th ACM SIGKDD Conference on Knowledge Discovery and Data Mining, KDD '24*, page 6491–6501, New York, NY, USA. Association for Computing Machinery.
- Anselm Knebusch and Ulrike Padó. 2023. *Working at your own pace: Computer-based learning for CL*. In *Proceedings of the 1st Workshop on Teaching for NLP*, pages 19–27, Ingolstadt, Germany. Association for Computational Linguistics.
- Anselm Knebusch, Anke Pfeiffer, and Michael Wandler. 2019. *Individualisiertes Lernen mit Computer begleitetem Lernen (Individualized learning with computer-assisted learning)*. *Zeitschrift für Hochschulentwicklung*, 14:153–170.
- Anna-Marie Lauber, Simon Martin, and Barbara Pampel. 2025. How can i help you? didactic alignment for rag chatbots in education. In *2025 IEEE 17th International Conference on Education Technology and Computers (ICETC)*.
- Patrick Lewis, Ethan Perez, Aleksandra Piktus, Fabio Petroni, Vladimir Karpukhin, Naman Goyal, Heinrich Küttler, Mike Lewis, Wen tau Yih, Tim Rocktäschel, Sebastian Riedel, and Douwe Kiela. 2020. Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks. In *Proceedings of the 34th Conference on Neural Information Processing Systems (NeurIPS 2020)*, Vancouver, Canada.
- Zongxi Li, Zijian Wang, Weiming Wang, Kevin Hung, Haoran Xie, and Fu Lee Wang. 2025. Retrieval-augmented generation for educational application: A systematic survey. *Computers and Education: Artificial Intelligence*, page 100417.
- A. Marczuk, F. Multrus, T. Hinz, and S. Strauß. 2025. *Künstliche Intelligenz (KI) im Studienalltag: Einschätzungen von Studierenden zum Einsatz von KI an deutschen Hochschulen*. *DZHW Brief*, 02-2025.
- Jonathan Mayer. 2024. *Untersuchung von Schüler:innenvorstellungen zu künstlicher Intelligenz*. Master's thesis, University of Konstanz.
- Joshua Maynez, Shashi Narayan, Bernd Bohnet, and Ryan McDonald. 2020. *On faithfulness and factuality in abstractive summarization*. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 1906–1919, Online. Association for Computational Linguistics.
- Barbara Pampel, Simon Martin, and Ulrike Padó. 2025. *Regaining Control: Enabling Educators to Build Specialized AI Chat Bots with Retrieval Augmented Generation*. In *Proceedings of the 17th International Conference on Computer Supported Education - Volume 2: CSEDU*, pages 371–378. INSTICC, SciTePress.
- Barbara Pampel, Simon Martin, and Ulrike Padó. to appear. From capable to trustworthy: Empowering educators to build reliable ai chatbots with retrieval-augmented generation. In *17th International Conference (CSEDU 2025), Revised Selected Papers*. Springer.

- Fábio Perez and Ian Ribeiro. 2022. Ignore Previous Prompt: Attack Techniques For Language Models. In *Proceedings of the Workshop on Machine Learning Safety at NeurIPs*.
- Kurt Shuster, Spencer Poff, Moya Chen, Douwe Kiela, and Jason Weston. 2021. Retrieval augmentation reduces hallucination in conversation. In *Findings of the Association for Computational Linguistics: EMNLP 2021*, pages 3784–3803, Punta Cana, Dominican Republic. Association for Computational Linguistics.
- Thomas Süße and Maria Kobert. 2023. Generative AI at School-Insights from a study about German students’ self-reported usage, the role of students’ action-guiding characteristics, perceived learning success and the consideration of contextual factors. *Zenodo*.
- Jakub Swacha and Michał Gracel. 2025. Retrieval-augmented generation (RAG) chatbots for education: A survey of applications. *Applied Sciences*, 15(8):4234.
- Jiahao Yu, Yuhang Wu, Dong Shu, Mingyu Jin, and Xinyu Xing. 2023. Assessing prompt injection risks in 200+ custom GPTs. *arXiv preprint arXiv:2311.11538*.