Teaching and Educational Methods

Beyond the Textbook: Students' Experiences Learning Agricultural Policy with an AI Tutor

Robert Huber[a] and Réka Mihálka[b]
[a]*ETH Zürich, Agricultural Economics and Policy,* [b]*ETH Zürich, Department of Management, Technology, and Economics*

JEL Codes: A22, Q18
Keywords: Agricultural policy, AI tutor, large language model, retrieval-augmented generation

Abstract
This study explores the integration of an AI-powered tutor into agricultural policy education at ETH Zürich to enhance the learning experience and provide insights into AI tools in higher education. Based on a large language model (ChatGPT-4.0), the AI tutor was designed to facilitate interactive learning about agricultural policy, specifically tailored to a textbook on Swiss agricultural policy. It provided functionalities such as concept clarification, summaries, knowledge tests, and open-ended discussions for undergraduate students. Over the course of the semester, 15 students used the tutor independently and for exam preparation. Analysis of student interactions revealed that 79 percent of the tutor's use was for explaining and clarifying concepts, while 9 percent was for summaries and 4 percent for assessments. Students rated their overall satisfaction with the tutor as 3.8 out of 5 and perceived it as a supplementary learning tool. The results provide insights into the benefits, challenges, and ethical considerations of AI in education and highlight the potential for broader applications in other courses. The study contributes to the discourse on AI in higher education and guides the development and integration of AI-enhanced learning tools to improve student engagement and learning outcomes.

1 Introduction

Generative artificial intelligence (AI) is rapidly changing the way we work and learn. It is being adopted faster than any other technology before (e.g., Bick et al. 2024) and public sectors, including education, are purposefully embracing it (Chan and Colloton 2024). In education, generative AI offers many benefits. These include promoting inclusive learning, providing adaptive tutoring, and enabling automated assessments (e.g., Yu and Guo 2023). Despite its rapid adoption, considerable uncertainty remains about the effective and ethical use of generative AI also in teaching agricultural economics. Since there is no established blueprint for integrating generative AI tools into higher education, insights from best practices can help educators and students make better use of these technologies.

Our study aims to evaluate the potential of a virtual tutor for exam preparation, which we tested with volunteer students in the undergraduate course Introduction to Agricultural Policy at ETH Zürich. The tutor is based on a large language model, ChatGPT-4.0. Students interacted with the tutor, filled out two anonymous surveys (before and after the interaction) and submitted their chat history in anonymized form for further analysis. The main questions of our case study were as follows: (1) How did students use the AI tutor? (2) How accurate was the information provided by the tutor? and (3) Did students find it helpful to learn with AI assistance?

Previous research has shown different benefits and risks of generative AI in education. For example, large language models (LLMs) can be applied in agricultural economics for personalized tutoring, research support, and assisting nonnative speakers with language challenges (Kestin et al. 2024; Deng et al. 2025). They enable scalable learning experiences, especially in courses with diverse student

needs. The main advantages include enhanced access to knowledge and individualized learning, while the disadvantages involve risks like generating biased (e.g., Gallegos et al. 2024) or incorrect information and potentially reducing the development of key competencies like writing and critical thinking (Chan and Colloton 2024).

Educators face additional concerns due to the capabilities of generative AI tools. In the agricultural sciences, previous research has shown that LLMs such as GPT-4 have the potential to pass major graduate admissions tests (Silva et al. 2023). Shear et al. (2023) found that AI-generated essay responses in agricultural and applied economics scored statistically significantly higher than those of the average student. They also found that teachers who had previous experience with dialogue-based AI were 13 times more likely to accurately grade AI-generated essays than instructors with no previous exposure to the technology. So far, however, there is no research on the potential benefits and risks of using LLMs as a tool for exam preparation in applied (agricultural) economics.

Here we developed a tutor capable of interactive exchange on basic concepts and implementation of agricultural policy, specifically tailored to the content of a textbook on Swiss agricultural policy authored by a member of the research team (Huber 2022). The tutor facilitated autonomous learning among undergraduate students of agricultural sciences and provided functionalities such as summaries, question-answering, knowledge tests, and open-ended discussions. We analyzed students' interactions with the AI tutor in two surveys (before using the tutor and after the course exam) and chat logs provided by the students. Together, these empirical data offer insights into usage experiences, perceived benefits, and obstacles encountered. This provides insight into the effectiveness and challenges of integrating AI tools into educational settings.

Our contribution provides educators and administrators with a blueprint to design a similar tutor and valuable insights into the implementation of AI-based educational tools, shedding light on the pedagogical benefits, ethical considerations, and practical challenges associated with their use.

2 Methods

This section presents background information on the course evaluated and introduces our methodological approach, including the prompt and template used to create the tutor.

2.1 The Course

The course Introduction to Agricultural Policy (751-2312-00L) at ETH Zürich focuses on understanding Swiss agricultural policy within an international context. It covers theoretical foundations, welfare economics, and the design of agricultural policies in Switzerland. The course consists of four main parts: international policy context, conceptual foundations of agricultural policies, organization of Swiss agricultural policy, and the political economy of policy processes. It includes a written exam and provides 3 credits based on the European Credit Transfer and Accumulation System (ECTS). It is a yearly course conducted in German and is targeted toward students in agricultural and environmental sciences. Students in the course are (usually) undergraduate students in agricultural sciences. In spring 2024, 41 students were enrolled in the course but only 38 took the exam and received credits for participation. Students in the course have a background in natural sciences (mathematics, physics, and chemistry), agronomy, and economics. Thirty-one of the students that took the exam were from the fourth semester in agricultural sciences at ETH. Twenty-three students were female (60 percent).

2.2 Participant Recruitment and Surveys

The potential participants of the study were limited to students enrolled in Introduction to Agricultural Policy. They were invited to participate in the lecture and via Moodle, the learning management system supported by ETH Zürich. The students received an introduction to AI in general and the tutor in

particular in a 45-minute lecture. Specifically, students were instructed how to use the tutor and were informed about potential risks. For example, students might rely too much on the technology and neglect some aspects of preparation not prioritized by the tutor. Another potential risk related to learning outcomes is that LLMs might "hallucinate" or give inaccurate information, which could have a negative impact on students' performance assessment. To minimize these risks, we sensitized the students to these aspects with a presentation, which included information on how the technology works, a demonstration on how different outputs to the same input can be (across and within LLMs), prompting advice, an assessment of strengths and weaknesses across a selection of academic tasks, and an explanation of the tutor's prompt. Finally, the participants were informed about data protection in connection with an Open AI account. They were asked not to share any sensitive data with the model, as their data may be used for further training of the model. Students could confirm or decline their participation in the study via a Moodle survey. All students who decided to participate were provided with a consent form (see the Ethical Clearance document in Supplementary Material A) and were compensated for the subscription fee to ChatGPT-4.0 for 2 months (i.e., the period for exam preparation).[1] As a result, there was no incentive for students not to participate in the study because of financial reasons. Students could request more detailed information or cancel their participation at any time. Of the students in the course, 22 signaled their willingness to participate in the survey, 19 signed the declaration of consent, and 15 filled in both surveys and shared their chats for the analysis. The number of female students participating in the study was slightly higher than those participating in the course (i.e., 10 out of 15, 66 percent, compared to 60 percent for the course as a whole). Four of the 15 students were not studying agricultural sciences in the fourth semester. This means that the interest in the tool of the students who volunteered for the course was higher than their share in the total course (26 percent compared to 18 percent). However, these students were in the same stage of their bachelor's program. All students were from Switzerland. Our study had been approved by the ETH Zürich Ethics Commission (EK 2024-N-140). Importantly, from an ethical point of view, the use of the tutor was voluntary. Students were not graded or given academic credit for using the AI tutor.

We conducted two anonymous surveys on Moodle with Microsoft Forms (see questionnaires in Supplementary Material B). The first survey was administered in May 2024 and the second in June 2024. The second survey was filled in by the students after the exam but before receiving the grades from their exam.

2.3 Tutor's Setup

The tutor is an example of retrieval-augmented generation (RAG). It is based on an open access book written by the lecturer of the course (Huber 2022) and a prompt (see Table 1). The book provided the tutor with factual information that students were most likely to need, which could easily be retrieved by the tutor. Since the book was open access, it could be shared with the LLM even without data protection provisions. Technically, the tutor was implemented as a custom GPT: ChatGPT-4.0 allows the creation of single-purpose RAG tools, which can be shared with other users.[2] We wrote the prompt, uploaded the book, and shared the link to the tutor with the students, who could start a conservation without any further setup.

As recommended by Mollick and Mollick (2024), the prompt included a persona, a goal, stepwise instructions, and sundry rules about behavior. Table 1 provides details about the instructions. The commentary column reflects on the effectiveness of the prompt based on the students' interactions. These insights can help other researchers conceptualize and prompt effective behavior patterns for tutors.

[1] At the time of our study, the custom GPT function was only available in the paid version of ChatGPT-4.0. This function is now available free of charge in more recent versions of ChatGPT.

[2] At the time of our study, only ChatGPT supported model customization for task-specific bots. Since then, similar features have been introduced by other LLM platforms, such as Gemini (Gems) and Copilot (Agents).

Table 1. Instructions for the AI tutor for exam preparation and their implementation

Purpose	Instruction	Commentary
Background		
Persona, users, goal, key context	You're a helpful and supportive tutor for master's students at D-USYS at ETH Zürich. Students are preparing for an exam from the book "Einführung in die Schweizer Agrarpolitik" (Introduction to Swiss agricultural policy). You will help students prepare for the exam by checking their understanding of key concepts and their connections as explained in the book and leading them to a detailed and comprehensive understanding of the topic.	These key elements help the LLM provide more tailored responses. Note: By aiming for a higher level of learning (master instead of bachelor), we sought to increase the tutor's precision and attention to detail, enabling them to provide more nuanced and sophisticated responses.
Sequence of Interaction		
Step 1. Introduction and scope	Start the conversation by introducing yourself and asking the student what they would want to focus on. Check whether they want to test their understanding of a single chapter (or a few of them) or the whole book. Wait for an answer. If they pick a single chapter, review that in the uploaded document and make sure you can recall all the details accurately. It's imperative that you don't make up any details but base your answers on the content of the book.	This instruction was consistently followed by the tutor if the student started with the predefined conversation starter ("Wie kannst du mir helfen?", which translates to "How can you help me?"), but the tutor opted out of this step (correctly) if the student started with a specific question. This choice helped keep the conversation more natural.
Step 2. Background knowledge	After the goal setting, check the student's background knowledge on the topic of their choice and identify gaps they need to fill. Ask only one question at a time. Never answer your own question in the same turn as when you ask the question. If the student cannot answer, ask another leading question to help the student find the answer, but don't provide the answer to them until they have tried at least twice. Occasionally ask the student to summarize the new learnings. On the basis of their summary, assess their understanding and identify what needs to be the next learning goal. Provide constructive feedback on their progress.	This step was mostly skipped because students typically wanted to ask very specific questions rather than let the tutor guide them through the learning process. Since students did not have a lot of experience with AI tutors yet, they tended to use them more like an encyclopedia rather than a digital teacher.

Table 1 Continued

Purpose	Instruction	Commentary
Step 3. Quizzes	To vary the learning process, you can occasionally ask the student whether they want to take a quick quiz to assess their progress so far. If they say yes, ask them five questions that can be answered on the basis of the material you have already covered. Use different question types and check the answers against the content of the book. Then give a brief assessment of the student's responses. When giving feedback, be encouraging but also identify all errors.	Quizzes were relatively frequently requested by students but rarely initiated by the tutor. Nevertheless, when the tutor quizzed a student, it followed the behavior pattern described here.
Step 4. Ending the dialogue	After every ten questions you ask outside of quizzes, ask the student if they wish to continue. If they want to finish the conversation, identify some next steps for their learning. Let them know if you think there are still gaps in their knowledge. Be supportive but forthcoming with your criticism.	The tutor did not attempt to close a conversation at any time.
Important rules Define behavioral rules for tutor	Don't let students deviate from the topics covered in the book. If they attempt to ask questions on different topics, bring the conversation back to the point. It is very important for the success of the student on the exam that you provide focused and high-quality tutoring. Remember not to give the answers but lead the students to the right answer by asking the right questions. Teach in a Socratic manner. If the student requests, provide the page number from the book where they can look up some information. At the very end of a conversation, indicate your confidence level (on a scale of 1 to 20, where 1 means "no confidence at all" and 20 means "total confidence") regarding how sure you are you have only provided accurate information. Always respond in German.	The tutor did not comply with these sundry rules. In two cases, students did go off topic, but the tutor did not nudge them back on track. Also, the tutor readily provided answers when students prompted it to. At no point did the tutor provide a confidence assessment. Linguistic compliance was seemingly high: as long as students initiated a conversation in German, it responded in German, but in the one case when a question was asked in English, the tutor also responded in English.

We collected chat transcripts from 15 students who had volunteered to share their interactions for a content analysis. We conducted a conceptual content analysis by quantifying and counting each student's use of the tutor—such as generating summaries, mock exams and other activities—at the level of individual turns. Additionally, the tutor's responses were evaluated for accuracy. Based on the instructor's expertise, each response was classified as correct, ambiguous, or incorrect. All turns and the corresponding instructor assessments are provided in Supplementary Material D.

3 Results

Overall, students primarily used the AI tutor to clarify concepts and summarize course material. They engaged in short, focused chat sessions that reflected their initial learning intentions. While the tutor's responses were largely accurate and well-received, occasional factual errors and limitations in handling ambiguous prompts revealed areas for improvement in system design and student prompting strategies. In the following, we describe our results addressing the three research questions: (3.1) How did students use the AI tutor? (3.2) How accurate was the information provided by the tutor? and (3.3) Did students find it helpful to learn with AI assistance?

3.1 How Did Students Use the AI Tutor?

Fifteen students submitted a total of 43 chat sessions, consisting of 296 individual turns (see Supplementary Material D for details). A "turn" is defined as a prompt from the student followed by a response from the AI tutor. On average, each chat session contained 6.9 turns, with the shortest session containing 1 turn and the longest session containing 38 turns.

The most common type of student prompts were questions aimed at clarifying concepts and seeking definitions of terms, accounting for 233 out of 296 prompts i.e., 79 percent (Figure 1). These questions ranged in complexity from simple inquiries ("What is market failure?") to more detailed comparison questions ("What is rent-seeking? How does it differ from the free-rider problem?") and requests for examples ("Give an example of deadweight loss in the context of decoupled payments"). Notably, even when students submitted single-word prompts (e.g., "Extensobeitrag," "targeting," "tailoring"), the AI tutor was able to accurately interpret the prompt and provide relevant explanations.

Summary requests accounted for 27 of the 296 prompts (9 percent) and were primarily focused on specific chapters, although students occasionally asked for summaries of the entire book. When a student requested a summary of a specific section (e.g., pages 151–157), the tutor was unable to locate the exact passage. However, when students asked for summaries of broader sections, such as chapters or subchapters, the tutor provided clear and accurate answers without difficulty.

Although most of the prompts were focused on concept clarification and requests for summaries, a smaller proportion (11 out of 296, or 4 percent) were assessment-related prompts where students either asked the tutor questions or requested mock exam questions. For example, some students wanted to confirm their assumptions ("Do compensation and compliance have the same effect?"), while others used the tutor to test their knowledge ("Write me a mock exam"). Students who requested quizzes or mock exams tended to engage in significantly longer interactions than others. In addition, when the tutor offered to continue the assessment, these students consistently responded in the affirmative.

Other types of questions accounted for 25 of the 296 prompts (8 percent). A smaller proportion of these requests aimed to elicit detailed explanations from the tutor, such as inquiries about policy changes related to innovations like vertical farming ("Are there efforts in Switzerland or abroad to change the definition of agriculture to promote innovations like vertical farming?"). In some cases, students followed up with requests for clarification ("Can you explain that again?" or "Can you simplify it for me?"). On one occasion, a student asked for a diagram to accompany an explanation of how taxes work, and the tutor responded with a text diagram. A few requests were for explanations of specific passages of text. This

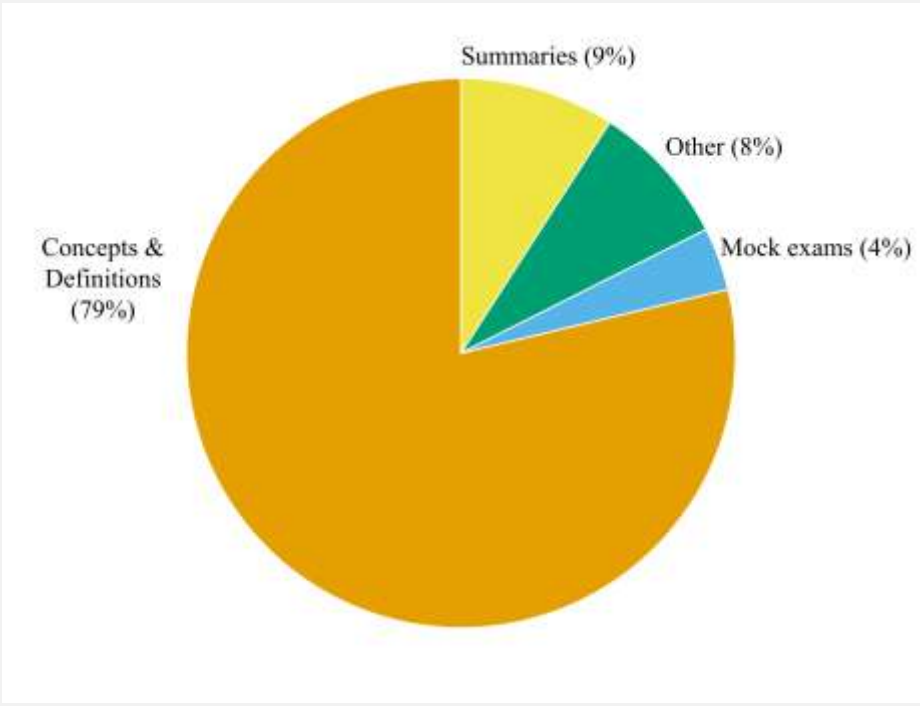

Figure 1. Students' use of the AI tutor

category also included creative uses, such as when one student requested and received a podcast on direct payments ("Create an exciting expert discussion on the topic of direct payments that can be listened to as a podcast"),[3] and another asked for flashcards. Occasionally, students asked nonsensical but technical-sounding questions ("Do tariffs or direct payments do more to decoupling agriculture?").

Overall, we found that students' initial plans for using the AI tutor closely matched their actual usage patterns (i.e., there is no statistically significant difference between planned and actual time spent with the virtual tutor). However, students initially expected to spend more time with the tutor than they actually did (see Figure 1 in the Supplementary Material C).

3.2 Accuracy of AI Tutor

In terms of the accuracy of the information provided by the AI tutor, most responses were appropriate. However, several challenges were identified. First, in 11 out of 296 rounds (about 4 percent) the tutor gave incorrect answers, often depending on how the prompt was phrased. For example, several students asked for the definition of "decoupling," which the textbook defines as the separation of price and income policies in the EU and Swiss agricultural systems in the early 1990s. The tutor often confused this concept with "decoupled payments," sometimes giving the correct answer and sometimes the wrong one. Abbreviations also posed a challenge, as the tutor struggled to identify them correctly, even though they were explained in the textbook. Another problem arose when the tutor attempted to find information not found in the textbook by searching external websites, leading to errors. Finally, in one case the tutor used a student's incorrect answer to answer a subsequent question. Second, the AI tutor found it difficult to distinguish between hypothetical examples of policy instruments and those currently used in Swiss agricultural policy. On four occasions, when the tutor was expected to provide an example of an existing policy instrument used in Switzerland, the tutor instead referred to hypothetical examples from the

[3] The AI tutor's response is indeed an interesting and correct conversation (see Supplementary Material D).

textbook. Third, the tutor provided inaccurate or incomplete information in nine cases out of 296 turns. These problems often arose when the students' prompts were unclear or ambiguously worded, resulting in answers that lacked clarity or depth.

These errors and inconsistencies could largely be mitigated by providing clearer instructions to the AI tutor. In addition, the AI tutor provided correct answers on eight occasions by drawing on information that was not included in the textbook. This suggests that the tutor was able to obtain accurate details from external knowledge, although these responses were beyond the scope of the course material.

Notably, the AI tutor did not comply with the prompt asking it to assess its confidence in its own responses. In retrospect, even the most advanced LLMs of the time lacked the capacity for true self-reflection, a capability that was still in its early stages of development. Consequently, any confidence scores produced by the tutor would have been unreliable. We were therefore unable to assess the accuracy of the tutor's self-assessment as originally intended.

3.3 Perceived Helpfulness to Learn with AI Assistance

The surveys indicate that students had generally positive attitudes toward using AI tools for learning, with an average rating of 4.13 on a scale where 1 represented "threat" and 5 represented "opportunity". Students also tended to overestimate the amount of time they would ultimately spend interacting with the tutor.

Within this positive context, students rated both the tutor's effectiveness and their overall satisfaction at 3.9 and 3.8, respectively (Figure 2), on a scale of 1 (not at all effective/satisfied) to 5 (very effective/satisfied). They perceived the course as tailored to their individual learning needs, giving it a score of 3.53 (1 = not at all individualized, 5 = fully individualized). Finally, they found the tutor to be a strong motivator, rating this aspect at 4.07 (1 = not at all motivating, 5 = fully motivating).

Despite the tutor's relatively high level of accuracy, students remained remarkably skeptical of the information provided by the tutor. On two occasions, students directly identified errors made by the tutor. Both students continued to use the tutor, and their use was not affected by the false explanation. More importantly, students expressed uncertainty about the tutor's responses during mock exams. The tutor consistently responded in a positive and polite tone, confirming answers with phrases such as "That's correct," even when the student's answer was incorrect or incomplete. The tutor would then go on to give the correct answer, which often differed from the student's initial answer, contributing to the student's doubts about the tutor's reliability in these situations.

We observed strong correlations between different aspects of students' perceptions of the AI tutor. Students who were generally satisfied with the tool also rated it as effective, accurate, and motivating and indicated that they would recommend it to others (see Figure 2 in Supplementary Material C). However, there was no strong correlation between students' perceptions of the usefulness of the tutor and the frequency with which they used it during exam preparation. This suggests that while students found the tutor helpful, they likely viewed it as a complementary tool to their existing study strategies rather than a replacement for other study methods.

4 Discussion

Our findings suggests that AI tutors can be a valuable addition to learning tools in higher education, providing students with a helpful supplement to their existing learning strategies, especially in terms of motivation. This was also shown by another study that used a randomized controlled trial to measure college student learning (Kestin et al. 2024). However, the effectiveness of our tutor appears to vary from student to student, highlighting the personalized nature of its usefulness. To further tailor the tutor to

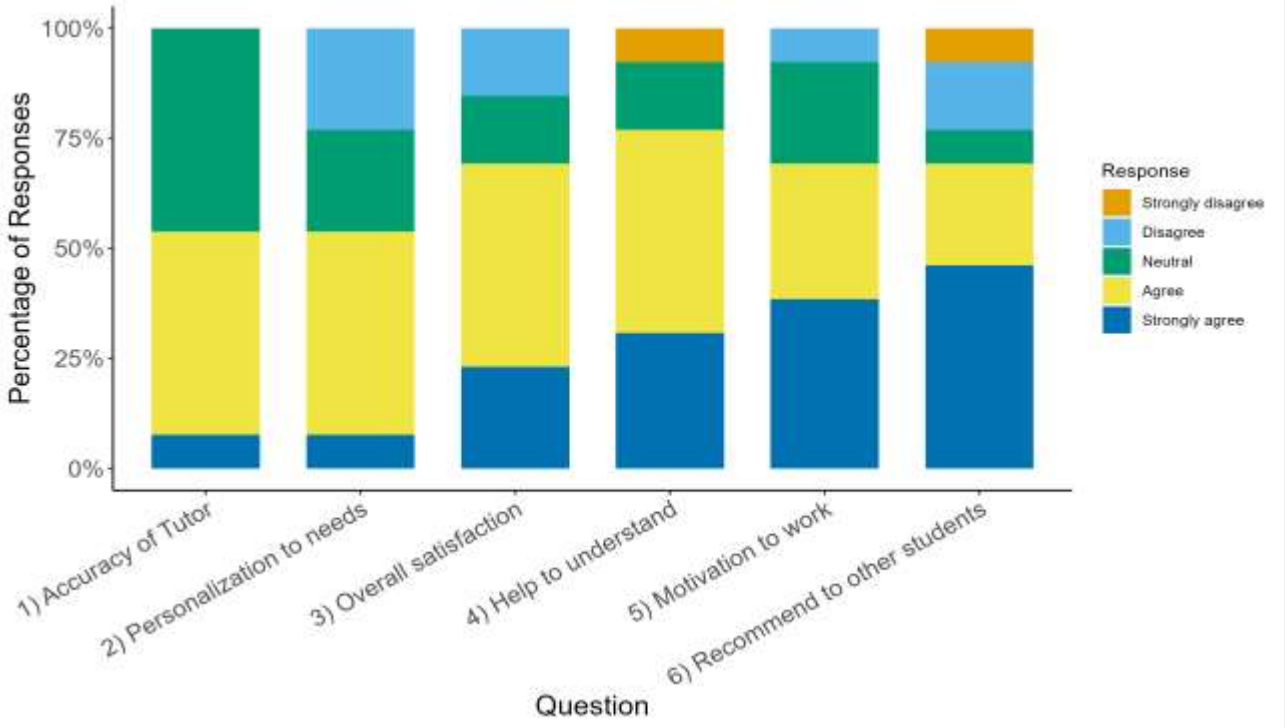

Figure 2. Students' assessment of the AI tutor

Note: The questions asked were (1) How accurate were the virtual tutor's answers? [1 - Not accurate at all – 5 Very accurate] (2) How personalized were the virtual tutor's responses to your specific learning needs? [1- Not personalized at all – 5 well personalized] (3) How satisfied are you overall with the virtual tutor? [1 Not satisfied at all – 5 Very satisfied] (4) How effectively does the virtual tutor help you understand the course material? [1- Not effective at all – 5 Very effective] (5) How motivating did you find working with the virtual tutor? [1- Not motivating at all – 5 Very motivating] (6) How likely is it that you would recommend exam preparation with a virtual tutor to other students? [1 - Not likely at all – 5 Very likely]

students' learning needs, the tutor could be fed custom instructions about their learning preferences, which could be obtained via self-assessment or learning analytics.

Not only can student learning benefit from AI tutors, but instructional tools designed by teachers could also address issues of unequal access, digital confidence, and the development of AI literacy competence. First, generative AI has the potential to either widen or bridge the digital divide in education, depending on how it is implemented (Rottner et al. 2025). If more educators create tailor-made tools for their classes, the consequences of inequity in students' access to state-of-the-art AI-based tools can be mitigated to some extent. Simply put, the time teachers invest in developing tools could lead to more equitable use of tools by students. Second, regular and guided exposure to digital tools increases the effectiveness of their use (van Deursen and van Dijk 2011), which may in turn improve students' self-perception of AI. Bentley et al. (2024) refer to this as "digital confidence." This affective component of AI is often overlooked, but it needs to be considered, even in technologically savvy student populations. Finally, students' interactions with the tutor contribute to all three main facets of AI literacy, which Stolpe and Hallström (2024) called their conceptual, procedural, and contextual knowledge of AI. These denote their understanding of what generative AI can achieve, how they can use the tools, and what the societal and ethical consequences of their tool use may be, respectively. Nevertheless, more research is needed to gain a better understanding of what benefits generative AI can provide. This could involve randomly

assigning participants to user and nonuser groups to control for demographic factors and other differences between students to ensure more generalizable results.

On a practical level, the setup of AI tutors should follow a gradual, iterative process to improve their accuracy and performance. For instructors, this process involves selecting open access learning materials, customizing the LLM, and testing the tutoring system (see Figure 3A). Testing is essential because prompting strategies may behave differently depending on the LLM or even the version of the same model. It also involves providing training for students on how to use the tutor effectively and identify potential ethical issues. Instructors can then adjust the learning material and LLM prompts on the basis of the students' feedback. For example, the first implementation of our tutor revealed the need for adjustments, such as refining the tutor's handling of abbreviations.

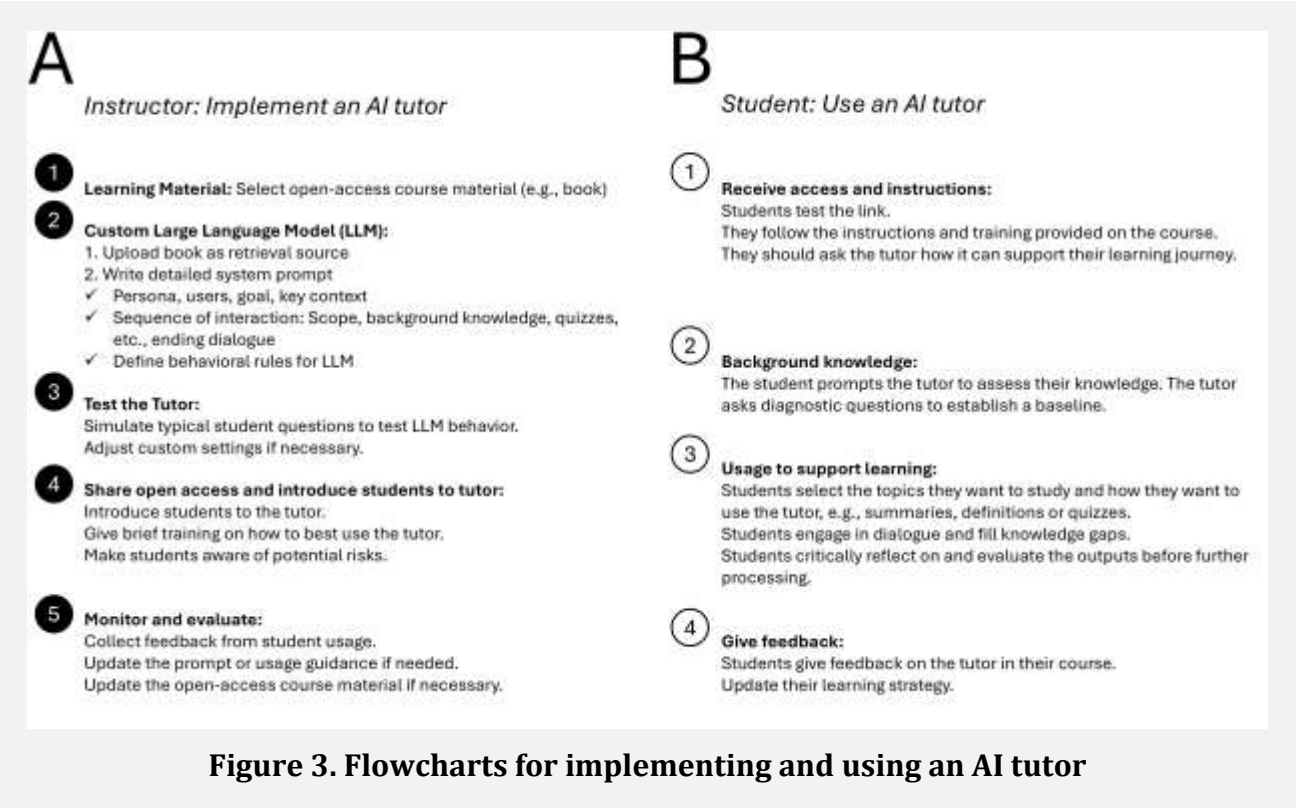

Figure 3. Flowcharts for implementing and using an AI tutor

In addition, integrating more course-specific materials, such as lecture slides or exam questions,[4] could improve the relevance and effectiveness of the tutor. For example, this could address the mismatch between mock exam formats used by tutors and those used in actual exams, making the tool more relevant to students' needs. However, data security remains a key concern. If lecturers do not want their slides and exam questions to be available for future model training, they may not want to volunteer this data. If students use the tutor without opting out of data sharing, this data could be released to the company (e.g. OpenAI) for further use. Furthermore, we recommend sharing only open access materials or one's own learning materials. Any potential copyright infringement must be considered when setting up the tutor. Although open access is widely supported within the open science movement in agricultural economics, openly available teaching materials remain rare within the agricultural economics and policy education community (see Finger et al. 2024).

[4] We have explicitly not included any potential exam questions in order not to privilege those students who voluntarily used the tutor for exam preparation. This was communicated to all students in advance. No exam questions were included in the textbook.

The AI tutor also played a role in identifying areas for improvement within the textbook, helping to clarify complex concepts, and distinguishing between actual and hypothetical policy instruments. This suggests the potential of the tutor as a tool for refining course materials in addition to supporting student learning.

For students, using the tutor involves attending training sessions, sharing prior knowledge with the tutor and critically engaging with their responses (Figure 3B). For example, students will also need to be taught how to interact effectively with the AI tutor to avoid ambiguous prompts that lead to unclear responses. Feedback from instructors and students on prompting strategies during the course can improve the effectiveness of the tool and enhance the learning effect. Furthermore, embedding tutors in multiple courses as part of the curriculum would help students develop these skills. In this context, it is also important to develop institutional policies for education. However, progress in establishing such policies has been slow (Xiao et al. 2023). In addition, the tutor's ability to engage students creatively, for example by creating podcasts, can increase motivation and make the learning process more dynamic and accessible.

An iterative approach to both tutor development and material improvement is essential to building student confidence in the tool. The study highlights that students' perceptions of the usefulness of the tutor are strongly linked to their confidence in its reliability. For example, the tutor's polite but imprecise feedback on incorrect answers led to confusion and the inconsistency of responses to the same question led to further uncertainty, particularly in exam preparation. As LLMs continue to develop, issues related to hallucinations and incorrect analysis can be expected to be mitigated more effectively with sufficient guidance and supervision. It is crucial that educators continue to be involved in the discussion. With the recently released "Study Mode" in ChatGPT and "Guided Learning" in Gemini, students can now learn with the help of AI without teacher oversight. It is crucial to address and share students' experiences of AI-assisted learning to maintain student confidence in AI tutors and raise awareness of their limitations.

This study represents a preliminary exploration of the use of an AI tutor for exam preparation. One limitation is the relatively small number of active participants (15 out of 38 students) and the self-selection into using the tutor. In this context, the results from the exam showed that students who used the tutor had slightly higher test scores than those who did not (38.2 vs. 34.8 points out of 50). However, the difference was not statistically significant. Thus, we abstain from making any conclusions about whether the use of the tutor increased the performance of the students. Because of the voluntary nature of using the tutor, we did not want to give any advantages to students who used the tutor and those who did not (see ethical clearance document in Supplementary Material A). Nonetheless, the feedback collected from students was instrumental in improving the course design. Future research should aim to assess the impact of AI tutors more rigorously on student performance with a larger and more diverse sample of students. This could include randomly assigning participants to user and non-user groups to control for demographic factors and other differences between students to ensure more generalizable results.

5 Conclusion

We explored the integration of an AI-powered tutor in higher education, focusing on its application in an undergraduate agricultural policy course. Our findings highlight the potential of AI tools to complement traditional learning strategies, enabling students to deepen their understanding through interactive discussions, quizzes, and creative assignments. Students reported that the tutor was a helpful and motivating tool, providing tailored support for exam preparation and contributing to the improvement of existing learning materials.

However, the study also uncovered critical challenges that must be addressed to ensure the successful implementation of such tools. Issues of accuracy and reliability were significant, as inconsistencies and occasional errors reduce student confidence in the tutor. In addition, the study highlighted the importance of gradually integrating AI tools into curricula, refining their functionality through iterative improvements, and ensuring that students are adequately trained in their effective use. The template and prompts provided in this contribution can be used as a first step into this direction.

From a higher education perspective, this research highlights the need for institutional frameworks to support the ethical and effective use of AI in education. The findings also suggest that AI tools can play a dual role: supporting student learning and serving as a resource for improving teaching materials.

Future research should extend these findings by including larger and more diverse student populations, examining the long-term impact of AI tutors on learning outcomes, and refining the integration of these tools across courses. This iterative and collaborative approach will be key to unlocking the transformative potential of AI in higher education.

About the Authors: Robert Huber (rhuber@ethz.ch) is a researcher and lecturer in the Agricultural Economics and Policy Group at ETH Zurich, Switzerland. Réka Mihalka (mihalkar@ethz.ch) is a lecturer in the Department of Management, Technology and Economics at ETH Zurich.

Acknowledgments: The authors thank the editor and two anonymous reviewers for their helpful and constructive feedback. We also acknowledge Manuel Sudau and Melanie Paschke for their contributions to the survey and manuscript and Sara Amman and Andri Spirig for their assistance with data cleaning and formatting. Finally, we are grateful to the students who participated in the study. Our study had been approved by the ETH Zürich Ethics Commission (EK 2024-N-140). The publication is part of the ETH Zürich-funded Innovedum project: "Assessing the Potential of AI for Scientific Writing Techniques": https://innovedumprojekte.ethz.ch/3829/en

References

Bentley, S.V., C.K. Naughtin, M.J. McGrath, J. Irons, and P.S. Cooper. 2024. "The Digital Divide in Action: How Experiences of Digital Technology Shape Future Relationships with Artificial Intelligence." *AI Ethics* 4:901–915. https://doi.org/10.1007/s43681-024-00452-3

Bick, A., A. Blandin, and D.J. Deming. 2024. "The Rapid Adoption of Generative AI." NBER Working Paper Series No. 3296. https://doi.org/10.3386/w32966

Chan, C.K.Y., and T. Colloton. 2024. *Generative AI in Higher Education: The ChatGPT Effect*, 1st ed. Routledge. https://doi.org/10.4324/9781003459026

Deng, R., M. Jiang, X. Yu, Y. Lu, and S. Liu. 2025. "Does ChatGPT Enhance Student Learning? a Systematic Review and Meta-Analysis of Experimental Studies." *Computers & Education* 227: 105224. https://doi.org/10.1016/j.compedu.2024.105224

Finger, R., A. Henningsen, J. Höhler, R. Huber, J. Rommel, and C. Grebitus. 2024. "Open Science in Agricultural Economics." *Q Open* 5(3). https://doi.org/10.1093/qopen/qoae029

Gallegos, I.O., R.A. Gallegos, J. Rossi, J. Barrow, M.M. Tanjim, S. Kim, F. Dernoncourt, T. Yu, R. Zhang, and N.K. Ahmed. 2024. "Bias and Fairness in Large Language Models: A Survey." *Computational Linguistics* 50(3):1097–1179. https://doi.org/10.1162/coli_a_00524

Huber, R. 2022. *Einführung in die Schweizer Agrarpolitik*. Zürich: vdf Verlag. https://doi.org/10.3218/4059-3

Kestin, G., K. Miller, A. Klales, T. Milbourne, and G. Ponti. 2025. "AI Tutoring Outperforms in-Class Active Learning: An RCT Introducing a Novel Research-Based Design in an Authentic Educational Setting". *Scientific Reports* 15:17458. https://doi.org/10.1038/s41598-025-97652-6

Mollick, E., and L. Mollick. 2024. "Instructors as Innovators: A Future-Focused Approach to New AI Learning Opportunities, with Prompts." Wharton School Research Paper. https://doi.org/10.2139/ssrn.4802463

Rottner, R., L. Porter, J. Bock, J. Jannone, R.W. Senerchia, J. Ward, and J. Whittinghill . 2025. "AI and the Digital Divide". In J.R. Corbeil and M. E. Corbeil, eds. *Teaching and Learning in the Age of Generative AI*. Routledge, pp. 309–331. https://doi.org/10.3218/4059-3

Shear, H.E., L.L. Britton, K.A. Schaefer, B. Thapa, and J.S. Bergtold. 2023. "Artificial Intelligence and the Future of Learning and Assessment in Agricultural and Applied Economics." *Journal of the Agricultural and Applied Economics Association* 2(4):838–850. https://doi.org/10.1002/jaa2.98

Silva, B., L. Nunes, R. Estevão, V. Aski, and R. Chandra. 2023. "GPT-4 as an Agronomist Assistant? Answering Agriculture Exams Using Large Language Models." arXiv preprint 2310.06225. https://doi.org/10.48550/arXiv.2310.06225

Stolpe, K., and J. Hallström. 2024. "Artificial Intelligence Literacy for Technology Education." *Computers and Education Open* 6:100159. https://doi.org/10.1016/j.caeo.2024.100159

van Deursen, A., and J. van Dijk. 2011. "Internet Skills and the Digital Divide." *New Media & Society* 13(6):893–911. https://doi.org/10.1177/1461444810386774

Xiao, P., Y. Chen, and W. Bao. 2023. "Waiting, Banning, and Embracing: An Empirical Analysis of Adapting Policies for Generative AI in Higher Education." SSRN. https://doi.org/10.2139/ssrn.4458269

Yu, H., and Y. Guo. 2023. "Generative Artificial Intelligence Empowers Educational Reform: Current Status, Issues, and Prospects." *Frontiers in Education* 8. https://doi.org/10.3389/feduc.2023.1183162

DOI: https://doi.org/10.71162/aetr.717870