

# Improving AI assistants embedded in short e-learning courses with limited textual content

Jacek Marciniak, Marek Kubis, Michał Gulczyński  
Adam Szpilkowski, Adam Wieczarek, Marcin Szczepański

Faculty of Mathematics and Computer Science  
Adam Mickiewicz University, Poznań  
ul. Uniwersytetu Poznańskiego 4, 61-614 Poznań, Poland  
{jacekmar,mkubis,marcin.szczepanski}@amu.edu.pl  
{micgull,adaszp,adawiel}@st.amu.edu.pl

## Abstract

This paper presents a strategy for improving AI assistants embedded in short e-learning courses. The proposed method is implemented within a Retrieval-Augmented Generation (RAG) architecture and evaluated using several retrieval variants. The results show that query quality improves when the knowledge base is enriched with definitions of key concepts discussed in the course. Our main contribution is a lightweight enhancement approach that increases response quality without overloading the course with additional instructional content.

## 1 Introduction

AI assistants based on large language models (LLMs) are increasingly used to support learning and access to educational content. Most research in this area assumes access to large-scale textual resources, such as entire textbooks or extensive document collections. To reduce hallucinations and improve grounding, many approaches rely on techniques such as Retrieval-Augmented Generation (RAG; Lewis et al., 2020), where relevant documents are retrieved from a knowledge base and passed to the model at inference time. However, the effectiveness of such methods typically depends on the availability of rich textual input—an assumption that often does not hold in real-world educational contexts.

In practice, modern educational programs often rely on short e-learning modules designed to teach narrowly defined learning objectives within a limited timeframe. These modules—especially in higher education—are intentionally concise to preserve instructional clarity and reduce cognitive load. When AI assistants are embedded in such courses, they are expected to provide accurate, context-aware support without relying on large external corpora or hallucinating irrelevant content.

Despite the growing popularity of LLM-based assistants, there is a lack of research on how to

design such systems when instructional content is minimal. Existing work typically targets high-resource settings, and it remains unclear whether techniques developed for large-scale retrieval transfer effectively to low-resource educational contexts. Moreover, instructors often have limited time and must make strategic decisions about which concepts or materials are worth covering. Expanding materials solely to meet model requirements is pedagogically undesirable.

This paper investigates how to improve the effectiveness of AI assistants embedded in short e-learning courses with limited textual content. Rather than expanding the course, we propose a lightweight enhancement strategy: injecting definitions of key course concepts into the assistant’s knowledge base. We evaluate this approach using a real-world e-learning course on machine learning fundamentals (approx. 30 learning objects) and a benchmark of 94 questions collected from students who completed the course.

Our main contributions are as follows:

1. We identify and address the challenge of building AI assistants for short e-learning courses with limited instructional content.
2. We show that augmenting the knowledge database with definitions of key course concepts improves response quality, even without modifying the course itself.
3. We demonstrate that retrieval method variants have relatively little impact compared to content enrichment, providing a practical and scalable solution for educators with limited time and resources.

## 2 Related work

Recent research on AI-powered educational assistants has largely relied on large-scale datasets. For example, Wang et al. (2024) introduced

Book2Dial, which generates synthetic teacher-student dialogues from 35 textbooks to fine-tune chatbots—though issues like hallucinations and repetitive content remain. Similarly, [Fernandez et al. \(2024\)](#) proposed SyllabusQA, a 5k QA-pair dataset from 63 course syllabi, aimed at handling logistical queries. Despite high similarity scores, factual accuracy remained a challenge.

[Huang et al. \(2025\)](#) presented RAM2C, a RAG-based system generating pedagogically grounded dialogues in liberal arts education. The method depends on rich, curated knowledge bases, which limits applicability to low-resource contexts. [Garcia \(2025\)](#) combined RAG and LLMs to help instructors analyze student reflections and identify course-wide learning challenges through topic modeling. While RAG offered valuable insights, it did not consistently outperform standalone LLMs.

To the best of our knowledge, no prior work has addressed how to design AI assistants for courses with limited textual content, where expanding the material is not feasible due to instructional constraints.

### 3 AI-assisted course

The study was conducted using the e-learning course *Introduction to Machine Learning*, designed to provide foundational knowledge and develop practical skills in constructing and analyzing simple machine learning models. ([Szczepański et al., 2025](#)). The course emphasizes applied learning through examples and hands-on exercises using Google Teachable Machine. Figure 1 shows an excerpt from the course materials.

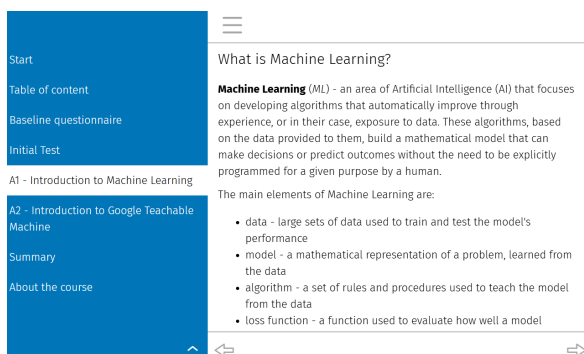


Figure 1: Fragment of the learning materials.

The course is organized into four modules: (1) introduction, (2) data preparation, (3) model training, and (4) evaluation metrics. It is used in AI-related computer science classes as a preparatory resource

to align students' baseline knowledge in machine learning. Estimated completion time is 3–4 hours.

The content is intentionally concise, focusing on the fundamentals of classification. While broader AI and machine learning topics are briefly mentioned, they are not developed in detail. As an introductory resource, it is used by students with varying prior knowledge, leading to diverse questions during the learning process. To support this, an AI assistant was introduced to help in four areas: (1) clarifying course content, (2) deepening understanding of key concepts, (3) addressing related but uncovered topics, and (4) summarizing material.

Expanding the core content was considered pedagogically inappropriate. The course was purposefully limited to foundational topics, with advanced material reserved for later stages in the curriculum. Nonetheless, students may still raise more complex questions. The course is a validated educational resource, positively received by students in earlier editions, and modifying it solely to enhance AI assistant performance was not an option.

Designed for self-paced learning outside of class—where instructor support may be unavailable—the course positions the AI assistant as a key element of the learning experience, offering targeted guidance as students navigate the material independently.

### 4 Method

The AI assistant embedded in the e-learning course follows RAG architecture, which combines the strengths of large language models with the ability to incorporate domain-specific knowledge—in our case, the textual content of the course.

To develop the system, we compared several RAG variants. After a series of preliminary tests, we decided to adopt AdvanceRAG approach with a query routing mechanism as it yielded the best results. This model involves classifying the user query into a predefined type—such as requesting a citation, paraphrase, summary, or elaboration—and dynamically selecting a tailored prompt accordingly. Based on the chosen prompt, the system retrieves relevant data from a knowledge database, which is then passed to the language model along with the prompt to generate the final response. We chose *LLaMA 3.1 8B* as our foundation model due to its balance between output quality and hardware requirements.

To evaluate how different retrieval strategies im-

pact response quality, we tested four configurations of increasing complexity and contextual coverage:

**Baseline** The initial setup used only the course text as the knowledge base, paired with a dense retriever. To improve retrieval precision, the material was preprocessed to remove auxiliary or transitional content (e.g., phrases such as “*Let’s move on to the next section*”) that could degrade semantic relevance. This minimal configuration reflects a real-world scenario where the assistant operates solely on content provided by course authors.

**Reranking** This variant introduced a reranking stage using the ms-marco-MiniLM-L-6-v2 cross encoder to improve semantic relevance.

**Extended** In this configuration we added curated Wikipedia articles containing definitions of key concepts presented in the course to the document set, to increase the breadth of available information.

**Combined** The final configuration employs both reranking and the extension of the document set.

## 5 Experiments

### 5.1 Data

To evaluate the performance of the proposed system, we collected a set of real user questions related to a short e-learning course on the basics of machine learning. The questions were formulated by students who had previously completed the course. To help participants simulate realistic interactions with an AI assistant, they were instructed to first ask their question and then obtain an answer from ChatGPT, followed by an evaluation of whether the response was satisfactory.

A total of 94 questions were collected from 14 students. Each student submitted between 2 and 10 questions, covering all four modules of the course (28 questions from module 1, 19 from module 2, 29 from module 3, and 18 from module 4).

The resulting dataset consists of natural, goal-oriented queries and can be categorized into four main types: (1) clarifying course content (45 questions); (2) deepening understanding of key concepts (33 questions); (3) addressing related but uncovered topics (10 questions); (4) summarizing material (6 questions).

This dataset forms the basis for evaluating retrieval configurations under realistic student-like usage scenarios.

### 5.2 Retrieval evaluation

For the purpose of evaluating retrieval performance we measured Reciprocal Rank@K (**RR@K**), Normalized Discounted Cumulative Gain@K (**nDCG@K**), Average Precision@K (**AP@K**), Recall@K (**R@K**), Precision@K (**P@K**) and determined F1 scores. The results are given in Table 1. The *Baseline* solution while competitive in precision for top-ranked results ( $RR@3 = 0.2579$ ), lacked contextual depth, limiting the assistant’s ability to handle more complex or exploratory queries. In case of *Reranking* model the performance improved for  $K=1$ , however all metrics for  $K=3, 5$  decreased. This suggests that reranking narrows the focus at the cost of contextual diversity—an undesirable trade-off in educational settings, where broader context is often beneficial for comprehension. The *Extended* configuration significantly improved context diversity and precision, especially for  $K=5$ , where precision rose from 0.1167 to 0.1663. However,  $nDCG@5$  declined, likely due to the added noise from general-purpose content. The *Combined* approach yielded the best performance for  $K=1$  but consistently underperformed for higher values of  $K$ , indicating a trade-off between precision at the top and overall contextual coverage. Among all tested configurations, the *Extended* configuration proved most effective. It provided the best balance between precision and recall at  $K=3$  and  $K=5$  (e.g.,  $F1@3 = 0.2306$ ;  $F1@5 = 0.2179$ ), making it well-suited for educational assistants that must deliver context-rich responses aligned with instructional goals.

### 5.3 End-to-end assessment

To measure end-to-end performance of the AI assistant we asked a group of three experts to assess the quality of the responses yielded by the system. Each expert was provided with reference answers, responses predicted by the system and the contextual information retrieved from the knowledge base for the given question. The experts were requested to verify, if the answer returned by the system is adequate given the provided reference answer and the context being retrieved, with three options available *Yes*, *No* and *Don’t know*. Three configurations were evaluated *Baseline*, *Extended* and *Combined* to measure the impact of enhancing the AI assistant with the definitions of key course concepts on re-

Table 1: Retrieval results

Model	K	RR@K	nDCG@K	AP@K	R@K	P@K	F1 Score
Baseline	1	0.1904	0.1904	0.1785	0.1785	0.1904	0.1843
	3	0.2579	0.3104	0.2569	0.4642	0.1626	0.2410
	5	0.2000	0.2878	0.2012	0.5595	0.1166	0.1931
Reranking	1	0.2142	0.2142	0.2023	0.2023	0.2142	0.2082
	3	0.1607	0.1932	0.1488	0.3134	0.1111	0.1641
	5	0.1644	0.2165	0.1481	0.4146	0.0904	0.1485
Extended	1	0.2530	0.2530	0.0957	0.0957	0.2530	0.1390
	3	0.3092	0.2323	0.1478	0.2574	0.2088	0.2306
	5	0.2263	0.2095	0.1307	0.3160	0.1662	0.2179
Combined	1	0.2650	0.2650	0.1128	0.1128	0.2650	0.1583
	3	0.2168	0.1530	0.0913	0.1720	0.1405	0.1547
	5	0.1903	0.1452	0.0788	0.2192	0.1108	0.1473

Table 2: End-to-end expert evaluation

System type	% Yes	% No	% Don't know	Fleiss' $\kappa$
Baseline	47	41	12	0.14
Extended	53	38	9	0.26
Combined	55	31	14	0.24

sponse quality.<sup>1</sup> The evaluation results along with inter-annotator agreement measured with Fleiss'  $\kappa$  are reported in Table 2. The results show that extending the knowledge base with definitions of key concepts lead to the rise of the percentage of adequate answers from 47% for *Baseline* system to 53% in case of *Extended* configuration. The introduction of reranking resulted in additional 2% performance rise as shown by *Combined* configuration, however it has to be noted that it also lead to 5% increase of responses that were considered by the experts to be ambiguous.

#### 5.4 Challenges in human annotation of end-to-end AI assistant responses

The annotation process highlighted how challenging human evaluation can be in the development of AI assistants. System outputs were assessed by annotators (two instructors and one student) familiar with the course content and its pedagogical intent, and the resulting labels reflect the complexity of the task — yet they show that the adopted labeling scheme provided a solid basis for meaningfully capturing the quality of the assistant's responses.

<sup>1</sup>Due to the limited availability of human annotators, the *Reranking* configuration was not included in end-to-end evaluation. Nevertheless, the impact of reranking can be observed by contrasting the *Extended* configuration, which does not employ reranking, with the *Combined* configuration, which incorporates both reranking and the extended document set.

For example, the following response to Question 14 (“*What is the reward and punishment system in machine learning?*”) generated by the *Combined* configuration was labeled *No*: *The answer to the student's question: The reward and punishment system in machine learning is a model that learns by interacting with the environment, through a system of rewards and punishments.* The sentence is incoherent, repetitive, and lacks both clarity and educational value; it relies on a tautological structure, effectively rephrasing the question instead of providing a meaningful or instructive definition.

Another example is the response to Question 38 (“*Does class imbalance matter?*”), generated by the *Combined* configuration, which was labeled *Don't know*: *Yes, class imbalance does matter. In the case of a very rare disease and an extremely uneven class distribution — between sick and healthy individuals — both in the training set and in the test set randomly drawn from the population, there is only one sick person for every 99 healthy ones. In such a scenario, the model ignored the minority class and automatically assigned all examples to the dominant one, achieving a very high accuracy of 99%, yet failing to fulfill its primary purpose — correctly diagnosing the sick individual. Consequently, it was to be expected that the model, having ‘seen’ only a single instance of that digit during training, would be unable to learn to recog-*

nize it, and would instead simply ignore it. While the response is generally correct, the explanation is lengthy and somewhat dense. It relates to course content but does not explicitly reference it or direct the learner to definitions provided in the material. As a result, it may hinder comprehension rather than support it, especially for learners unfamiliar with the concept.

A third case illustrates a different kind of difficulty — when the relevant information is simply missing. In response to Question 65 (“*What is an edge case in machine learning?*”), the *Combined* configuration returned: *I cannot answer this question because the context does not contain any passages that could be quoted as a valid answer.* This response was annotated as *No*, although it is, in principle, correct — the concept of edge cases was not covered in the course content nor included in the extended knowledge base. The annotator’s negative judgment likely resulted from the system’s failure to provide a pedagogically helpful justification, such as: *This topic falls outside the scope of the course and is therefore not addressed here.* This example demonstrates that analyzing student queries alongside system responses can help identify gaps in coverage and motivate iterative updates to the knowledge base.

The full answers returned by the system across all three configurations discussed above are included in Appendix C.

## 6 Discussion

The results of our experiments show that even modest extensions to the assistant’s knowledge base—specifically, the inclusion of carefully selected definitions of course-relevant concepts—can lead to measurable improvements in response quality. However, while the improvements were consistent, they remained moderate in scope. Expert assessments showed only fair agreement (Fleiss’  $\kappa = 0.26$ ), highlighting the inherent challenges of evaluating AI-generated responses in educational contexts, where interpretation often depends on the perceived intent behind a student’s question.

Instructors providing feedback to students must often determine whether a question stems from confusion, a need for clarification, or simple curiosity. The experts participating in our evaluation may have applied similarly critical reflection when judging the assistant’s answers. During annotation, they likely evaluated the responses based on cri-

teria such as factual correctness, relevance to the question, and linguistic clarity, as well as pedagogical usefulness, alignment with course terminology, and the ability to communicate uncertainty when appropriate.

Additionally, some limitations in response quality likely stem from the assistant’s lack of access to richer content. This may have particularly affected questions aimed at deepening understanding (e.g., through examples beyond those given in the course) or exploring topics that, while present in the instructional material, were not discussed in sufficient detail due to being outside the intended scope of instruction. In these cases, although the retrieved context included terms relevant to the student’s question, the absence of detailed explanation or clear definitions reduced the educational usefulness of the assistant’s response. Such cases highlight the need for a more nuanced expansion of the knowledge base, especially when dealing with boundary concepts that are implicitly acknowledged in course materials but not explicitly explained.

## 7 Future work

This study did not examine the impact of enriching the assistant’s context with broader resources, such as domain-specific books or curated examples from outside the course scope. Future work should also explore how different segmentation strategies for content added to the knowledge base influence AI assistant performance. Another important direction for future work is expanding the knowledge base with content addressing topics raised by students that are currently missing from both the course and the extended resources. It is also planned to collect feedback on the usefulness of AI assistants during learning, with particular attention to their perceived limitations.

## 8 Conclusions

Our study shows that AI assistants embedded in short e-learning courses can be improved without expanding the core instructional content. Instead of increasing course length or adding in-line material—which could compromise clarity and coherence—instructors can enhance assistant performance by supplying concise, reference-style content directly to the RAG knowledge base.

## 9 Limitations

The use of a general-purpose dense retriever not tailored to educational content represents a limitation of this study. Future research should investigate task-adapted or hybrid retrieval methods more closely aligned with instructional needs.

The course materials utilized in the experiments cover only one specific STEM subject. To what extent the presented results can be generalized to social sciences and humanities coursework requires further investigation.

## References

- Nigel Fernandez, Alexander Scarlatos, and Andrew Lan. 2024. [SyllabusQA: A course logistics question answering dataset](#). In *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 10344–10369, Bangkok, Thailand. Association for Computational Linguistics.
- Frank Ley Garcia. 2025. [Llm+rag driven topic modeling](#). In *Proceedings of the 56th ACM Technical Symposium on Computer Science Education V. 2, SIGCSETS 2025*, page 1754, New York, NY, USA. Association for Computing Machinery.
- Haoyu Huang, Tong Niu, Rui Yang, and Luping Shi. 2025. [RAM2C: A liberal arts educational chatbot based on retrieval-augmented multi-role multi-expert collaboration](#). In *Proceedings of the 31st International Conference on Computational Linguistics*, pages 448–458, Abu Dhabi, UAE. Association for Computational Linguistics.
- 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 International Conference on Neural Information Processing Systems, NIPS '20*, Red Hook, NY, USA. Curran Associates Inc.
- Marcin Szczepański, Grzegorz Gapiński, and Jacek Marciniak. 2025. [Ai tutor: Adaptive e-learning system using expert fuzzy controllers](#). In *Proceedings of the 17th International Conference on Computer Supported Education - Volume 2: CSEDU*, pages 96–107. INSTICC, SciTePress.
- Junling Wang, Jakub Macina, Nico Daheim, Sankalan Pal Chowdhury, and Mrinmaya Sachan. 2024. [Book2Dial: Generating teacher student interactions from textbooks for cost-effective development of educational chatbots](#). In *Findings of the Association for Computational Linguistics: ACL 2024*, pages 9707–9731, Bangkok, Thailand. Association for Computational Linguistics.

## A Sample Student Questions Used for Evaluation

### 1) Clarifying course content:

- Explain the difference between a training set and a test set in Machine Learning.
- Explain in one sentence what overfitting and underfitting mean in machine learning.

### 2) Deepening understanding of key concepts:

- Why is precision worse than accuracy?
- When is the F-score a good evaluation metric? In what situations should it be used?

### 3) Addressing related but uncovered topics:

- What are large language models?
- Why can't you switch between browser tabs during training in Google Teachable Machine?

### 4) Summarizing material:

- Shorten it by half: ...
- Write it out for me shorter, so that each definition takes up one sentence. ...

## B RAG prompts

### Common system context (used in all prompts):

You are an educational assistant in an e-learning course on deep learning and artificial intelligence. Your task is to answer student questions only when they are relevant to the course topics. You are capable of analyzing text, answering questions, explaining concepts, summarizing, and paraphrasing content.

**Citation prompt:** Answer the student's question using only quotations from the context below.

**Context:** {context}

**Chat history:** {chat\_history}

**Student question:** {question}

Respond by quoting only fragments of text found in the context. Citations should include full sentences or coherent passages. You may use multiple quotes if needed to fully answer the question. Do not add any words, interpretations, or paraphrasing of your own. If the context does not contain suitable quotes, state this clearly.

## C End-to-end expert evaluation

**Summarization prompt:** Answer the student's question by creating a concise but informative summary based solely on the context.

**Context:** {context}

**Chat history:** {chat\_history}

**Student question:** {question}

Craft a summary that synthesizes key information from the context using your own words. The summary should be complete, accurate, and based only on the context. If the context lacks sufficient data, state this clearly.

**Paraphrasing prompt:** Answer the student's question by explaining the concept using different words, based on the context.

**Context:** {context}

**Chat history:** {chat\_history}

**Student question:** {question}

Provide a paraphrased or simplified explanation using only information from the context. Do not quote directly, add new information, or change the original meaning. If the context lacks sufficient information, state this clearly.

**General fallback prompt:** Answer the student's question using only the context and chat history provided.

**Context:** {context}

**Chat history:** {chat\_history}

**Student question:** {question}

Respond clearly and precisely based solely on the available information. Do not use external knowledge. If the context lacks sufficient information, state this clearly.

**Routing prompt (for classification):** You are an intelligent e-learning assistant that analyzes user questions and classifies them into one of four categories: citation, summarization, paraphrasing, or other.

**Input:** {question}

Return a valid JSON object with two keys: "option" and "explanation".

**Example output:**

```
{
  "option": "citation",
  "explanation": "The user is asking for a
precise definition."
}
```

Return only valid JSON, with no extra text.

Table C.1: Example 1 of end-to-end expert evaluation

<b>Question 14:</b> What is the reward and punishment system in machine learning?		
<b>System type</b>	<b>System answer</b>	<b>Evaluation</b>
Baseline	The answer to the student’s question: The reward and punishment system in machine learning is a model that learns by interacting with the environment, through a system of rewards and punishments.	No
Extended	Here’s the answer to the student’s question: Reinforcement learning is a subfield of machine learning focused on how software agents should take actions in an environment to maximize a notion of cumulative reward. Reinforcement learning is a way of machine learning where the agent receives feedback in the form of rewards or punishments. In machine learning, the reward and punishment system is part of the learning process, where the agent is rewarded for actions that lead to desirable outcomes and punished for those that lead to undesirable ones.	Yes
Combined	The answer to the student’s question: The reward and punishment system in machine learning is a model that learns by interacting with the environment, through a system of rewards and punishments.	No



Table C.2: Example 2 of end-to-end expert evaluation

<b>Question 38:</b> Does class imbalance matter?		
<b>System type</b>	<b>System answer</b>	<b>Evaluation</b>
Baseline	Yes, class imbalance does matter. In the case of a very rare disease and an extremely uneven class distribution — sick versus healthy individuals — both in the training set and in the test set randomly sampled from the population, there may be only one sick person for every 99 healthy ones. In such a scenario, the model simply ignored the minority class and automatically assigned all examples to the dominant one, achieving a seemingly high accuracy of 99%, while entirely failing at its main task: correctly diagnosing the sick individual. It was therefore to be expected that a model exposed to only a single instance of that digit during training would be unable to learn to recognize it — and would instead simply ignore it. This example clearly shows how important it is to ensure that all classes in the training set are represented in roughly equal numbers.	Don't know
Extended	Yes, class imbalance does matter. For instance, in the case of a very rare disease and an extremely uneven distribution between the sick and the healthy, the model may ignore the minority class and automatically assign all examples to the dominant one — achieving very high accuracy, yet failing to accomplish its primary objective. In such a case, the model will not be able to correctly diagnose the sick individual.	Yes
Combined	Yes, class imbalance does matter. In the case of a very rare disease and an extremely uneven class distribution — between sick and healthy individuals — both in the training set and in the test set randomly drawn from the population, there is only one sick person for every 99 healthy ones. In such a scenario, the model ignored the minority class and automatically assigned all examples to the dominant one, achieving a very high accuracy of 99%, yet failing to fulfill its primary purpose — correctly diagnosing the sick individual. Consequently, it was to be expected that the model, having "seen" only a single instance of that digit during training, would be unable to learn to recognize it, and would instead simply ignore it.	Don't know

Table C.3: Example 3 of end-to-end expert evaluation

<b>Question 40:</b> What is an edge case in machine learning?		
<b>System type</b>	<b>System answer</b>	<b>Evaluation</b>
Baseline	There is no answer to this question in the given context.	No
Extended	There is no answer to this question in the context.	No
Combined	I cannot answer this question because the context does not contain any passages that could be quoted as a valid answer.	No

Table C.4: Original Polish version of Example 1

<b>Question 14 in Polish: Co to jest system kar i nagród w uczeniu maszynowym?</b>	
<b>System type</b>	<b>System answer in Polish</b>
Baseline	Odpowiedź na pytanie kursanta: System kar i nagród w uczeniu maszynowym to model, który uczy się wchodząc w interakcje ze środowiskiem, poprzez system kar i nagród.
Extended	Oto odpowiedź na pytanie kursanta: Reinforcement learning jest obszarem uczenia maszynowego, który koncentruje się na tym, jak agenty software powinni podejmować działania w środowisku, aby maksymalizować pewną koncepcję nagród kumulatywnych. Reinforcement learning to sposób uczenia maszynowego, w którym agent otrzymuje feedback w postaci nagród lub kar za podejmowane działania. W uczeniu maszynowym system kar i nagród jest częścią procesu uczenia się, w którym agent otrzymuje nagrody za podejmowane działania, które prowadzą do pożądaných wyników, a kary za działania, które prowadzą do niepożądanych wyników.
Combined	Odpowiedź na pytanie kursanta: System kar i nagród w uczeniu maszynowym to model, który uczy się wchodząc w interakcje ze środowiskiem, poprzez system kar i nagród.

Table C.5: Original Polish version of Example 2

<b>Question 38 in Polish: Czy ma znaczenie liczebność klas</b>	
<b>System type</b>	<b>System answer in Polish</b>
Baseline	Tak, znaczenie ma liczebność klas. W przypadku bardzo rzadkiej choroby i skrajnie nierównej liczebności klas: chorych i zdrowych - zarówno w zbiorze uczącym, jak i w zbiorze testowym, losowo wybranym z populacji, mamy tylko jednego chorego na 99 zdrowych. Model w takim przypadku zignorował klasę mniej liczną i automatycznie przypisał wszystkie przykłady do klasy dominującej, osiągając przy tym bardzo wysoką dokładność (accuracy) na poziomie 99%, nie spełniając swojego głównego zadania, czyli prawidłowego zdiagnozowania osoby chorej. W związku z tym, należało spodziewać się, że model, "widząc" w procesie treningu tylko pojedyncze wystąpienie tej cyfry, nie będzie w stanie nauczyć się go rozpoznawać, a zamiast tego po prostu będzie je ignorował. Na tym przykładzie widać, jak ważne jest to, by wszystkie klasy w zbiorze uczącym miały podobną liczbę reprezentantów.
Extended	Tak, znaczenie ma liczebność klas. Na przykład, w przypadku bardzo rzadkiej choroby i skrajnie nierównej liczebności klas: chorych i zdrowych, model może zignorować klasę mniej liczną i automatycznie przypisać wszystkie przykłady do klasy dominującej, osiągając przy tym bardzo wysoką dokładność, ale nie spełniając swojego głównego zadania. W takim przypadku model nie będzie w stanie prawidłowo zdiagnozować osoby chorej.
Combined	Tak, znaczenie ma liczebność klas. W przypadku bardzo rzadkiej choroby i skrajnie nierównej liczebności klas: chorych i zdrowych - zarówno w zbiorze uczącym, jak i w zbiorze testowym, losowo wybranym z populacji, mamy tylko jednego chorego na 99 zdrowych. Model w takim przypadku zignorował klasę mniej liczną i automatycznie przypisał wszystkie przykłady do klasy dominującej, osiągając przy tym bardzo wysoką dokładność (accuracy) na poziomie 99%, nie spełniając swojego głównego zadania, czyli prawidłowego zdiagnozowania osoby chorej. W związku z tym, należało spodziewać się, że model, "widząc" w procesie treningu tylko pojedyncze wystąpienie tej cyfry, nie będzie w stanie nauczyć się go rozpoznawać, a zamiast tego po prostu będzie je ignorował.

Table C.6: Original Polish version of Example 3

<b>Question 40 in Polish: Czym jest przypadek graniczny w uczeniu maszynowym?</b>	
<b>System type</b>	<b>System answer in Polish</b>
Baseline	Nie ma odpowiedzi na to pytanie w podanym kontekście.
Extended	Nie ma odpowiedzi na to pytanie w kontekście.
Combined	Nie mogę udzielić odpowiedzi na to pytanie, ponieważ w kontekście nie ma fragmentów, które mogłyby być zacytowane jako odpowiedź na to pytanie.