Executive Summary

Generative AI and Large Language Model (LLM) chatbots can be useful tools in teaching and studying. However, they use a lot of energy, and as AI is becoming increasingly common and integrated in daily life, its energy use and emission is growing at an alarming rate. Accordingly, the University of Amsterdam (UvA) is finding a sustainable, least-energy-consuming way for professors and students to use their developing UvA chatbot. The UvA chatbot is powered by pre-existing LLMs. One component of the chatbot is the "persona" feature, enabling teachers to establish an AI persona with information specific to their course and materials. This involves the selection of a LLM (including Llama 3.1 8b, Llama 3.1 70b, GPT 35 turbo, GPT 4o mini, GPT 4o, GPT 4 vision, GPT 4 turbo, and GPT 4) for their persona. Each model has a different energy consumption and level of performance, which professors have to consider when picking out the LLM for their personas. Accordingly, this project aims to 1) compare the different models on their energy consumption and performance on specific tasks (determined by students' need for the chatbot) and 2) communicate this finding to professors, providing them information to guide their decision in selecting an appropriate LLM for their persona feature.

Through literature review, we estimated the order of the different LLMs is from the least energy exhaustive to most is Llama 3.1 8b Instruct, Llama 3.1 70b , GPT 35 turbo, GPT 4 Instruct mini, GPT 4o, GPT 4 vision, GPT 4 turbo, and GPT 4. The "rule-of-thumb" is that the more sophisticated the model, the more energy it uses. However, the complexity of the model does not always translate to better performance of tasks. This can be helpful for professors to keep in mind when picking a more-sustainable model. Additionally, some literature also suggested that switching models may not be the most efficient way to conserve energy; instead, optimisation of the models may be the preferred method. As such, the ICT department can look into the feasibility and different ways of chatbot optimisation to reduce energy consumption.

The interviews show that although students from different programmes may have different needs, there is also potential for overlapping uses of the chatbot across programs. Knowing student needs proves to be helpful when ranking chatbot performance and their energy consumption, allowing professors to pick the best model for the "persona" according to student needs. As such, studies should be done on more models as the UvA chatbot expands, and further research on performances of tasks for different courses (identifying different student needs for

different subjects) may be helpful to inform professors. This may be possible as the UvA ICT is slowly starting to allow usage of the UvA chatbot for specific programmes in the next few years. Looking at students' needs for these specific programmes and courses may aid the UvA ICT in understanding how professors can cater their models for their "persona" to student needs. This can influence the ICT's department message and training for professors as the UvA chatbot is expanded to other programmes. Additionally, as of now, professors are mostly thinking of simple uses and functions for their "persona." Because of this, the UvA chatbot can start with "personas" that use smaller models first to gauge their performance and impact, further finding out whether there is a large trade-off between LLM performance and energy consumption, informing the next steps of sustainably incorporating the chatbot.

The findings above, when communicated to professors, can allow them to assess an appropriate model for "persona" by considering environmental factors (energy consumption of the different models), what usage of the chatbot is relevant to their course, and their students' needs. Professors will be more equipped to evaluate the best model for their course(s) while keeping in mind the environmental trade-off, allowing to minimise the environmental impact while maximising AI's advantage for their teachings. After being informed of the environmental impact of AI, Professors interviewed said they will keep it in mind when using the UvA chatbot and designing their personas. This highlights the importance of finding clear, effective ways to communicate AI's environmental effects to faculty. Because AI is more integrated in education than ever, it is important to carefully assess and use it as a tool when continuing to incorporate it. The environmental repercussions are dire, so being well-informed when using these tools for teaching and learning can be a large step in using AI responsibly.

Relevance

The consideration of the environmental cost of AI chatbot usage is increasingly important as the use of AI tools becomes more common. Finding the best way to optimise the energy consumption of LLM chatbots is crucial to make AI more sustainable. Making the information about the environmental costs of AI chatbot use more readily accessible to the users, and in this case the UvA professors, is important in order to ensure the transparency of these technologies for everyone.

The project is also relevant to our partner (the UvA ICT department), as the city of Amsterdam's electricity grid has reached its maximum capacity, meaning that any additional electricity supply is not currently possible. This means that the university must utilise the energy it has access to in the most optimal way. UvA also aims to produce 10% of the energy it uses sustainably by 2026, so unnecessary energy use should be minimised to make this goal more realistic.

Therefore, the research questions we formed are as follows:

- How can the UvA chatbot's AI models' performance and energy consumption be compared?
    - What is the estimated trade-off between chatbot performance and energy consumption of inference?

- How should FMG UvA professors be informed on their choice of model for their own chatbot persona?
    - What do FMG UvA students and professors want from the chatbot?
    - How can the FMG UvA professors best be informed on the optimal AI model choice for their chatbot persona?

Methodology

Literature Review

Information on all AI energy consumption will be public in August 2026 when the rules of the EU Artificial Intelligence Act begin to be enforced. Until then, only an estimation of the energy usage of the models for this project can be made. While artificial intelligence has been studied

widely, there is limited research on the comparison of different large language models' energy consumption. We looked for any literature mentioning the energy usage, cost or size of the models we are researching. Cost and size were relevant as they could potentially be related to the energy usage of the model. Helpful articles we found were "The Price of Prompting: Profiling Energy Use in Large Language Models Inference" (2024)—which included a breakdown of the energy consumption of different Llama versions—and "From Words to Watts: Benchmarking the Energy Costs of Large Language Model Inference" (2023). The project requires the comparison of seven GPT and Llama models, and these sources help us to estimate the Llama models' energy consumption. "Reducing the Carbon Impact of Generative AI Inference (today and in 2035)" (2023) discusses the energy cost of the inference of ChatGPT, which is based on GPT models. "Carbon Emissions and Large Neural Network Training" (2021) includes GPT3 in the models studied. These sources still leave us with no information on GPT4's energy usage, making the complete comparison and analysis of the UvA chatbot's model options challenging. OpenAI's own research releases gave some insight into the differences in performance between their GPT models, but energy usage is not stated. The comparison frameworks in the different papers also cannot be unified, posing a challenge in comparing the different models for this project. Hence, the comparisons of the different LLMs in this paper are mere estimates.

Interview Methodology and Interviewee Selection

Following the stakeholder analysis (see in appendix), professors and students were identified as the primary users of the UvA chatbot persona feature to be interviewed. We designed an interview guide for both groups, as attached in the appendix. The interview guide for students was centred on previous usage of LLM chatbots and probable use of the UvA chatbot persona feature. The interview guide for teachers centred on the incorporation of LLMs in academic environments, the possible use of the UvA chatbot, and gauging the tradeoff between sustainability and performance when choosing a LLM for their persona feature. Interviewee selection was determined initially using purposive selection, as professors and students were selected based on faculty; as part of this, we selected the faculty of Social and Behavioral sciences (FMG). Furthermore, for students, additional criteria established was if they were bachelors students. This was done in order to establish a clear scope, as possible use of the chatbot would feasibly vary depending on faculty and degree level. A clearer scope was preferred, as focused data collection was preferable. Snowballing sampling was utilised whilst conducting interviews with professors within the FMG faculty, as when conducting interviews

with professors we asked interviewees to refer to other professors with possible greater expertise with technical use and incorporation in education. We utilised this sampling technique in order to gain a greater breadth within the narrower scope of the faculty of FMG.

Interview Analysis

For the student and teacher interview analysis, a “code guide” and two “code networks” are made. The “code guide” contains one “code book” and two separate “code-occurrence and code-document analyses”. These are all made in the software of ATLAS.ti and can be found in the “Interview analysis” appendix.

ATLAS.ti Report: Code Book

When making the codes, we wanted a clear analysis of the answers given by the student. We approached the analysis at the start from a deductive standpoint. We had some preconceived notions of the themes and answers, by writing and conducting the interviews. We had predetermined what kind of quotes we wanted as an outcome and wrote our code based on that. All the codes, except for what study programme the student follows, are divided between two code categories: expectations and potential needs of UvA chatbot and experience of alternative LLM chatbots. These were the two major themes of the interview. This way the codes are structured and clear. Each code in the code categories represents a question we think is important to analyse its associated quote. In addition, all codes made for the student interview belong to the code group “student interview”. When we had made the code categories and code groups, we started analysing inductively. We went through the documents of the student interviews and made codes based on the answers that were given by the students. We choose to make codes that fit the answers of the interviewees, so the code analysis would be more precise. The teacher interview codes are made in the same fashion as the student interviews. First the codes are made deductively, with the information available beforehand, then were created inductively through writing codes while reading the interviews. The code book is made to answer our two interview research questions: “What is UvA students’ experience with using existing LLM chatbots, and what is their potential interest/initial perception in the UvA chatbot?” and “To what extent professors are willing to sacrifice performance for more sustainable LLM models when choosing a LLM chatbot for their persona?”. The structure of the codes makes it

easy to look at the main themes, sub-questions, and fitting answers from interviewees and to conclude an answer. To answer the research question, we will look at the quotations that we highlighted through the codes.

ATLAS.ti Report: Co-Occurrence and Code-Document Analyses

In the co-occurrence analysis of the student interviews, all the student interview codes are mapped against the exact same codes to check for co-occurrences in the documents of the transcribed student interviews. This is also done in the teacher interview co-occurrence analysis. For the code-document analysis, both student and teacher, all the codes are plotted against the documents. This scheme shows how often each code occurs in each document.

ATLAS.ti: Code Networks

The student code network is structured through the framework of the code book and the way in which the interviews were conducted. There are two main themes in the interviews, and these have associated nodes in the code network. There is a symmetrical arrow between the two. The main theme nodes start with a capital Y or Z and have arrows pointing to the sub questions that fit specifically them. The sub questions have asymmetrical arrows among themselves, called "Follow up Q" showing the flow of the interview. The sub questions, these all start with Y1, Y2 etc. or Z1, Z2 etc., have arrows pointing to the codes that represent the answers of the students. The study programme node starts with a capital X and is separate from the rest, except from the node Y1 that represents the code, which is the follow-up question to provide further insights on the overarching themes.

The code network for the teacher interviews is structured in the same way as the student's is structured. In these interviews there were three themes, so there are three associated nodes in the network. The nodes start with Q, R, and S, and the nodes for the interview questions with Q1, R1, S1, etc. The interview question nodes branch out into the nodes of the codes that represent the teachers' answers. There are asymmetrical arrows that represent the flow of the interview called "Follow up Q" also seen in the student interview code network. The study programme that the interviewee teaches has a node, whose name starts with a capital P.

Results

With the literature review, we aimed to answer our research question “What is the estimated trade-off between chatbot performance and energy consumption of inference?”. The literature helped us understand that generally the larger (more parameters) and more complex models use more energy than smaller and simpler models. This “rule-of-thumb” was used in creating our estimated guideline for the UvA professors for our infographic. However, we were unable to properly benchmark all the models as available data and research was limited and establishing a unified comparison framework was not possible given expertise limits and time constraints.

Using both the interviews conducted amongst 11 FMG students and 3 FMG professors, we sought to answer the aforementioned research question. Said interviews allowed us to understand the needs and expectations of students regarding their possible use of the UvA chatbot. This allowed us to inform teachers and gauge their expectations and willingness to sacrifice performance for more sustainable choices when choosing one of the several LLMs available for the persona feature. Using a code co-occurrence generated by ATLAS.ti we are able to examine and analyse prevalent answers and themes that emerged across interviews from both stakeholders, which help us better understand the perceived trade-off. The codes and themes generated from the teacher interviews are most suitable to answering said research question. Due to student interviews being conducted as supplemental research and added information for the design of the infographic. The first major theme to emerge from the code co-occurrence, surrounded the incorporation of LLMs into higher education (see appendix). Many teachers stated efficiency as a result of the use of LLMs as a benefit to their incorporation. Furthermore, a prevailing sentiment that LLMs are part of the future of education emerged. However, a sub-theme emerged regarding concerns about the use of LLM’s and plagiarism. Teachers stated that the overuse of LLMs as a replacement for authentic work, rather than as a tool, posed a problem for academic integrity. In response to questions regarding the possible use of the UvA chatbot, there is a low co-occurrence as all teachers interviewed stated various uses of the persona feature. For example, possible uses included research assistance, practical communication with students, and inspection of exams. However, a theme that the teachers simply required simple uses and functions was prevalent. One teacher stated, “It doesn’t have to be fancy at all.” This leads to the final theme, specifically surrounding the trade-off between chatbot performance and energy consumption. As to how teachers weighed sustainability versus performance, one teacher stated that “it is definitely something I would keep in mind if there

were alternatives." Another teacher stated that "if the benefits outweighed the costs, and it was good, and it could be done sustainably, then I'm fine for it." However, said teacher would not use the LLM if they believed the inverse to be true. The final teacher interviewed stressed performance and specifically accuracy as the utmost important attribute when choosing an LLM for their persona feature. Having said that, in response to the question "Would you choose a more sustainable LLM given the choice" there was a high co-occurrence, with all teacher interviewees responding "Yes". However, one interviewee once again stated accuracy was the most important criterion, and that their choice of LLM pertained more to accuracy rather than sustainability.

Discussion

With the literature review, a "rule-of-thumb" was created to estimate the model complexity and performance regarding its energy consumption, allowing UvA chatbot's AI models' performance and energy consumption to be loosely compared. The literature also informs about the model performance for specific tasks. However, the degree of the difference in energy consumption remains indeterminate and unquantifiable. Additionally, due to the lack of literature, inferences drawn in this study, particularly when determining model effectiveness for specific student needs, are drawn from claims of OpenAI, developer of ChatGPT and the GPT models. The lack of third-party study to support OpenAI's claim regarding its LLMs performances makes this research (with its product—the infographic) an evolving project as more studies on this are conducted. It opens up room for further study on specific models' speciality, which can aid with optimisation. This may be helpful as many articles are focused on lowering the energy consumption of AI models by different optimisation strategies, suggesting optimisation of AI may also be an impactful action that addresses the question of model performance and energy consumption. Looking further into the technicalities of making this feasible may also be useful to allow the UvA chatbot to become more environmentally friendly.

Using the interviews, students' needs for the UvA chatbot are identified. While there may be different student needs with different student programs, there are also overlapping potential usage of the chatbot. Professors have to know this for their programme and course to effectively weigh the models' performance and energy usage when creating their persona pertaining to what students in their course need. Because of this, there can be more extensive studies or

surveys on students' needs in the particular course or programme to better inform professors. This is particularly important because it aids with the sustainable incorporation and use of AI chatbot. By knowing what their students specifically want and having information about which model is best suited for the task while using the least energy, professors can choose the most appropriate model, instead of solely choosing the most sophisticated (and often more energy consuming) model. Professors would not simply pick the best model because “it gets the job done” or because picking other models can “hinder” their students’ abilities, especially when they are aware that some task does not require a complex model and that for some task, the more complex models perform the same way as the less complex model. (For example, for student needs like summaries of lessons, re-explaining concepts, and defining terms, GPT 4 turbo and Llama 3.1 70b are equally suited for the tasks, even though GPT 4 turbo consumes more energy as a more complex model). From the professors interviewed, after being informed about the environmental effects of AI, they have stated they would keep in mind when using the UvA chatbot and creating their persona. This response may not be representative of the entire faculty and may not be a reflection of what professors actually would do as it may be response bias. Nevertheless, considering that some professors did not even know about the impact of AI on the environment, letting them know about it allows them to be a more well-informed user that may use chatbots more responsibly. Hence, letting professors know about what their students need from the chatbot, the energy consumption of the LLMs, and the environmental impact of generative AI allows them to select the most appropriate model for their courses. As an extension, finding out ways to effectively and concisely communicate this information to professors also plays a large role.

Conclusion

This study underlines the importance of balancing energy consumption in relation to performance in Large Language Models (LLMs) for educational needs and general use. The findings suggest that larger and more complex LLMs with more parameters typically consume more energy than the models with fewer parameters.

This insight is quite relevant as the UvA is in the process of developing an AI chatbot, where teachers are obligated to choose an LLM for their so-called “Persona”. This Persona feature makes this UvA chatbot unique by presenting a personalised experience for the student through the upload of, for example, course material by the teacher, so it is accessible for the student at

any time. We had to establish a guideline for teachers due to this feature to ensure they are using this chatbot sustainably by choosing the right LLM for their use to implement in the persona. We used a scale as a guideline showcasing energy consumption corresponding to the student needs from the UvA chatbot, which we discovered through our research in interviews. Beyond the student needs, the interviews indicated the impact of AI usage in an academic field, which was mainly positive.

This study's limitations, particularly the lack of data on the newer models of OpenAI (ChatGPT), underscore the need for further research in the future. Since all data on energy consumption must be published by 2026 due to the AI act, the study could then be carried out more precisely and in more detail. Moving forward, the continuous usage of AI tools in an academic environment will need guidelines to reinforce more sustainable decision-making.

References

Chien, A. A., Lin, L., Hai Binh Nguyen, Rao, V., Sharma, T., & Rajini Wijayawardana. (2023). Reducing the Carbon Impact of Generative AI Inference (today and in 2035). *Reducing the Carbon Impact of Generative AI Inference (Today and in 2035)*, *9798400702426*. https://doi.org/10.1145/3604930.3605705

Husom, E. J., Goknil, A., Shar, L. K., & Sen, S. (2024). *The Price of Prompting: Profiling Energy Use in Large Language Models Inference*. ArXiv.org. https://www.arxiv.org/abs/2407.16893

OpenAI. (2024). *Research*. Openai.com. https://openai.com/news/research/

Patterson, D., Gonzalez, J., Le, Q., Liang, C., Munguia, L.-M., Rothchild, D., So, D., Texier, M., & Dean, J. (2021). Carbon Emissions and Large Neural Network Training. *ArXiv:2104.10350 [Cs]*. https://arxiv.org/abs/2104.10350

Samsi, S., Zhao, D., Mcdonald, J., Li, B., Michaleas, A., Jones, M., Bergeron, W., Kepner, J., Tiwari, D., & Gadepally, V. (2023). *From Words to Watts: Benchmarking the Energy Costs of Large Language Model Inference*. https://arxiv.org/pdf/2310.03003

Strubell, E., Ganesh, A., & McCallum, A. (2019). Energy and Policy Considerations for Deep Learning in NLP. *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*. https://doi.org/10.18653/v1/p19-1355

Universiteit van Amsterdam. (2023, November 17). *Sustainable energy generation on campus*. University of Amsterdam. https://www.uva.nl/en/about-the-uva/about-the-university/sustainability/sustainable-operations/sustainable-campus/sustainable-energy-generation-on-campus.html

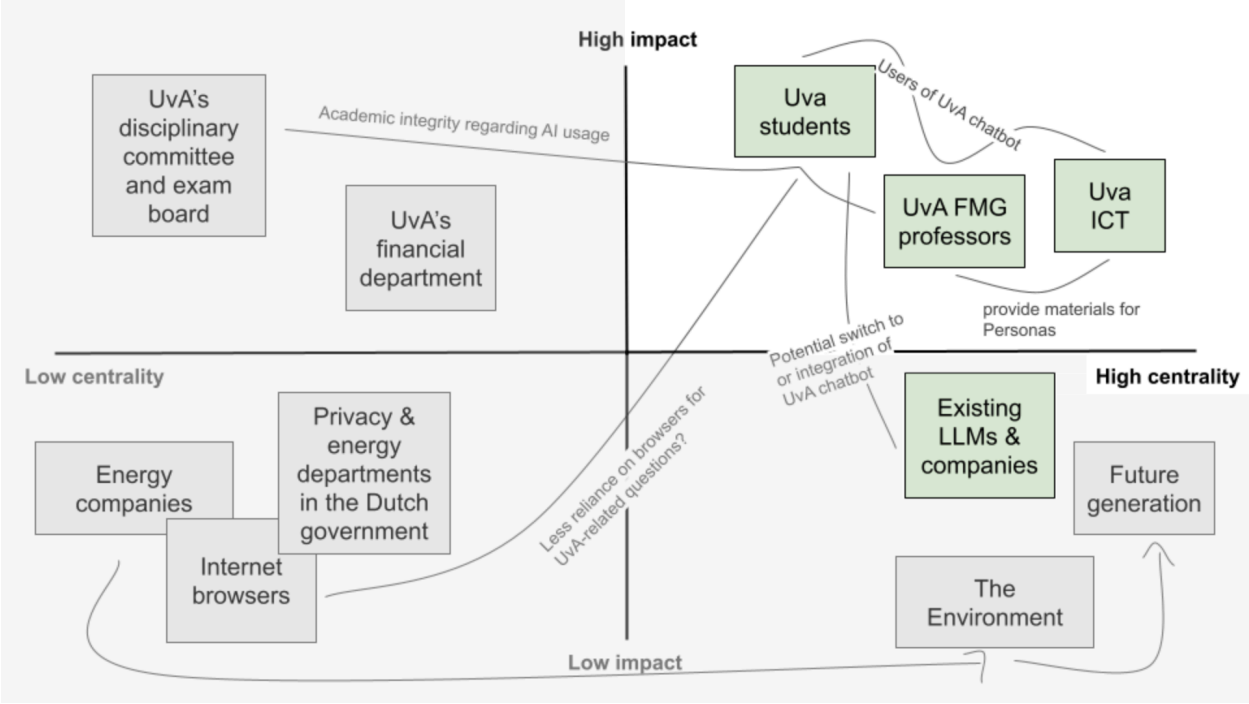
Appendices

Stakeholder analysis (Week Goal 1)

The University of Amsterdam (UvA) ICT department is developing a UvA-specific chatbot, powered by different existing large-language models (LLMs), namely ChatGPT by OpenAI and Llama by Meta. Within the UvA chatbots, there is a mini-chat or "Persona" feature, where teachers can add information specific to their course (slides, canvas information, readings, etc.). This way, students can have answers more specific to their studies.

However, generative AI and chatbot models use a lot of energy. As such, one of the challenges for UvA ICT is how to inform and educate professors to select a model that is most suitable to their purpose while keeping in mind the energy usage of the different models.

When evaluating which stakeholders would be a part of this analysis, we started by looking at the potential entities at various levels of analysis. We have identified some relevant stakeholders, shown in the map below:



Since the topic of efficient usage of the UvA chatbot is specific to the university, it is more relevant if we look at a local level, with a focus on the Faculty of Social and Behavioural Sciences (FMG). This chatbot is developed by the UvA ICT for the UvA community, so naturally, UvA FMG students, professors, and other departments are high on the impact and centrality. Since this chatbot can impact how they navigate around and interact with the university, making them high on the impact scale; they are also our targeted audience, making them high on the centrality scale. Accordingly, our rule for determining the stake-holders is to focus on Quadrant 2, where the entities are the most central to and impacted by our project. The only exception to this rule will be existing LLMs; this is because while this project will not

impact them, existing LLMs and their companies are extremely central to the project and can be influential as existing LLMs are used for UvA chatbot.

We have identified and analysed their needs, constraints (challenge to the project), interests, and diverging perspectives in the stakeholder matrix below:

Stakeholder Group	**Needs**	**Constraints**	**Interests**	**Insights**
UvA chatbot developers	- professors understand the energy usage of different LLMs, influencing their decision when creating their "Persona"	- may add more LLMs in the future	- developing a chatbot that can be used within the UvA community	- research required: environmental impact of AI & benchmarks for different models - should have a well-documented methodology for future use
UvA FMG professors	- give students knowledge and materials	- limited time -> may choose the "best" model because it gets the work done	- transferring knowledge to students and answering questions most time-efficiently	- target audience - learning the different models might be time-consuming - keep in mind their limited time when designing infographic => user-friendly interface
Existing generative AI chatbots/LLMs	- more data & infrastructure to train chatbots - more users to be profitable	- companies are not transparent with their energy consumption data	- the companies would want to widen their profit margin	- as companies are not transparent with their energy consumption data, benchmarking and comparing the different models is a challenge

UvA students	- access to information to help them answer questions about university (i.e. assignments, campus, etc.)	- the chatbot will have to be meet student needs => professors need to consider this when choosing models	- easy ways to access information	- they are the target audience that professor will keep in mind when training the models - eventual users of the product
Insights	- common need: transfer of knowledge -> should investigate ways to make sure information transfer well (ex: subjects & AI compatibility)	- research needed to simplify information about AI & to design good user interface	- companies interest leads to constraints as there is a lack of information on the energy consumption of chatbots, going against the interest of the project, UvA ICT, and teachers who want to know this information when picking a chatbot => different way of measure environment impact of LLMs is needed	

While there are little to no conflicting needs, the interests and constraints may act as challenges that should be remembered and accommodated to successfully achieve our goal and get our message across with our infographic. Namely, the UvA professors (main target audience) need to transfer knowledge that aligns well with the purpose of the chatbot, which in extension can make them more receptive to information about it, including our infographic. Yet simultaneously, as they have limited time, they may want to find the quickest route to train the models. This may mean picking the "best" model instead of the most energy-efficient one, as it takes more time to know which one is the most energy-efficient model, and the "best" model gets the job done. As such, when creating the product, time efficiency should be considered when designing the user interface to encourage more professors to learn about the information. Additionally, companies do not provide information on energy consumption of their LLMs, may be due to interest of profit or lack of ability to measure; however, the lack of information go against the interest of this project, the UvA ICT when developing their chatbots with the environment in might, teachers when picking the models, and so on. Accordingly, looking at different ways to estimate energy consumption of LLMs is needed; for example, the location of data centres, different parameters implemented when developing the LLMs, size of data infrastructure, etc.

Additionally, for this project to be successful, we have to rely on UvA Chatbot developers to develop an effective chatbot, who also depend on professors to develop and train these chatbots. This creates a power dynamic wherein their interest should be considered for the chatbot's success and as an extension, our project. While this detail allows them to be ranked high on the scale of centrality (and power if included), it is also something that we need to keep in mind as being beyond the scope of our control for this project. Our reliance on the ICT department should also be noted when we interact with them as partners, as they are also relying on us to make sure that students understand how to use this chatbot in a responsible and energy-efficient manner.

Agreement with the partner (Week Goal 2)

Research proposal:

Research question
How can the UvA chatbot's AI models' performance and energy consumption be compared, and how should UvA professors be informed on their choice of model for their own chatbot persona?

- What do UvA students and professors want from the chatbot?
- *What is the trade-off between chatbot performance and energy consumption of inference?* (likely we will only be able to offer an estimation instead of a clear answer due to lack of existing sources discussing specific AI models)
- How can the professors best be informed on the optimal AI model choice for their chatbot persona? (we will limit our focus to one department)

Relevance of project
The consideration of the environmental cost of AI chatbot usage is increasingly important as the use of AI tools becomes more common. Finding the best way to optimise the energy consumption of large language model chatbots is crucial to make AI more sustainable while society transitions to using renewable energy sources. Making the information about the environmental costs of AI chatbot use more readily accessible to the users, and in this case the UvA professors, is important in order to ensure the transparency of these technologies for everyone.

The project is also relevant to our partner (the UvA ICT department) as the city of Amsterdam's electricity grid has reached its maximum capacity, meaning that any additional electricity supply is not possible. This means that the university must utilise the energy it has access to in the most optimal way. UvA also aims to produce 10% of the energy it uses sustainably by 2026 and unnecessary energy use should be minimised to make this goal more realistic.

Background research
While artificial intelligence has been studied widely, there is limited research on the comparison of different large language models' energy consumption. The most helpful article regarding this was "The Price of Prompting: Profiling Energy Use in Large Language Models Inference"

(2024), which included a breakdown of the energy consumption of different Llama versions. As the project requires a comparison of seven different Llama and GPT models, this source is useful in benchmarking the energy usage of some. The results of the study show that the smallest version of the model uses only a fraction of the energy during inference when compared to the largest Llama model. Another similar paper, “From Words to Watts: Benchmarking the Energy Costs of Large Language Model Inference” (2023) also studied the energy use of Llama. “Reducing the Carbon Impact of Generative AI Inference (today and in 2035)” (2023) discusses the energy cost of the inference of ChatGPT, which is based on GPT models. “Carbon Emissions and Large Neural Network Training” (2021) includes GPT3 in the models studied. These sources still leave us with no information on GPT4’s energy usage, making our comparison and analysis of the UvA chatbot’s model options more challenging. Many articles are focused on lowering the energy consumption of AI models by different optimisation strategies, but as our project is focused on choosing the smallest effective model for the professors creating their personas on the chatbot, we will only focus on comparing the models as they are. In the paper “ChatGPT—A Challenging Tool for the University Professors in Their Teaching Practice” (2023) it’s stated how university professors use artificial intelligence: “They view ChatGPT as a means to support time-consuming teaching activities, provoke interest, activate and engage students, and stimulate their critical thinking and creativity” (Kiryakova and Angelova, 2023, p.17).

Schedule for project

Week 1:

Due: Stakeholder Analysis - Mai

- Progress Check - 18/9
- Final - 29/9

- ☑ ~~Tuesday 17/9: Team Charter~~
- ☑ ~~Wednesday 18/9: Individual Research & Questions~~
- ☑ ~~Wednesday 18/9: Stakeholder draft~~
- ☑ ~~Wednesday 18/9: Partner meeting agenda~~
- ☑ ~~Thursday 19/9: Stakeholder Analysis~~

Week 2:

Due: Project Proposal - Maria & (research help: Phoebe)

- Progress Check - 26/9
- Final - 27/9

- ☑ ~~Wednesday 25/9: Project Proposal draft & comments~~
- ☑ ~~Thursday: Progress check & amend~~
- ☑ ~~Friday: “Final” proposal & send to Danny~~
    - may have edits depending on Danny & Regina

Week 3:

Due: Design Interview Guide - Niko

- Progress Check - 2/10
- Final - 3/10

-> Meeting with Danny on Tuesday at 16

- Schedule interviews

➔ Goal: Narrow down to a specific department

<u>Week 4:</u>
- Conduct & transcribe interviews - Niko

➔ Goal:
- Have a brief understanding of students' interest, needs, and wants regarding UvA chatbot
- Learn about prior similar projects through interview

<u>Week 5:</u>
- Conduct & transcribe interviews - Niko

➔ Goal: Have a brief understanding of teachers' interest, needs, and wants regarding UvA chatbot

Week 6:
Due: Analysis of Interviews - Phoebe
- Progress Check - 23/10
- Final - 24/10

Meeting with Danny

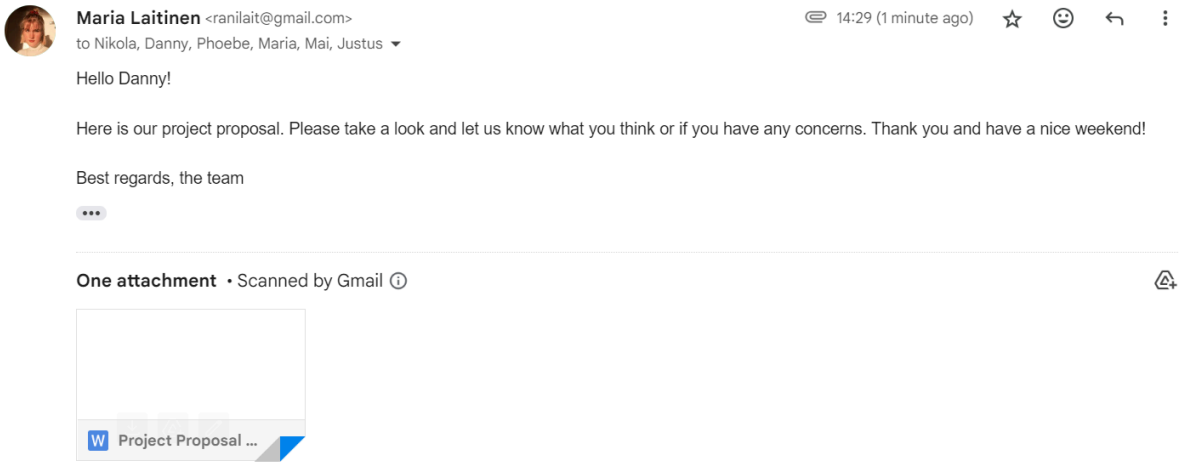
➔ Goal:
- Transcribe and conduct analysis on interviews to draw reasonable conclusions for reports and infographics
- Have a potential outline/draft for the infographic

<u>Week 7:</u>
Due: INFOGRAPHIC 30/10 - Justus

➔ Goal:
- Have a complete infographic
- Have an outline/draft of the Research Report

<u>Week 8:</u>
Due: Research Report 4/11

Communication with project partner:



The project proposal has been adapted according to the project partner's requests after initial communication.

Bibliography

Berthelot, A., Caron, E., Jay, M., & Lefèvre, L. (2024). Estimating the environmental impact of Generative-AI services using an LCA-based methodology. *Procedia* CIRP, 122, 707–712. https://doi.org/10.1016/j.procir.2024.01.098

De Vries, A. (2023). The growing energy footprint of artificial intelligence. *Joule*, 7(10), 2191–2194. https://doi.org/10.1016/j.joule.2023.09.004

From Words to Watts: Benchmarking the energy costs of large language model inference. (2023, September 25). IEEE Conference Publication | IEEE Xplore. https://ieeexplore.ieee.org/document/10363447

Husom, E. J., Goknil, A., Shar, L. K., & Sen, S. (2024, July 4). *The price of prompting: Profiling energy use in large language models inference*. arXiv.org. https://arxiv.org/abs/2407.16893

Chien, A. A., Lin, L., Nguyen, H., Rao, V., Sharma, T., & Wijayawardana, R. (2023). Reducing the Carbon Impact of Generative AI Inference (today and in 2035). *2nd Workshop on Sustainable Computer Systems*. https://doi.org/10.1145/3604930.3605705

Liu, V., & Yin, Y. (2024, April 1). Green AI: *Exploring carbon footprints, mitigation strategies, and trade offs in large language model training*. arXiv.org. https://arxiv.org/abs/2404.01157

Luccioni, S., Jernite, Y., & Strubell, E. (2024). Power Hungry Processing: Watts Driving the Cost of AI Deployment? *FAccT '24: Proceedings of the 2024 ACM Conference on Fairness, Accountability, and Transparency*. https://doi.org/10.1145/3630106.3658542

Patterson, D. A., Gonzalez, J., Le, Q. V., Liang, C., Munguía, L., Rothchild, D., So, D. R., Texier, M., & Dean, J. (2021). *Carbon emissions and large neural network training*. https://www.semanticscholar.org/paper/Carbon-Emissions-and-Large-Neural-Network-Training-Patterson-Gonzalez/7a16d9b4e04300d034502dc7dd58428714594e2c

Kiryakova, G., & Angelova, N. (2023). ChatGPT—A challenging tool for the university professors in their teaching practice. *Education Sciences*, 13(10), 1056. https://doi.org/10.3390/educsci13101056

IS PROFESSORS' PERSPECTIVES ON AI-ASSISTED PROGRAMMING. (2023). *Issues in Information Systems*. https://doi.org/10.48009/2_iis_2023_115

Universiteit van Amsterdam. (2024, July 3). *Impact of overloaded electricity grid on UvA's new construction and sustainability.* Universiteit Van Amsterdam. https://www.uva.nl/shared-content/uva/en/news/news/2024/07/impact-of-overloaded-electricity-grid-on-uvas-new-construction-and-sustainability.html

Universiteit van Amsterdam. (2023, November 17). *Sustainable energy generation on campus*. University of Amsterdam. https://www.uva.nl/en/about-the-uva/about-the-university/sustainability/sustainable-operations/sustainable-campus/sustainable-energy-generation-on-campus.html

Verma, P., & Tan, S. (2024, September 18). A bottle of water per email: the hidden environmental costs of using AI chatbots. *The Washington Post.* https://www.washingtonpost.com/technology/2024/09/18/energy-ai-use-electricity-water-data-centers/

Interview guide (Week Goal 3)

Interview guide 1 for students:

RQ: What is UvA students' experience with using existing LLM chatbots, and what is their potential interest/initial perception in the UvA chatbot?

<table>
<tr><td colspan="2">Introduction - Briefly explain the following to the interviewee (do not make it too "heavy" or fraught):
□ Goal of the interview: Gauge students potential need/interest in the UvA chatbot to inform Professors
□ Briefly go through the key points of the Consent Form
□ Ask for recording the interview ("Is it OK for you if I record the interview?")</td></tr>
<tr><td>Subtheme 1: Experience of alternative LLM chatbots

Goal: Understand how and if students use LLM chatbots already.</td><td>Introduction: "As part of our research we are interested in UvA students' experience using large language model chatbots e.g. Chatgpt, Gemini and Llama."

Initial question:
"Have you used AI chatbots for your studies?"<br>"If yes, for what purposes?"

Potential follow-up questions and probes:
"Do you feel that using AI impacts your academic skills?"

"Do you think that AI chatbots have influenced your study habits?"

"Do you think AI chatbots have improved your university experience?"</td></tr>
</table>

	“**Why did you first try ChatGPT/AI chatbots?”**
Demo run of UvA chatbot	
Subtheme 2: Expectations and potential needs of the UvA chatbot **Goal:** Gain information on students expectations and potential needs to inform professors	**Introduction:** “Having quickly gone through a demo run of the UvA chatbot, I would like to know more about your initial thoughts, and looking forward more about your expectations and needs of the chatbot.” **Initial question:** “Do you think the UvA chatbot, specifically the persona feature, will impact your studies?” **Potential follow-up questions and probes:** “Do you think you would use the persona feature?” “What would you want the persona feature to be able to tell you?” “What impact do you think the UvA chatbot will have on you? If none, why?" “Would you imagine yourself using this chatbot? How?” “If this is rolled out for students to use, do you think you would actually use this over existing chatbot models?”
Wrapping up the interview. Thank the interviewee for taking time to sit down with you.	

Interview guide 2 for Professors:

RQ: To what extent professors are willing to sacrifice performance for more sustainable LLM models when choosing a LLM chatbot for their persona?

<table>
<tr><td colspan="2">Introduction. Briefly explain the following to the interviewee (do not make it too “heavy” or fraught):
□ Goal of the interview: Understand the trade off point between performance and sustainability for professors in choosing the LLM persona.
□ Briefly go through the key points of the Consent Form
□ Ask for recording the interview (“Is it OK for you if I record the interview?“)</td></tr>
<tr><td>Subtheme 1: Possible use of LLM chatbots in educational environment

Goal: Gauge teachers willingness to incorporate LLM chatbots into their teaching</td><td>Introduction: “To begin the interview I would like to understand your thoughts on potentially using LLM chabots in academic environments.”

Initial question: “With the rise in use of LLM chatbots, do you believe they should be incorporated into the academics of higher education?”

Potential follow-up questions and probes:
“In what ways do you think LLM chatbots can be used as an effective and useful academic tool?”

“Do you think the increased use of LLM chatbots has impacted the studies of students in your course?”

“What are some of your primary concerns surrounding the use of</td></tr>
</table>

	LLM chatbots in academic environments?"
Demo run of UvA chatbot	
Subtheme 2: Expectations and potential needs of the UvA chatbot **Goal:** Gain information on teachers expectations and needs to understand their use of the chatbot.	**Introduction: "**Having demonstrated the UvA chatbot, I would now like to understand your possible expectations and needs of the chatbot." **Initial question:** "Do you think you would use the UvA chatbot?" "What would be your primary use?" **Potential follow-up questions and probes:** "How do you think the chatbot, more specifically the persona feature, could impact both yourself and your students?" "Do you think you would use this tool with your course?"
Subtheme 3: The trade off between performance and sustainability Goal: Determine whether or not teachers would be willing to sacrifice performance for more sustainable options	**Introduction:** Despite the functionality of such LLM's, they also pose a serious issue in terms of sustainability and energy efficiency, which is why I would like to know more about your willingness to sacrifice the performance of your persona by choosing a more sustainable LLM." **Initial question:** "Would you be willing to choose a more energy efficient LLM over perhaps a more powerful LLM?" **Potential follow-up questions and probes:**

	"How would you weigh sustainability versus effectiveness?" "For which functions of the chatbot would you be willing to lower performance standards in favour of increased energy efficiency?"
Wrapping up the interview. Thank the interviewee for taking time to sit down with you.	

Consent form:

CSSci Standard Consent Form

Energy usage of UvA AI Chatbots

Aim of the study

This project is part of the educational program at the Computational Social Science bachelor program of the University of Amsterdam. The results of this study will be used for educational purposes.

By the end of the project, students will develop an informative visual aid regarding sustainable use of the UvA LLM chatbot.

The purpose of this study is to assess the energy consumption of the varying LLMs available to UvA professors when choosing their "Persona" and in turn create a visual aid to allow professors to make an informed choice, balancing performance, needs and energy efficiency.

Participation will involve responding to questions regarding potential needs, interest and use of the UvA chatbot (If applicable), as well as questions concerning general expertise of LLMs, more specifically energy consumption of LLMs.

Confidentiality

Your privacy is protected as a participant in this study. Any reports generated might use paraphrased wording or quotes and can be attributed to your name. If you would not like your responses to be identifiable, you have a right to mention this to the interviewer. In this case, we will make your responses unidentifiable and use only paraphrased wording or quotes that cannot be used to identify you.

Withdrawal

Participation in this study is completely voluntary; it is entirely up to you to choose whether to participate or not to participate. You can withdraw at any time or refuse to answer any question without consequences of any kind. If you experience discomfort, you may discontinue the interview at any time.

Contact

The study is conducted by Justus Tödtmann, Maria Laitinen, Mai Bui, Phoebe Poort and Nikola Kresojevic, students of the University of Amsterdam. To contact the researchers, Nikola Kresojevic can be reached at +44 7761715031 or by email nikola.kresojevic@student.uva.nl This project is supervised by Regina Nockerts, r.a.nockerts@uva.nl whom you can contact for additional questions or concerns about this research.

Signature: Date:

Interview analysis (Week Goal 4)

- Code Book
- Code Co-Occurrence and Code-Document Analyses
- Code Networks

ATLAS.ti Report: Code Book

UvA Chatbot (7)

Codes grouped by Smart codes

Report created by Phoebe Poort on 4 Nov 2024

Codes (102)

102 Codes:

P: Study programme of teacher

Created: 31/10/2024 by Phoebe Poort, **Modified:** 04/11/2024 by Phoebe Poort

Comment:
What programme the professor is a part of.

P: Study programme of teacher: PPLE professor

Created: 31/10/2024 by Nikola Kresojevic, **Modified:** 31/10/2024 by Nikola Kresojevic

Comment:
Professor of PPLE

P: Study programme of teacher: Professor in Political Economy of Finance

Created: 02/11/2024 by Nikola Kresojevic, **Modified:** 02/11/2024 by Nikola Kresojevic

Comment:
Professor in political economy of finance

P: Study programme of teacher: Professor of Political science

Created: 02/11/2024 by Nikola Kresojevic, **Modified:** 02/11/2024 by Nikola Kresojevic

Comment:
Professor of political science

Q1: How teachers think LLM chatbots should be incorporated into higher education

Created: 31/10/2024 by Phoebe Poort, **Modified:** 04/11/2024 by Phoebe Poort

Comment:
Teachers perspective on the use of LLM chatbots in higher education

Q1: How teachers think LLM chatbots should be incorporated into higher education: Could increase efficiency

Created: 02/11/2024 by Nikola Kresojevic, **Modified:** 04/11/2024 by Phoebe Poort

Comment:
This code is used when a teacher thinks that all facets of teaching could be more efficient because of LLMs.

● Q1: How teachers think LLM chatbots should be incorporated into higher education: Makes research more efficient

Created: 02/11/2024 by Nikola Kresojevic, **Modified:** 04/11/2024 by Phoebe Poort

Comment:
This code is used when a teacher thinks that research could be more efficient as a result of LLM.

● Q1: How teachers think LLM chatbots should be incorporated into higher education: Should find a way to teach skills for LLM

Created: 31/10/2024 by Nikola Kresojevic, **Modified:** 04/11/2024 by Phoebe Poort

Comment:
This code is used when a teacher thinks that the skills to use LLM's as an effective tool should be teached, as they are a part of the future.

● Q1: How teachers think LLM chatbots should be incorporated into higher education: Students learn less using LLM

Created: 02/11/2024 by Nikola Kresojevic, **Modified:** 04/11/2024 by Phoebe Poort

Comment:
This code is used when a teacher thinks that students are learning less by using chatgpt etc.

● Q2: Professor view on impact of LLMs on students

Created: 31/10/2024 by Nikola Kresojevic, **Modified:** 04/11/2024 by Nikola Kresojevic

Comment:
The view of professors of the impact of LLM's on students

● Q2: Professor view on impact of LLMs on students: Can exploit LLMs for better grades

Created: 02/11/2024 by Nikola Kresojevic, **Modified:** 04/11/2024 by Phoebe Poort

Comment:
This code is used when a teacher thinks that students who understand how LLMs algorithm works, can use them effectively for better grades.

● Q2: Professor view on impact of LLM's on students: Help with editing and reviewing final papers

Created: 31/10/2024 by Nikola Kresojevic, **Modified:** 04/11/2024 by Phoebe Poort

Comment:

This code is used when a teacher thinks that LLMs can help with editing and making sure essays are concise.

● Q2: Professor view on impact of LLM's on students: Part of the way new students learn

Created: 02/11/2024 by Nikola Kresojevic, **Modified:** 04/11/2024 by Phoebe Poort

Comment:
This code is used when a teacher thinks that newer students have already been using LLM's for some time and it is ingrained into how they study now.

● Q2: Professor view on impact of LLM's on students: Possible plagiarism

Created: 31/10/2024 by Nikola Kresojevic, **Modified:** 04/11/2024 by Phoebe Poort

Comment:
This code is used when a teacher thinks that plagiarism could derive from the use of LLMs.

● Q3: Concerns about the use of LLMs

Created: 31/10/2024 by Nikola Kresojevic, **Modified:** 04/11/2024 by Phoebe Poort

Comment:
This code category is for all the codes made for the concerns of teachers surrounding the use of LLMs.

● Q3: Concerns about the use of LLMs: Issues surrounding intellectual property

Created: 02/11/2024 by Nikola Kresojevic, **Modified:** 04/11/2024 by Phoebe Poort

Comment:
This code is used when a teacher is concerned about intellectual property rights.

● Q3: Concerns about the use of LLMs: Legal standing created by use of LLM's

Created: 02/11/2024 by Nikola Kresojevic, **Modified:** 04/11/2024 by Phoebe Poort

Comment:
This code is used when a teacher thinks that questions answered by LLMs have legal standing and legitimacy.

● Q3: Concerns about the use of LLMs: LLMs don't need personal information about students

Created: 02/11/2024 by Nikola Kresojevic, **Modified:** 04/11/2024 by Phoebe Poort

Comment:

This code is used when a teacher thinks that LLM's don't need to know personal information about students.

• Q3: Concerns about the use of LLMs: LLMs lack accuracy

Created: 02/11/2024 by Nikola Kresojevic, **Modified:** 04/11/2024 by Phoebe Poort

Comment:
This code is used when a teacher thinks that LLMs are probability machines and not always 100% accurate.

• Q3: Concerns about the use of LLMs: Plagiarism

Created: 31/10/2024 by Nikola Kresojevic, **Modified:** 04/11/2024 by Phoebe Poort

Comment:
This code is used when a teacher thinks that plagiarism can derive from using LLMs as a replacement rather than as a tool.

• Q3: Concerns about the use of LLMs: Reduced creativity in written assignments

Created: 31/10/2024 by Nikola Kresojevic, **Modified:** 04/11/2024 by Phoebe Poort

Comment:
This code is used when a teacher thinks that students are being less creative in written assignments, because of LLM usage.

• Q: Possible use of LLM chatbots in educational environment

Created: 02/11/2024 by Phoebe Poort, **Modified:** 04/11/2024 by Phoebe Poort

Comment:
This main code category is for all the codes that fit in the first goal of the teacher interviews: to gauge teachers willingness to incorporate LLM chatbots into their teaching.

• R1: Primary use of Persona feature

Created: 31/10/2024 by Nikola Kresojevic, **Modified:** 04/11/2024 by Phoebe Poort

Comment:
This code category is for all the codes that describe the primary use of the persona feature for teachers.

• R1: Primary use of Persona feature: Helping students find course information

Created: 02/11/2024 by Nikola Kresojevic, **Modified:** 04/11/2024 by Phoebe Poort

Comment:
This code is used when a teacher thinks that LLMs help students to easily find relevant course information.

R1: Primary use of Persona feature: Helping with research

Created: 31/10/2024 by Nikola Kresojevic, **Modified:** 04/11/2024 by Phoebe Poort

Comment:
This code is used when a teacher thinks that LLMs can be used by students as a tool to find sources and general research etc.

R1: Primary use of Persona feature: Inspection of exams

Created: 02/11/2024 by Nikola Kresojevic, **Modified:** 04/11/2024 by Phoebe Poort

Comment:
This code is used when a teacher thinks that LLMs can help students understand why their questions on exams were wrong etc.

R1: Primary use of Persona feature: Practical communication with students

Created: 02/11/2024 by Nikola Kresojevic, **Modified:** 04/11/2024 by Phoebe Poort

Comment:
This code is used when a teacher thinks that easy communication with students can be easily answered by LLMs in the future.

R: Expectations and potential needs of the UvA chatbot

Created: 02/11/2024 by Phoebe Poort, **Modified:** 04/11/2024 by Phoebe Poort

Comment:
This main code category is for all the codes that fit in the second goal of the teacher interviews: to gain information on teachers expectations and needs to understand their use of the chatbot.

S1: How teachers weigh sustainability versus effectiveness

Created: 31/10/2024 by Phoebe Poort, **Modified:** 04/11/2024 by Phoebe Poort

Comment:
This code category is for all the codes that describe the perceived trade off between performance and sustainability.

S1: How teachers weigh sustainability versus effectiveness: Doesn't have to be fancy, just needs to perform well

Created: 02/11/2024 by Nikola Kresojevic, **Modified:** 04/11/2024 by Phoebe Poort

Comment:
This code is used when a teacher thinks that when the chatbot is able to execute its functions well, sustainability is preferred.

• S1: How teachers weigh sustainability versus effectiveness: If the benefits outweigh the costs in relation to students and environment

Created: 02/11/2024 by Nikola Kresojevic, **Modified:** 04/11/2024 by Nikola Kresojevic

Comment:
If it is good for students and can be done sustainably

• S1: How teachers weigh sustainability versus effectiveness: Would consider alternatives

Created: 31/10/2024 by Nikola Kresojevic, **Modified:** 04/11/2024 by Phoebe Poort

Comment:
This code is used when a teacher thinks that they would consider more sustainable alternatives of LLM usage.

• S2: Would you be willing to choose a more energy efficient LLM

Created: 31/10/2024 by Nikola Kresojevic, **Modified:** 04/11/2024 by Phoebe Poort

Comment:
This code category is for all the codes that describe whether teachers would be willing to choose a more energy efficient LLM over a better performing one.

• S2: Would you be willing to choose a more energy efficient LLM: Depends on the performance standards

Created: 02/11/2024 by Nikola Kresojevic, **Modified:** 02/11/2024 by Nikola Kresojevic

Comment:
Depends on the performance standards, would have to be accurate

• S2: Would you be willing to choose a more energy efficient LLM: Yes, would use a more energy efficient LLM

Created: 31/10/2024 by Nikola Kresojevic, **Modified:** 02/11/2024 by Nikola Kresojevic

Comment:
Would choose a more energy efficient LLM

• S: The trade off between performance and sustainability

Created: 04/11/2024 by Phoebe Poort, **Modified:** 04/11/2024 by Phoebe Poort

Comment:
This main code category is for all the codes that fit in the third goal of the teacher interviews: to determine whether or not teachers would be willing to sacrifice performance for more sustainable options.

X: Study programme of student

Created: 25/10/2024 by Phoebe Poort, **Modified:** 27/10/2024 by Maria Laitinen

Comment:
What bachelor or master programme the student follows.

X: Study programme of student: Anonymous student

Created: 25/10/2024 by Phoebe Poort, **Modified:** 28/10/2024 by Phoebe Poort

Comment:
Student wants to be anonymous.

X: Study programme of student: Bachelor Business Administration

Created: 25/10/2024 by Phoebe Poort, **Modified:** 28/10/2024 by Phoebe Poort

Comment:
Student does a bachelor's in Business Administration.

X: Study programme of student: Bachelor Communication Science, 1st year

Created: 25/10/2024 by Phoebe Poort, **Modified:** 28/10/2024 by Phoebe Poort

Comment:
Student does a bachelor's in Communication Science.

X: Study programme of student: Bachelor Computational Social Science, 1st year

Created: 25/10/2024 by Phoebe Poort, **Modified:** 28/10/2024 by Phoebe Poort

Comment:
Student does a bachelor Computational Social Science.

X: Study programme of student: Bachelor Economics

Created: 25/10/2024 by Phoebe Poort, **Modified:** 28/10/2024 by Phoebe Poort

Comment:
Student does a bachelor's in Economics.

• X: Study programme of student: Bachelor Interdisciplinaire sociale wetenschap, 4th year

Created: 25/10/2024 by Phoebe Poort, **Modified:** 28/10/2024 by Phoebe Poort

Comment:
Student does a bachelor's in Interdisciplinaire sociale wetenschap.

• X: Study programme of student: Bachelor PPLE

Created: 25/10/2024 by Phoebe Poort, **Modified:** 28/10/2024 by Phoebe Poort

Comment:
Student does a bachelor PPLE.

• X: Study programme of student: Bachelor Psychology, 1st year

Created: 25/10/2024 by Phoebe Poort, **Modified:** 28/10/2024 by Phoebe Poort

Comment:
Student does a bachelor Psychology.

• X: Study programme of student: Graduated bachelor Future Planet Studies

Created: 25/10/2024 by Phoebe Poort, **Modified:** 28/10/2024 by Phoebe Poort

Comment:
Student does a bachelor Future Planet Studies.

• Y1: If students have used LLMs before or not

Created: 25/10/2024 by Phoebe Poort, **Modified:** 28/10/2024 by Phoebe Poort

Comment:
This code category is for when students have previously used LLMs before or not.

• Y1: If students have used LLMs before or not: Did use LLMs before

Created: 25/10/2024 by Phoebe Poort, **Modified:** 30/10/2024 by Nikola Kresojevic

Comment:
Whether an interviewee has previously used an LLM.

• Y2: Students' previous purpose of LLM usage

Created: 25/10/2024 by Phoebe Poort, **Modified:** 28/10/2024 by Maria Laitinen

Comment:
This code category is for when students explain the previous purpose of their LLM usage.

• Y2: Students' previous purpose of LLM usage: explain concepts

Created: 28/10/2024 by Maria Laitinen, **Modified:** 30/10/2024 by Maria Laitinen

Comment:
If student has used LLMs to ask for explanations of concepts.

• Y2: Students' previous purpose of LLM usage: Get ideas for assignment

Created: 27/10/2024 by Mai Bui (2), **Modified:** 30/10/2024 by Maria Laitinen

Comment:
If student has used LLMs to get ideas for assignments.

• Y2: Students' previous purpose of LLM usage: Has used LLM's for new ideas/creativity

Created: 25/10/2024 by Phoebe Poort, **Modified:** 30/10/2024 by Maria Laitinen

Comment:
If student has used LLMs for new ideas or creativity.

• Y2: Students' previous purpose of LLM usage: Link articles to theories/concepts

Created: 25/10/2024 by Phoebe Poort, **Modified:** 30/10/2024 by Maria Laitinen

Comment:
If student has used LLMs to link articles to theories or concepts.

• Y2: Students' previous purpose of LLM usage: Recreational

Created: 25/10/2024 by Phoebe Poort, **Modified:** 30/10/2024 by Maria Laitinen

Comment:
If student has used LLMs recreationally.

• Y2: Students' previous purpose of LLM usage: Rephrase to easier words

Created: 27/10/2024 by Mai Bui (2), **Modified:** 30/10/2024 by Maria Laitinen

Comment:
If student has used LLMs to rephrase things into easier words.

• Y2: Students' previous purpose of LLM usage: summarise articles

Created: 28/10/2024 by Maria Laitinen, **Modified:** 30/10/2024 by Maria Laitinen

Comment:

If student has used LLMs to summarise articles.

● Y2: Students' previous purpose of LLM usage: Writing support

Created: 25/10/2024 by Phoebe Poort, **Modified:** 30/10/2024 by Maria Laitinen

Comment:
If student has used LLMs for writing support.

● Y3,1: Negative impact LLMs on students' education

Created: 25/10/2024 by Phoebe Poort, **Modified:** 28/10/2024 by Phoebe Poort

Comment:
This code category is for when students experienced a negative impact on their education from using LLMs.

● Y3,1: Negative impact LLMs on students' education: Cherry picking articles in research

Created: 25/10/2024 by Phoebe Poort, **Modified:** 30/10/2024 by Maria Laitinen

Comment:
If student has used LLMs to cherry pick articles for their research.

● Y3,1: Negative impact LLMs on students' education: Learned less

Created: 28/10/2024 by Phoebe Poort, **Modified:** 30/10/2024 by Maria Laitinen

Comment:
If student has learned less due to LLMs.

● Y3,1: Negative impact LLMs on students' education: Negatively impacts my skill searching for research papers

Created: 25/10/2024 by Phoebe Poort, **Modified:** 30/10/2024 by Maria Laitinen

Comment:
If LLMs have negatively impacted the student's skills for searching for academic papers.

● Y3,1: Negative impact LLMs on students' education: Starting later with deadlines

Created: 25/10/2024 by Phoebe Poort, **Modified:** 30/10/2024 by Maria Laitinen

Comment:
If student starts with deadlines later due to LLMs.

● Y3,2: Positive impact of LLMs on students' education

Created: 25/10/2024 by Phoebe Poort, **Modified:** 04/11/2024 by Phoebe Poort

Comment:
This code category is for when students experienced a positive impact on their education from using LLMs.

● Y3,2: Positive impact of LLMs on students' education: Improved academic career

Created: 25/10/2024 by Phoebe Poort, **Modified:** 30/10/2024 by Maria Laitinen

Comment:
If LLM usage has improved the student's academic career.

● Y3,2: Positive impact of LLMs on students' education: Makes studying less time consuming

Created: 25/10/2024 by Phoebe Poort, **Modified:** 30/10/2024 by Maria Laitinen

Comment:
If LLM usage has made student's studying less time consuming.

● Y3,2: Positive impact of LLMs on students' education: Understand in easier terms

Created: 27/10/2024 by Mai Bui (2), **Modified:** 30/10/2024 by Maria Laitinen

Comment:
If LLMs have helped student understand topics in easier terms.

● Y4: Students motivation first usage LLMs

Created: 25/10/2024 by Phoebe Poort, **Modified:** 28/10/2024 by Phoebe Poort

Comment:
This code category is for when students explain their motivation for first usage LLMs.

● Y4: Students motivation first usage LLMs: Curiosity

Created: 25/10/2024 by Phoebe Poort, **Modified:** 30/10/2024 by Maria Laitinen

Comment:
If student first tried an LLM due to curiosity.

● Y4: Students motivation first usage LLMs: Recommended by someone

Created: 25/10/2024 by Phoebe Poort, **Modified:** 30/10/2024 by Maria Laitinen

Comment:

If student first tried an LLM due to someone's recommendation.

• Y4: Students motivation first usage LLMs: Student was in a time shortage

Created: 25/10/2024 by Phoebe Poort, **Modified:** 30/10/2024 by Maria Laitinen

Comment:
If student first tried an LLM due to being in a time shortage.

• Y: Experience of alternative LLM chatbots

Created: 23/10/2024 by Phoebe Poort, **Modified:** 28/10/2024 by Phoebe Poort

Comment:
This is a code category of the first main theme of the interview: experience of alternative LLM chatbots. Put no codes here please.

• Z1: If students would use the UvA chatbot or not

Created: 25/10/2024 by Phoebe Poort, **Modified:** 30/10/2024 by Maria Laitinen

Comment:
This code category is for if the students would use the UvA chatbot.

• Z1: If students would use the UvA chatbot or not: Depends on the quality compared to chatgpt

Created: 28/10/2024 by Nikola Kresojevic, **Modified:** 28/10/2024 by Phoebe Poort

Comment:
If the student would use it, depends on the quality of the chatbot.

• Z1: If students would use the UvA chatbot or not: No, student would not use the UvA chatbot.

Created: 25/10/2024 by Phoebe Poort, **Modified:** 25/10/2024 by Phoebe Poort

Comment:
If student would not use the UvA chatbot.

• Z1: If students would use the UvA chatbot or not: Would try out but prefers chatGPT if same functions

Created: 27/10/2024 by Mai Bui (2), **Modified:** 30/10/2024 by Maria Laitinen

Comment:
If student would prefer ChatGPT over UvA chatbot given they have the same functions.

• Z1: If students would use the UvA chatbot or not: Yes, student would use the UvA chatbot

Created: 25/10/2024 by Phoebe Poort, **Modified:** 25/10/2024 by Phoebe Poort

Comment:
If students would use the UvA chatbot.

• Z2: If students would use the persona feature

Created: 25/10/2024 by Phoebe Poort, **Modified:** 30/10/2024 by Maria Laitinen

Comment:
This code category is for if the student's would use the UvA chatbot's persona feature.

• Z2: If students would use the persona feature: No, student would not use the persona feature

Created: 25/10/2024 by Phoebe Poort, **Modified:** 25/10/2024 by Phoebe Poort

Comment:
If the student would not use the persona feature.

• Z2: If students would use the persona feature: Yes, student would use the persona feature

Created: 25/10/2024 by Phoebe Poort, **Modified:** 25/10/2024 by Phoebe Poort

Comment:
If the student would use the persona feature.

• Z3: Students' desired information from UvA chatbot

Created: 25/10/2024 by Phoebe Poort, **Modified:** 30/10/2024 by Maria Laitinen

Comment:
This code category is for students' desired information from the UvA chatbot.

• Z3: Students' desired information from UvA chatbot: Content and assignments

Created: 28/10/2024 by Nikola Kresojevic, **Modified:** 30/10/2024 by Maria Laitinen

Comment:
When students use UvA chatbot, they want to ask about content and assignments.

• Z3: Students' desired information from UvA chatbot: explanations of taught concepts

Created: 27/10/2024 by Maria Laitinen, **Modified:** 30/10/2024 by Maria Laitinen

Comment:
When students use UvA chatbot, they want to receive explanations of taught concepts.

• Z3: Students' desired information from UvA chatbot: practice problems

Created: 27/10/2024 by Maria Laitinen, **Modified:** 30/10/2024 by Maria Laitinen

Comment:
When students use UvA chatbot, they want to receive practice problems.

• Z3: Students' desired information from UvA chatbot: Provide additional sources

Created: 28/10/2024 by Phoebe Poort, **Modified:** 30/10/2024 by Maria Laitinen

Comment:
When students use UvA chatbot, they want to receive additional sources.

• Z3: Students' desired information from UvA chatbot: provide examples

Created: 27/10/2024 by Maria Laitinen, **Modified:** 30/10/2024 by Maria Laitinen

Comment:
When students use UvA chatbot, they want to receive examples of concepts or topics.

• Z3: Students' desired information from UvA chatbot: small summaries of classes

Created: 27/10/2024 by Phoebe Poort, **Modified:** 30/10/2024 by Maria Laitinen

Comment:
When students use UvA chatbot, they want to get small summaries of classes.

• Z3: Students' desired information from UvA chatbot: Summarise texts

Created: 28/10/2024 by Phoebe Poort, **Modified:** 30/10/2024 by Maria Laitinen

Comment:
When students use UvA chatbot, they want to receive summaries of longer texts.

• Z3: Students' desired information from UvA chatbot: the same as other existing LLMs

Created: 28/10/2024 by Phoebe Poort, **Modified:** 30/10/2024 by Maria Laitinen

Comment:
When students use UvA chatbot, they want the same qualities that other LLMs have.

● Z4: UvA chatbot potential impact on students

Created: 25/10/2024 by Phoebe Poort, **Modified:** 30/10/2024 by Maria Laitinen

Comment:
This code category is for the UvA chatbot's potential impact on the student's studies.

● Z4: UvA chatbot potential impact on students: Make studying easier

Created: 27/10/2024 by Mai Bui (2), **Modified:** 30/10/2024 by Maria Laitinen

Comment:
Student believes UvA chatbot will make studying easier.

● Z4: UvA chatbot potential impact on students: Uncertain about impact

Created: 27/10/2024 by Mai Bui (2), **Modified:** 30/10/2024 by Maria Laitinen

Comment:
Student is uncertain about UvA chatbot's potential impact.

● Z4: UvA chatbot potential impact on students: UvA chatbot might have more right answers

Created: 27/10/2024 by Mai Bui (2), **Modified:** 30/10/2024 by Maria Laitinen

Comment:
Student believes UvA chatbot might be more accurate than other existing LLMs.

● Z4: UvA chatbot potential impact on students: UvA chatbot would give more fact-checked answers

Created: 28/10/2024 by Phoebe Poort, **Modified:** 28/10/2024 by Phoebe Poort

Comment:
Students think the UvA chatbot would impact them, because the information given would be more secure then given by other LLMs.

● Z4: UvA chatbot potential impact on students: UvA chatbot would impact student's studies

Created: 25/10/2024 by Phoebe Poort, **Modified:** 30/10/2024 by Maria Laitinen

Comment:
If student thinks UvA chatbot would impact their studies.

● Z4: UvA chatbot potential impact on students: Would have no impact on education

Created: 25/10/2024 by Phoebe Poort, **Modified:** 30/10/2024 by Maria Laitinen

Comment:
If student thinks UvA chatbot would have no impact on their studies.

• Z5: How students would use UvA chatbot

Created: 25/10/2024 by Phoebe Poort, **Modified:** 30/10/2024 by Maria Laitinen

Comment:
This code category is for how the students would use the UvA chatbot.

• Z5: How students would use UvA chatbot: Make summaries specific to course

Created: 27/10/2024 by Mai Bui (2), **Modified:** 30/10/2024 by Maria Laitinen

Comment:
Student would use the UvA chatbot to make summaries that are specific to their course.

• Z5: How students would use UvA chatbot: Understanding terminologies in class

Created: 27/10/2024 by Mai Bui (2), **Modified:** 30/10/2024 by Maria Laitinen

Comment:
Student would use the UvA chatbot to understand terminology used in class.

• Z5: How students would use UvA chatbot: Use to answer questions

Created: 27/10/2024 by Mai Bui (2), **Modified:** 30/10/2024 by Maria Laitinen

Comment:
Student would use the UvA chatbot to answer questions.

• Z5: How students would use UvA chatbot: Would use UvA chatbot for writing support

Created: 28/10/2024 by Phoebe Poort, **Modified:** 30/10/2024 by Maria Laitinen

Comment:
Student would use the UvA chatbot for writing support.

• Z5: How students would use UvA chatbot: Would use UvA chatbot to find literature sources

Created: 28/10/2024 by Phoebe Poort, **Modified:** 30/10/2024 by Maria Laitinen

Comment:
Student would use the UvA chatbot to find literature sources.

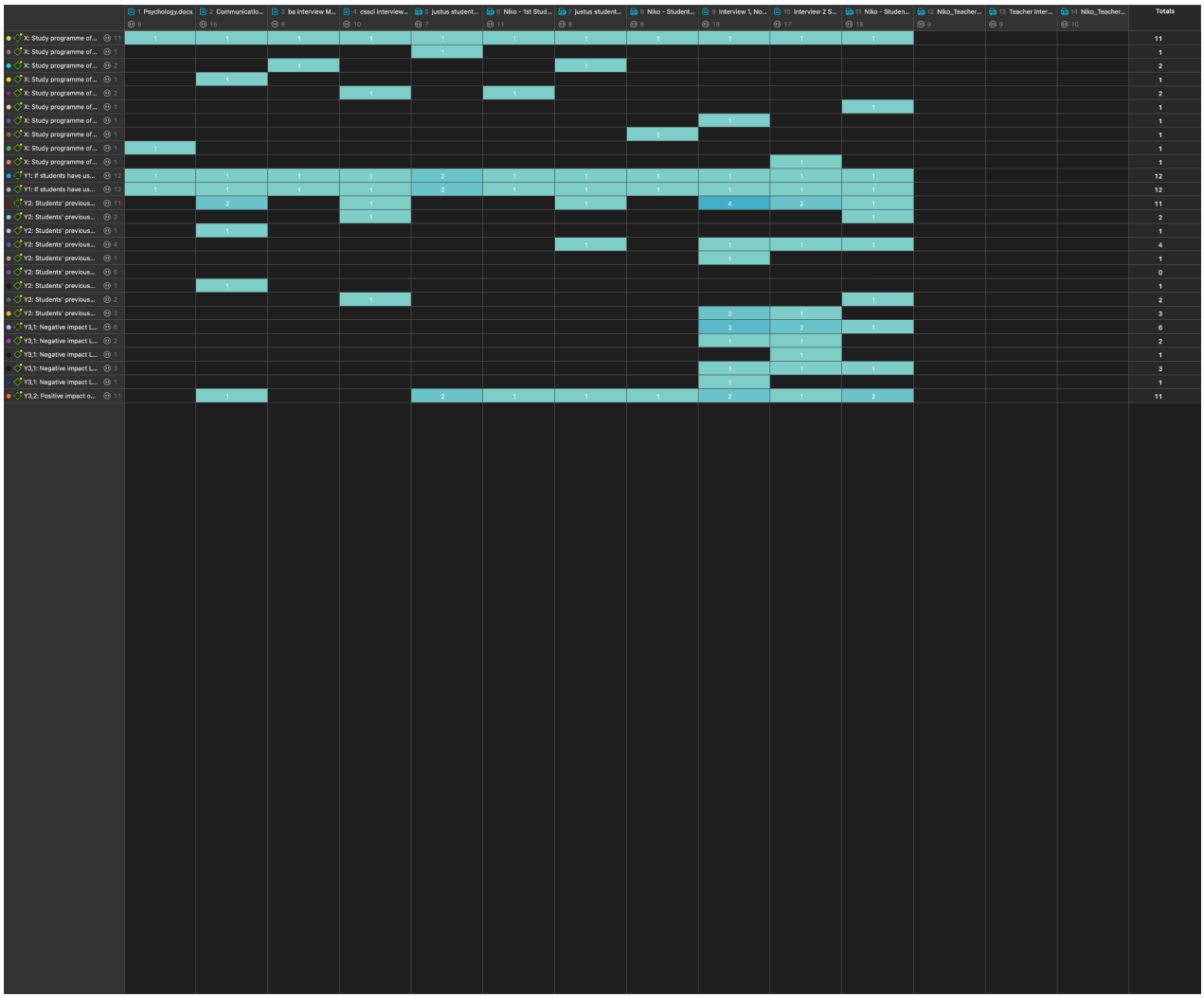
- **Z: Expectations and potential needs of UvA chatbot**

  Created: 23/10/2024 by Phoebe Poort, **Modified:** 28/10/2024 by Phoebe Poort

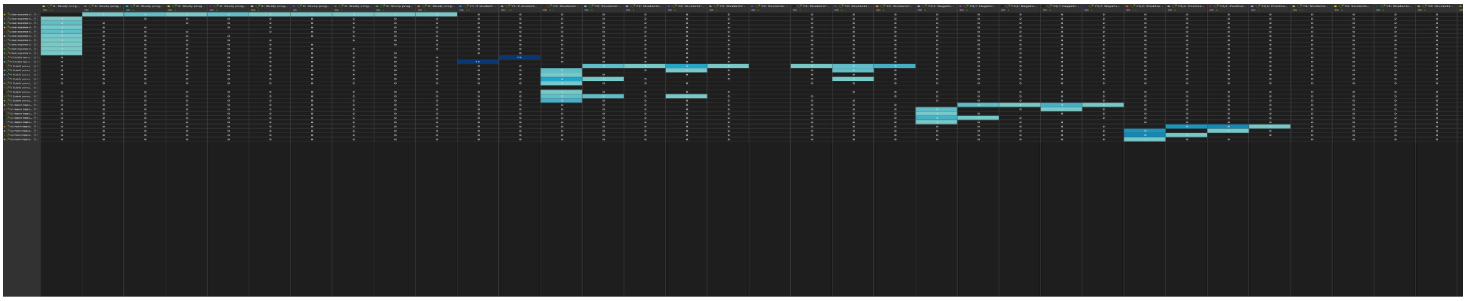
  Comment:
  This is the code category for the second main theme of the interview: expectations and potential needs of UvA chatbot.

ATLAS.ti Report: Student Interviews

Code-Document Analysis:



Code Co-Occurrence Analysis:

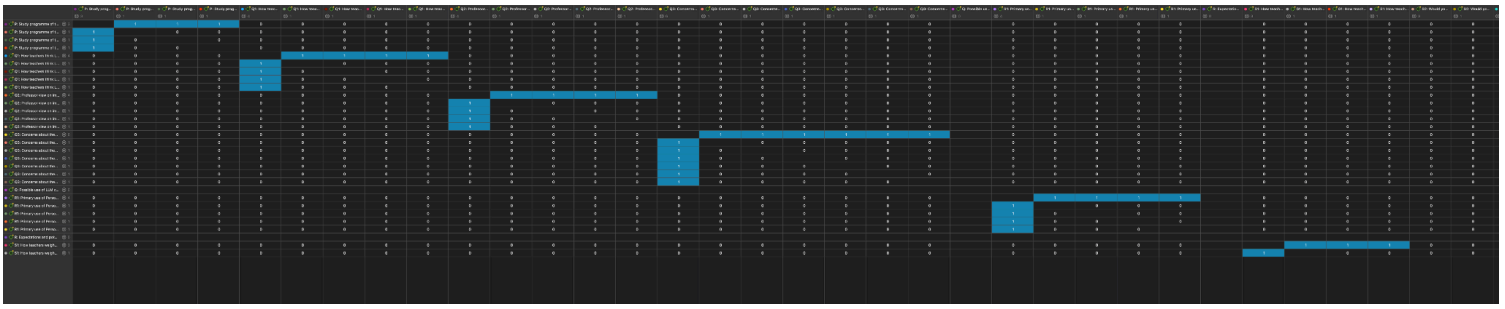


ATLAS.ti Report: Teacher Interviews

Code-Document Analysis:

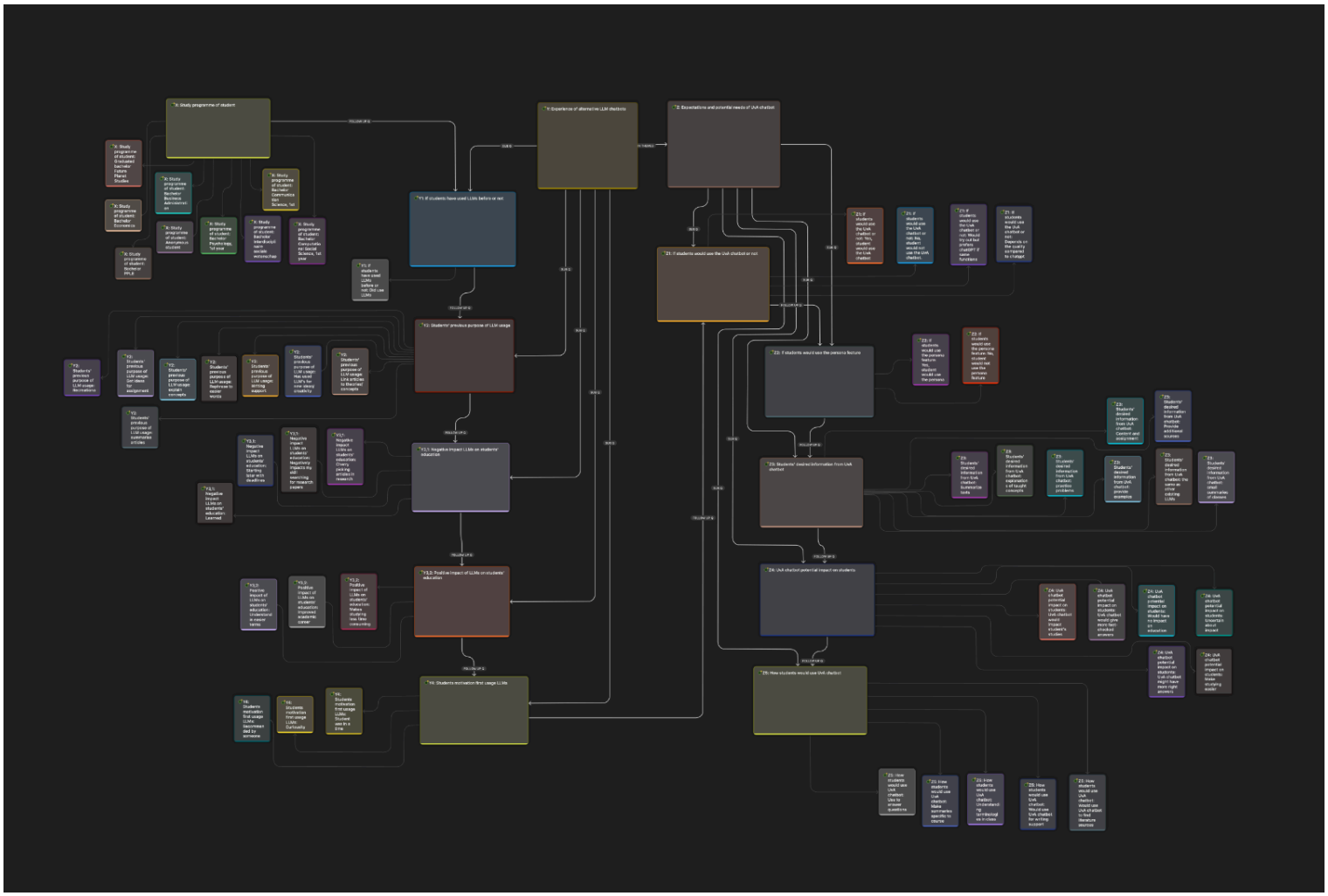
	12 Niko_Teacher... 9	13 Teacher Inter... 9	14 Niko_Teacher... 10	Totals
Q2: Professor view on im... 4	2	1	1	4
Q2: Professor view on im... 1			1	1
Q2: Professor view on im... 1	1			1
Q2: Professor view on im... 1		1		1
Q2: Professor view on im... 1	1			1
Q3: Concerns about the... 6	2	1	3	6
Q3: Concerns about the... 1			1	1
Q3: Concerns about the... 1			1	1
Q3: Concerns about the... 1			1	1
Q3: Concerns about the... 1		1		1
Q3: Concerns about the... 1	1			1
Q3: Concerns about the... 1	1			1
Q: Possible use of LLM c... 0				0
R1: Primary use of Perso... 4	1	1	2	4
R1: Primary use of Perso... 1		1		1
R1: Primary use of Perso... 1	1			1
R1: Primary use of Perso... 1			1	1
R1: Primary use of Perso... 1			1	1
R: Expectations and pot... 0				0
S1: How teachers weigh... 3	1	1	1	3
S1: How teachers weigh... 1			1	1
S1: How teachers weigh... 1		1		1
S1: How teachers weigh... 1	1			1
S2: Would you be willing... 3	1	1	1	3
S2: Would you be willing... 1			1	1
S2: Would you be willing... 2	1	1		2
S: The trade off between... 0				0
Q: Possible use of L... 4 0	5	4	5	14
R: Expectations an... 2 0	1	1	2	4
S: The trade off bet... 3 0	2	2	2	6
Teacher interview 10 0	9	8	10	27
Totals	35	31	39	105

Code Co-Occurrence Analysis:



ATLAS.ti: Code Networks

Student Interviews:



Teacher Interviews: