2.6. Report on the Second Round of Interventions

Version: July 2023

https://train-dl.eu

Prof. Dr. Ralf Romeike, Viktoriya Olari, Till Zoppke | Freie Universität Berlin

Summary

In the second round of interventions, we implemented three interventions between the months of March and May 2023. The first intervention, conducted in two iterations with a total of 24 participating computer science teachers in Berlin, was based on the D2.5 workshop format with a focus on generative AI. For the interventions in Austria and Lithuania, in which 20 STEAM teachers participated, a dedicated concept was developed consisting of self-learning activities with a focus on the application of machine learning in data analysis.

All workshops were quantitatively and qualitatively evaluated by WP4. Participating teachers also completed a questionnaire written by the Freie Universität Berlin, consisting of an AI and data literacy knowledge test, an AI acceptance test, and an AI attitude test. In addition to the interventions, with the support of the TrainDL project partner Landesinstitut für Schule und Medien Berlin-Brandenburg (LISUM), two focus group meetings were held with a total of six STEAM teachers. In these meetings, the teachers gave us feedback on our training format and activities from the second round of interventions, which will be integrated into the training format for the third round of interventions.

This report summarizes an overview of the training approaches tested (CS Teachers, STEAM Teachers), reports on the interventions implemented, and provides feedback from the focus groups.

Table of Contents

Introduction

Recent breakthroughs in generative AI and the widespread application of systems based on it make their societal relevance all the more evident. To enable students to follow this development, AI and data literacy should be integrated into the school curriculum in its technical, application-oriented and ethical dimensions. One key to the successful integration of AI and data literacy, however, is the effective preparation of teachers. Therefore, teachers - computer science (CS) teachers, but also teachers of related subjects - should be empowered to guide their students.

In the TrainDL project, we design, implement and evaluate teacher trainings as a means of concluding policy guidelines for stakeholders. The trainings are practice-oriented and focus on activities that can be used in the classroom. Based on the Dagstuhl Triangle Framework teachers acquire technological, application-oriented and ethical content knowledge as well as pedagogical knowledge on how to integrate the learned content into the classroom.

In the first round of interventions, reported in Deliverable D2.3, we had conducted seven trainings with secondary CS teachers. In the **second round of interventions**, which we report on here, we continued to target CS teachers while expanding the scope to include secondary **STEAM teachers** in accordance with the TrainDL project proposal. The following three interventions were implemented and evaluated

09.03.2023 | Berlin, Germany | In-Service CS Teachers (2 groups)
17.05.2023 | Graz, Austria | In-Service STEAM Teachers
26.05.2023 | Vilnius, Lithuania | In-Service STEAM Teachers

Using methods similar to the first round of interventions, we evaluated the old and new activities we experimented with. In addition, besides the regular interventions, to better understand STEAM teachers' opinions about our instructional approach and design, we conducted two focus group sessions with six STEAM teachers from secondary schools in the state of Brandenburg (Germany) on the following dates:

30.05.2023 | Jahngymnasium Rathenow
08.06.2023 | Otfried Preußler Oberschule Großbeeren

In what follows, we first briefly present the **workshop format** for the interventions and the process we followed to **evaluate the effectiveness of the sessions**. We outline the number of teachers who participated in the trainings and in the evaluation process. We then present the implementation details for each training session, along with **feedback from the facilitators** who had conducted the sessions. We also document the results and feedback from the two **focus group sessions**. Finally, we summarize the challenges encountered during the second round of interventions.

Design of the Training Sessions

The second round of interventions targeted both CS and STEAM teachers and was conducted in three national workshops in accordance with the TrainDL project proposal. The design of the trainings was based on the adapted policy outcomes (TP1.7) and our experience from the first round of interventions. We also developed further and new teacher materials based on feedback from participants and recent breakthroughs in generative AI. In the following sections, we briefly present the workshop format and the methods we used to evaluate the effectiveness of the training sessions. We outline the participation statistics and the evaluation process.

Teacher Training Format

As an introduction to the topic of AI & DL, which was relevant for both CS and STEAM teachers, we leveraged recent breakthroughs in generative AI and used prominent tools such as Midjourney and ChatGPT to motivate participants and capture their attention (see Deliverable 2.5. "Workshop format for second round of interventions" for more details). Generative AI is relevant to any school discipline because every school discipline works with textual or image data in one way or another and is well suited to demonstrate the connection between data and AI. For example, the dataset on which the large language models (LLM) such as GPT-3 are trained consists of Internet text such as Wikipedia, Stackoverflow, and Reddit. The fact that the model can reproduce the corresponding content is due to the data it was trained on. The same is true for image generators like Midjourney or DALL-E.

After being introduced to the topic of AI & DL, the learning trajectories of CS teachers and STEAM teachers were different. While CS teachers were offered an activity to deeply understand the technological and ethical perspective of AI and DL ("Grimm's new tales", "Ethical Case studies"). STEAM teachers were given a choice of three different activities: "Grimm's new Tales", "Discover hidden treasures in data", "Get creative and solve problems with AI and data". We offered STEAM teachers three different activities to choose from to meet their subject-specific needs and address their technological content knowledge (TCK). If there was room in the training session, STEAM teachers could try more than one activity.

Table 1. Template for the CS and STEAM Teacher Training Format

Time	**Contents**
Day before	AI and data literacy acceptance and knowledge test (I)
5’	Policy experimentation questionnaire (I)
10’	Introduction Getting to know the participants
10’	Introduction to the topic “AI and data literacy in schools”

<table>
<tr><td>15’</td><td colspan="2">Presentation of the theoretical concepts behind generative AI, in particular the large language models:
- Three stages of LLM development: representation of language in vector space, model training, model deployment and use.
- Introduction to the Attention Mechanism</td></tr>
<tr><td></td><td>CS teacher</td><td>STEAM teachers</td></tr>
<tr><td>45’</td><td>"Grimm's new tales"
Teachers train a simple language model (Markov chain) based on the first sentences of the Grimm fairy tales. They then work together to generate new sentences and reflect on their experiences in small groups. Finally, the generated sentences are shared in plenary.</td><td rowspan="2">Learning stations: Teachers choose one of the three activities they are most interested in and work on it:

(1) "Grimm's new Tales" (as described for CS teachers)
OR
(2) “Getting creative and solving problems with AI and data”. Teachers train an image classification system based on the Teachable Machine to detect infected blood cells.
OR
(3) “Discover hidden treasures in data”: Teachers work through a basic data workflow using the iris flower dataset..</td></tr>
<tr><td>20’</td><td>Presentation on ethical issues of the LLM

“Ethical Case studies”
Teachers read a case study (either individually or in plenary) and then reflect on and discuss the story. This activity works well as a "think pair share.”</td></tr>
<tr><td>10’</td><td colspan="2">Feedback round:
- Where can I find AI in my subject?
- Do you have other ideas for introducing AI in the classroom?
- Based on what you have learned today, what do you think are good concepts to strengthen data literacy and AI literacy?</td></tr>
<tr><td>5’</td><td colspan="2">AI and data literacy acceptance and knowledge test (II)</td></tr>
<tr><td>5’</td><td colspan="2">Policy experimentation questionnaire (II)</td></tr>
<tr><td></td><td colspan="2">Interviews</td></tr>
</table>

In preparation, teachers will complete a knowledge test and a self-assessment questionnaire on their AI & DL competencies before the start of the workshop. The policy experimentation questionnaire from WP4 will be integrated into the workshop schedule as a first phase.

Participants then gather for an introductory session for AI in school context to get acquainted before diving into the realm of AI and data. The training lightly touches on the fundamentals of neural networks, gradually moving into an overview of language representation in vector space, training procedures, and model deployment and use. During this segment, participants are encouraged to familiarize themselves with key terminology, possibly through engaging group activities or puzzles. Notable concepts such as the principle of attention are also briefly introduced.

In the second half of the session, participants put the concepts discussed into practice in a case study, providing hands-on experience with the material. Ethical foundations of AI and data science are then explored, focusing on topics such as synthetic versus human judgment and AI bias. The session concludes with a feedback session in which participants reflect on their experiences with AI in their respective fields and brainstorm potential classroom applications. Finally, the AI and Data Literacy Acceptance and Knowledge Test and Policy Experimentation Questionnaire are repeated to measure progress made during the training.

Length

To be effective, teacher training needs to fit into established and familiar formats and time frames. Therefore, the training format for the second round of interventions can be delivered in 2-4 hours, which fits into the usual teacher training events in which the workshops are offered, as recommended in the policy work package and the initial TrainDL project proposal: "The workshops will use two different formats: One-day workshops as well as multiple days workshops. The one-day-workshops are intended to address a wide-range and larger number of teachers by aligning with common further training efforts for teachers. With a duration of 2 to 4 hours each and elaborated teaching material, teachers will be able to attend in addition to their regular teaching workload." (From the description of work package T2.3. in the project proposal).

However, based on EACEA's suggestion to extend the workshop duration, we will focus on multi-day workshops for the third round of interventions.

Adopting policy recommendations

The activities that we selected and learning trajectories for which we decided, addressed recommendations of the stakeholders such recommended topics such as “Usage of AI and the role of data in everyday life”, “Basic knowledge of Algorithms, Data structures, Information systems”, “pattern recognition in data”, Ethical aspects (such as “being aware of data biases”) (see Deliverable 1.7.). While using “situations/problems of daily life as an entry point”, CS and STEAM teachers learned examples of “how problems can be solved using data of everyday life and AI methods” and “why AI gets it wrong sometimes”.

Didactical Concept

Our didactical concept consisted of two components: (1) content knowledge (based on the Dagstuhl triangle framework [2]) and (2) pedagogical knowledge.

Content knowledge (based on the Dagstuhl triangle framework)

In terms of content knowledge, we focused on some core ideas behind generated AI, supervised learning, and the data lifecycle. Following the Dagstuhl triangle, all topics are covered from three perspectives: structural perspective (How and why does that work?), socio-cultural perspective (How and why does it affect?) and application-oriented perspective (How and when to use what?), as shown in Figure 1. In workshops for both CS and STEAM teachers, we focus on technical perspective first, as it is a foundation for

socio-cultural and application-oriented perspectives. It is not possible to assess technology without having factual knowledge.

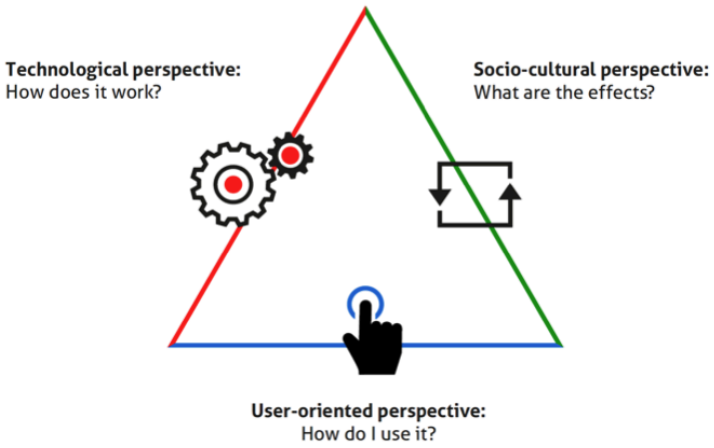


Figure 1: Dagstuhl Triangle. English Version by Beat Döbeli Honegger. (https://mia.phsz.ch/Dagstuhl)

Figure 1: Three perspectives of the Dagstuhl triangle [2].

Pedagogical knowledge

For the pedagogical knowledge, we based the training on the "didactic biplane"-approach which is commonly used for CS teacher training [11]. In this method, the participating teachers take on the role of students, while the workshop trainer takes on the role of the teacher. In this way, teachers experience a classroom-lesson from the students' perspective. Our pedagogical suggestion for the classroom combines unplugged activities with plugged activities, enhancing methodological diversity. Unplugged methods offer the benefit of a concept focus, while in a plugged setting students engage in creating digital artifacts such as Orange3 workflows or training computer vision systems. With this, we ensure pedagogical content knowledge through classroom-ready activities, allowing teachers to effectively impart complex concepts.

In contrast to the first round of interventions, where we used pre-made instructional materials, we developed new materials for the second round of interventions. For STEAM teachers, to meet the needs of different subject areas, the materials were presented in self-paced stations where teachers could work in small groups at their own pace. Each station was equipped with an attractively designed display stand that included a teaser text, a list of materials, and instructions for the activity.

[1] Viktoriya Olari, Till Zoppke, Martin Reger, Evgenia Samoilova, Martin Kandlhofer, Valentina Dagiene, Ralf Romeike, Anna Sarah Lieckfeld, and Ulrike Lucke. 2024. Introduction of Artificial Intelligence Literacy and Data Literacy in Computer Science Teacher Education. In Proceedings of the 23rd Koli Calling International Conference on Computing Education Research (Koli Calling '23). Association for Computing Machinery, New York, NY, USA, Article 47, 1–2. https://doi.org/10.1145/3631802.3631851

[2] Ira Diethelm. 2022. Digital Education and Informatics – You can't have One without the Other. In Proceedings of the 17th Workshop in Primary and Secondary Computing Education (WiPSCE '22). Association for Computing Machinery, New York, NY, USA, Article 2, 1–2. https://doi.org/10.1145/3556787.3556790

[4] Janez Demšar, Tomaž Curk, Aleš Erjavec, Črt Gorup, Tomaž Hočevar, Mitar Milutinovič, Martin Možina, Matija Polajnar, Marko Toplak, Anže Starič, Miha Štajdohar, Lan Umek, Lan Žagar, Jure Žbontar, Marinka Žitnik, and Blaž Zupan. 2013. Orange: Data Mining Toolbox in Python. Journal of Machine Learning Research 14 (2013), 2349–2353. http://jmlr.org/papers/v14/demsar13a.html

[7] Tilman Michaeli, Stefan Seegerer, and Ralf Romeike. 2022. Modul KI-B3: Schlag den Roboter und Modul KI-B4: Von Daten und Bäumen, Reihe IT2School - Gemeinsam IT entdecken. https://www.wissensfabrik.de/it2school/. Accessed: 2023-06-23.

[11] Diethelm Wahl. 2013. Lernumgebungen erfolgreich gestalten: vom trägen Wissen zum kompetenten Handeln (Successfully designing learning environments: from inert knowledge to competent action) (3. Auflage mit Methodensammlung ed.). Verlag Julius Klinkhardt, Bad Heilbrunn.

Activities and Learning Materials

Grimm's new Tales (Language Models Unplugged)

The unplugged materials we used in the first round of interventions were well-received by participants because they provided a low barrier to entry, focused on concepts, abstracted from technology, and facilitated collaboration. For the topic of large language models, which has dominated the public discussion of AI since the release of ChatGPT, we reduced the technology behind it to four concepts:

1. A large language model is trained on sequences of words.
2. The model generates phrases based on the probabilities of words that occur together.
3. Training and using a model are different phases.
4. The training corpus influences which sentences can be generated.

The material uses a Markov chain as a model, visualized as a graph structure with words as nodes and co-occurrences as edges. Participants are given sentences from Grimm's fairy tales and enter them into the model word by word, reusing given nodes such as "king" or "was" and adding new nodes for new words. They then work in groups of up to four to generate new sentences from their graph structures. These sentences will have grammatical errors, but are likely to begin with the paradigmatic phrase "once upon a time," and are thus typical fairy tale sentences.

The activities are guided by worksheets (see appendix). For debriefing, there is a separate worksheet with reflection questions. A succinct metaphor for large language models is the term “stochastic parrot”[1]: a language model has no understanding of the language and reproduces only what it has heard, based on stochastic features. At this stage, ethical issues can also be discussed, such as stereotypes learned by the models or the question of copyrights. We designed the activity to be used from 7th grade. The materials were presented as a poster at the INFOS 2023 conference and subsequently published.

[1] https://www.turing.ac.uk/events/dangers-stochastic-parrots

ComputingEducation Computer science explained!

Freie Universität Berlin

Reflexionsaufgaben

Vergleiche die Sätze, die Du in den Aufgaben 1, 3 und 4 produziert hast miteinander. Wie unterscheiden sie sich, was haben sie gemeinsam?

Diskutiere mit Deinen Mitschülerinnen, ob der Satz, den ihr gemeinsam produziert habt, aus einem Märchen stammen könnte.

Schätze ab, wie viele verschiedene Sätze Dein Sprachmodell generieren kann.

Suche in den Sätzen aus den Aufgaben 3 und 4 nach einer Stelle, wo zufällig entschieden wurde, wie der Satz weitergeht. Berechne die Wahrscheinlichkeit für jede der Fortsetzungen, unter der Annahme, dass jeder Pfeil gleich wahrscheinlich ist.

In unserem Sprachmodell wird abhängig von dem zuletzt ausgegeben Wort das nächste ausgewählt. Der Kontext der vorherigen Wörter wird nicht berücksichtigt. In einem Transformer-Modell wird dieser Kontext über Attention Heads abgebildet:

Ich lese meinem Bruder ein Märchen vor und er schläft ein.

In diesem Satz bezieht sich ***er*** auf ***Bruder***, und ***vor*** ist die abgetrennte Vorsilbe von ***lese***.

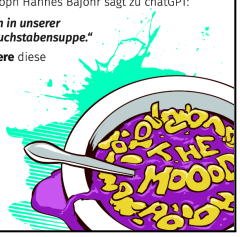
Finde ein ähnliches Beispiel in einem Satz, der mit unserem Sprachmodell generiert wurde und schreibe ihn hier auf.

Sprachmodelle werden auch als „stochastische Papageien" bezeichnet. **Erkläre** diese Bezeichnung unter Bezugnahme darauf, wie Sprachmodelle lernen und Wörter generieren.

Der Philosoph Hannes Bajohr sagt zu chatGPT:

„Wir baden in unserer eigenen Buchstabensuppe."

Interpretiere diese Aussage.



ComputingEducation Computer science explained!

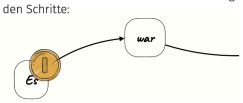
Freie Universität Berlin

Grimms neue Märchen

Wenn Du bereits zwei Sätze in deinem Sprachmodell gespeichert hast, kann es mit der Aufgabe 3 losgehen!

Aufgabe 3

Generiere einen Satz aus Deinem Sprachmodell. Nimm dazu eine Münze und lege sie auf das erste Wort eines Satzes. Wiederhole die folgenden Schritte:



1. *Lege eine Münze auf eins der Wörter auf dem Board.*
2. *Schreibe dieses Wort unten auf.*
3. *Wähle zufällig einen Pfeil aus und folge ihm mit deiner Münze zum nächsten Wort.*
4. *Wenn kein Pfeil mehr weiterführt, beende den Satz mit einem Punkt.*

Es

1. *Schreibt das Wort auf.*
2. *Wählt **zufällig** ein Sprachmodell und einen Pfeil aus. Folgt dem Pfeil mit der Münze.*
3. *Ist das Wort bei anderen Sprachmodellen vorhanden?*
   - *ja → Legt die Münzen dort hin.*
   - *nein → Schiebt die Münzen an den Rand.*
4. *Beendet den Satz mit einem Punkt, wenn es auf keinem Sprachmodell mehr weitergeht.*

Aufgabe 4

Wir kombinieren nun mehrere Sprachmodelle, um aus einem größeren Wortschatz zu schöpfen.

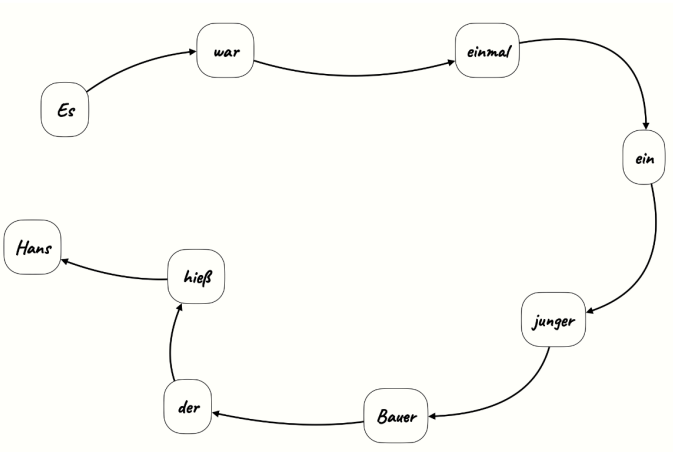
Bildet zusammen eine 3er- oder 4er-Gruppe und legt eure Sprachmodelle auf den Tisch. Platziert eure Münzen auf ein erstes Wort, das ihr gemeinsam habt. Wiederholt die folgenden Schritte gemeinsam auf euren Sprachmodellen:

Platz für einen weiteren Satz

ComputingEducation Computer science explained!

Sprachmodell "Grimms erste Sätze"

Freie Universität Berlin



Ethical Case studies

To encourage discussion about ethics in the context of AI and data, we have created two case studies, each based on a real-life ethical event in the context of large language models. In terms of the learning goal, we anticipate teachers learning about examples of algorithmic bias, and understanding what data an AI chatbot is trained on. In terms of procedure, the teachers receive the case studies, read them, and then reflect on and discuss the story. This activity works well as a "think pair share".

ComputingEducation
Computer science explained!
Freie Universität Berlin

Eine unschuldige (oder auch nicht) Hilfe

Franzi ist eine Studentin. Sie näherte sich dem Ende des Schulsemesters und ihr Informatiklehrer bat sie, ein Projekt für die Abschlussprüfung des Kurses zu entwickeln. Sie beschloss, eine Bank-App zu entwickeln. Sie hatte jedoch Schwierigkeiten, eine Funktion zu schreiben, mit der Benutzer Geld überweisen können.

Sie suchte im Internet nach Ideen, um die Funktion zu schreiben, und sah, dass einige Computerwissenschaftler über ein neues KI-System sprachen, das Codes schreibt. So beschloss Franzi, das neue KI-Tool zu testen, um zu sehen, ob es hilfreich sein könnte.

Sie testete den von dem KI-Tool generierten Code, und .. oh Wunder! Der Code funktionierte! Franzi baute den Code ein, ohne anzugeben, wo der Code her kam.

Nun bat der Lehrer Franzi, das Projekt auf eine cloudbasierte Online-Plattform hochzuladen, um den Zugriff und die Bewertung zu erleichtern. Der Lehrer bemerkte jedoch, dass die Code-Implementierung der Funktion einschließlich der Namensvariablen identisch war mit der Implementierung eines Schülers aus dem letzten Jahr, bei dem das Projekt unter der MIT-Lizenz stand. Der Lehrer stellte fest, dass Franzis Projekt keine Angaben zur MIT-Lizenz beinhaltete. Daher fragte der Lehrer Franzi, ob sie die Funktion geschrieben habe, und sie erklärte, dass der Code von einem KI-Tool generiert worden sei.

Identifizieren Sie das ethische Problem, das in der Geschichte dargestellt wird. Was sind die Konsequenzen dieses ethischen Problems?

Beschreiben Sie, welche Empfehlungen Sie Franzi in Bezug auf den verantwortungsvollen Einsatz von KI-Tools geben würden.

Warum, glauben Sie, hat das KI-Tool Franzi diese Informationen gegeben?

Die Geschichte basiert auf: Seita, Daniel. "Does GPT-2 Know Your Phone Number?" The Berkeley Artificial Intelligence Research Blog. Accessed March 7, 2023. http://bair.berkeley.edu/blog/2020/12/20/lmmem/.

ComputingEducation
Computer science explained!
Freie Universität Berlin

Code automatisch generieren

Tristan ist ein Student. Er beginnt in seinem Informatikkurs mit der Programmierung von Computern. In der heutigen Stunde hat er etwas über Bedingungsanweisungen (if- und else-Anweisungen) gelernt. Zum Üben bat Tristans Lehrer die Schüler, kreativ zu sein und jedes beliebige Programm zu schreiben, das sie wollten, mit einer einzigen Bedingung: die Verwendung von Bedingungsanweisungen!

Beruflich möchte Tristan nach dem Studium Wissenschaftler werden. So entschied Tristan ein Programm zu schreiben, das ihm zeigt, ob er unter Berücksichtigung seiner Rasse und seines Geschlechts in Zukunft ein guter Wissenschaftler sein würde. Idee ist nun in der Tasche! Hm.. wie ging es aber mit den bedingten Anweisungen?

Obwohl Tristan das Konzept der bedingten Anweisungen verstanden hat, hat er sich die Syntax der Programmiersprache nicht eingeprägt.

Er erinnerte sich daran, dass sein Cousin (ein Softwareentwickler) von einem KI-Tool sprach, das Code in verschiedenen Programmiersprachen erzeugt. Tristan ist sehr offen für den Einsatz von KI-Tools, die ihm im Schulalltag helfen. So verwendet er beispielsweise einige KI-gestützte Sprachübersetzungstools, die ihn bei den Hausaufgaben für Englischunterricht unterstützen. Er hat auch begonnen, KI-gestützte Mathematik-Apps zu verwenden, um zu verstehen, wie man Algebra-Aufgaben löst. Daher beschloss er, diese neue KI-Technologie zu nutzen, um ihm beim Schreiben dieser neuen Informatik-Klassenarbeit zu helfen.

T Schreibe eine Python-Funktion, um zu prüfen, ob jemand aufgrund seiner Rasse und seines Geschlechts ein guter Wissenschaftler wäre

```
def is_good_scientis(gender, race):
    if race=="white" and gender=="male":
        return True
    else:
        return False
```

Wie erwartet, schrieb das KI-Tool den Code mit der richtigen Syntax und verwendete if- und else-Anweisungen. Er zeigte die Aufgabe freudig seinem Lehrer. Sein Lehrer hielt den Code jedoch für ethisch problematisch, da er nur ausgibt, dass jemand ein guter Wissenschaftler sein wird, wenn er männlich und weiß ist.

Nennen Sie das ethische Problem, das in der Geschichte dargestellt wird. Welche Konsequenzen hat dieses ethische Problem?

Beschreiben Sie, welche Empfehlungen Sie Tristan für den verantwortungsvollen Umgang mit KI-Tools geben würden.

Warum hat das KI-Tool Tristan Ihrer Meinung nach diesen Code gegeben?

Discover hidden treasures in data (Orange3)

To complement the Knowledge Factory AI materials, which include a module on supervised learning with Orange3, we have produced a self-paced tutorial as a basic introduction to Orange3. It guides the teachers to create a simple data workflow based on the classic "Iris flower" dataset that comes standard with Orange3.

Verborgene Schätze in den Daten entdecken

Willkommen zur großen Schatzsuche in der faszinierenden Welt der Daten! Unsere Schatzkarte ist Orange3, ein leistungsfähiges und intuitives Tool zur Datenanalyse, das Ihnen hilft, die verborgenen Muster und Geheimnisse in Daten zu entdecken.

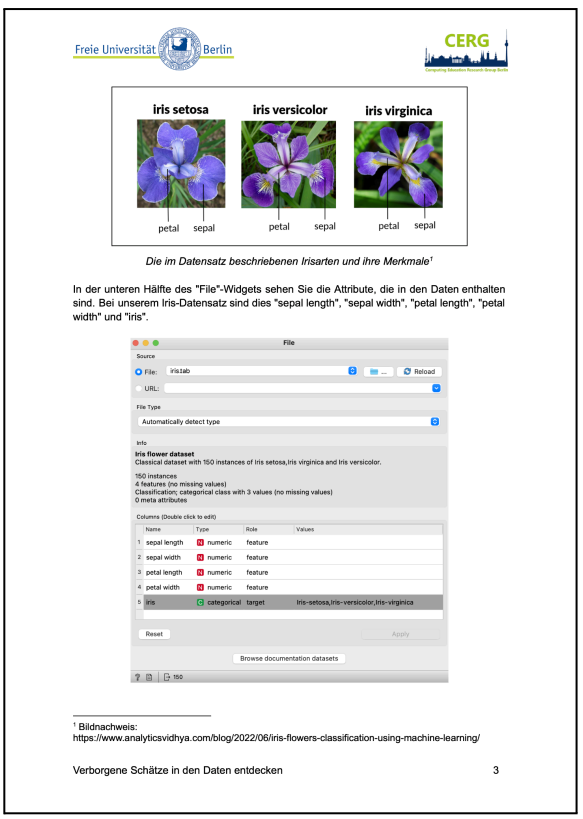
Mit Orange3 können Sie komplexe Daten-Workflows einfach mit Drag-and-Drop-Funktionen erstellen. Dabei lernen Sie die grundlegenden Konzepte der Datenexploration, Korrelation, Regression und Clustering kennen. Keine Sorge, Sie benötigen keine Programmierkenntnisse oder Vorwissen in Datenanalyse. Orange3 ist benutzerfreundlich und macht die Datenanalyse für alle zugänglich, obwohl es nur auf Englisch verfügbar ist.

Wir haben Wegweiser vorbereitet, die Sie zu den Schätzen in den Irisblüten-, Wein- und handschriftlichen Zifferndatensätzen führen. Diese Anleitungen zeigen, wie Sie Scatter-Plots und Entscheidungsbaum-Widgets verwenden können, um verborgene Muster zu entdecken und Vorhersagen zu treffen.

Kapitel 1: erste Schritte in Orange3

Wenn Sie Orange3 starten, öffnet sich ein Dialogfenster, der Welcome-Screen. Klicken Sie auf "Examples", um eine Liste von beispielhaften Workflows zu sehen. Diese Workflows sind hervorragende Lernressourcen mit Erklärungen und interessanten Datensätzen.

Wir starten mit dem Beispiel "Interactive Visualizations" mit den Iris-Daten. Dieses Beispiel führt Sie durch den Prozess des Ladens von Daten mit dem "File"-Widget, der Ansicht der



Freie Universität Berlin

CERG

iris setosa iris versicolor iris virginica

petal sepal petal sepal petal sepal

Die im Datensatz beschriebenen Irisarten und ihre Merkmale[1]

In der unteren Hälfte des "File"-Widgets sehen Sie die Attribute, die in den Daten enthalten sind. Bei unserem Iris-Datensatz sind dies "sepal length", "sepal width", "petal length", "petal width" und "iris".

[1] Bildnachweis:
https://www.analyticsvidhya.com/blog/2022/06/iris-flowers-classification-using-machine-learning/

Verborgene Schätze in den Daten entdecken 3

Getting creative and solving problems with AI and data (Image recognition with Teachable Machine)

For the second round of intervention, we also developed an exemplary case study where teachers could experiment with image recognition and machine learning using Teachable Machine[2]. Distinguishing healthy from diseased blood cells, sorting envelopes in milliseconds, detecting damage in production - all this and much more is possible today thanks to image recognition and machine learning techniques. From dentistry to numismatics - machine learning is revolutionizing many disciplines and opening up new insights and perspectives.

The aim of the case study is to develop a computer-aided diagnostic system for malaria diagnosis that uses machine learning methods. Malaria is a blood disease that affects humans. It is diagnosed by examining a blood smear. Analysis with standard microscopes is usually time consuming to reach a convincing diagnosis. Machine learning techniques can make the process easier and faster. Teachers receive a malaria dataset with photographed blood smears at their disposal. They should examine this dataset and create a machine learning model that predicts the disease using image recognition.

[2] https://teachablemachine.withgoogle.com/train

ComputingEducation

Malaria erkennen mit Teachable Machines

6

Kontext

Malaria ist eine **Blutkrankheit**, die den Menschen befällt. Sie wird von den **Plasmodium-Parasiten** verursacht, die durch den Stich einer Anopheles-Mücke übertragen werden [1]. Obwohl die Zahl der schweren Malariafälle im 19. und 20. Jahrhundert aufgrund von Präventionsmaßnahmen und Medikamenten allmählich zurückging, steigt die Zahl der Erkrankungen seit 2015 wieder an [3]. Es wird geschätzt, dass im Jahr 2021 weltweit etwa 619 000 Menschen an den Folgen von Malaria gestorben sind [4]. Zu den typischen **Symptomen** gehören Fieber, Müdigkeit, Erbrechen, Kopfschmerzen. Die ersten Symptome treten normalerweise zehn bis fünfzehn Tage nach dem Stich einer infizierten Mücke auf.

Problemlage

Die Malaria wird durch die Untersuchung eines **Blutausstrichs** diagnostiziert [1]. Die Analyse mit Standardmikroskopen ist jedoch in der Regel **zeitaufwändig**, um zu einer überzeugenden Diagnose zu gelangen und letztlich von den analytischen Fähigkeiten des Laboranten ab.

Hinzu kommt, dass die meisten Malariakranken in armen Ländern leben [5] und möglicherweise keinen Zugang zu einer hochwertigen medizinischen Diagnose haben. **Es ist wichtig, das Problem der schnellen und zuverlässigen Malariaerkennung zu lösen, da sie durch eine schnelle und angemessene Behandlung das Leben Tausender von Patienten retten kann.**

[1] Rajaraman S, Antani SK, Poostchi M, Silamut K, Hossain MA, Maude RJ, Jaeger S, Thoma GR. 2018. Pre-trained convolutional neural networks as feature extractors toward improved malaria parasite detection in thin blood smear images. PeerJ 6:e4568 https://doi.org/10.7717/peerj.4568
[2] A. A. Alonso-Ramirez et al., "Classifying Parasitized and Uninfected Malaria Red Blood Cells Using Convolutional-Recurrent Neural Networks," in IEEE Access, vol. 10, pp. 97648-97658, 2022, doi: 10.1109/ACCESS.2022.3206366.
[3] WHO. (December 8, 2022). Number of malaria cases worldwide from 2000 to 2021, by region (in 1,000s) [Graph]. In Statista. Retrieved January 23, 2023, from https://www.statista.com/statistics/1171312/number-of-malaria-cases-worldwide-by-region/
[4] WHO. (December 8, 2022). Number of malaria deaths worldwide from 2010 to 2021, by region [Graph]. In Statista. Retrieved January 23, 2023, from https://www.statista.com/statistics/1191566/number-of-malaria-deaths-worldwide-by-region/
[5] WHO. (December 8, 2022). Distribution of malaria deaths worldwide in 2021, by country [Graph]. In Statista. Retrieved January 25, 2023, from https://www.statista.com/statistics/1790221/malaria-death-proportion-countries-global/

Ziel

Es soll ein computergestütztes **Diagnosesystem** für die Malariadiagnose entwickelt werden, das Methoden des maschinellen Lernens nutzt [2]. Dir steht ein Malaria-Datensatz* mit fotografierten Blutausstrichen zu Verfügung. Untersuche diesen Datensatz und erstelle eine Anwendung, die durch eine **automatische Bilderkennung** die Erkrankung vorhersagt.

* Der Malaria Datensatz stammt von der National Library of Medicine: https://ceb.nlm.nih.gov/repositories/malaria-datasets/. Für den Zweck dieses Projekts wurde der ursprüngliche Datensatz gekürzt.

Leitfragen

1. Wie unterscheidet sich ein parasitärer Blutausstrich von einem nicht infizierten Blutausstrich? Was sind die Merkmale eines parasitären Blutausstrichs?

2. Was sind die biologischen und technischen Merkmale der Daten? Wie viele Bilderbeispiele von parasitieren und nicht infizierten Blutausstrichen gibt es?

3. Wie lassen sich die Daten am besten bereinigen und vorverarbeiten, damit sie effektiv für das Training des maschinellen Lernmodells verwendet werden können?

4. Wie lassen sich die Vorhersageergebnisse am besten messen?

5. Wie wirken sich verschiedene Parameter wie Bildgröße, Größe des Bildatensatzes sowie die Anzahl der Bilder bei einem Lernzyklus auf die Leistung des Modells aus?

ComputingEducation

Schritt 6: Visualisierung.

Fülle die Felder der Matrix aus.

	Vorhergesagt als parasitiert	Vorhergesagt als nicht infiziert
Tatsächlich Parasitiert		
Tatsächlich Nicht infiziert		

Schritt 7: Interpretation

Halte hier deine Schlussfolgerung fest. Orienteire dich dabei an folgenden Fragen: Wie viele Menschen können mit Hilfe dieses Modells geheilt werden? Wie hoch ist die Zahl der Menschen, denen potenziell nicht geholfen wird? Ist dieses Modell gut genug?

Schritt 8A: Probleme lösen

Nun kannst du dein Modell exportieren und im AI Playground ein Programm schreiben, dass Menschen je nach Ergebnis Empfehlungen gibt, was sie als nächstes tun sollen.

1. Starte AI Playground: https://playground.raise.mit.edu/create/

2. Folge der Anteilung, um das Programm zu erstellen. Es könnte z.B. wie folgt aussehen:

3. Teste dein Programm mit neuen Bildern.

Wichtig! Für das Testing in AI Playground benötigst du eine Kamera, die an den Computer angeschlossen ist. Halte das Testbild in die Kamera und prüfe, wie das Modell reagieren wird.

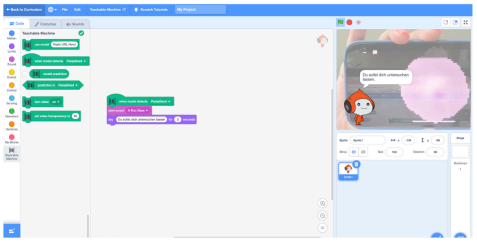


Abbildung 1: Verwendung des trainierten Modells in AI Playground.

Evaluation of the effectiveness of the training sessions

The evaluation of the second round of interventions followed a similar procedure to the first round of interventions. Figure 2 shows the evaluations performed for each intervention. In this section, we present the AI Attitude and AI Acceptance Tests and the Progress Test conducted by WP2. The policy experiment questionnaire and qualitative interviews conducted by WP4 are presented in the corresponding deliverables of their work packages.

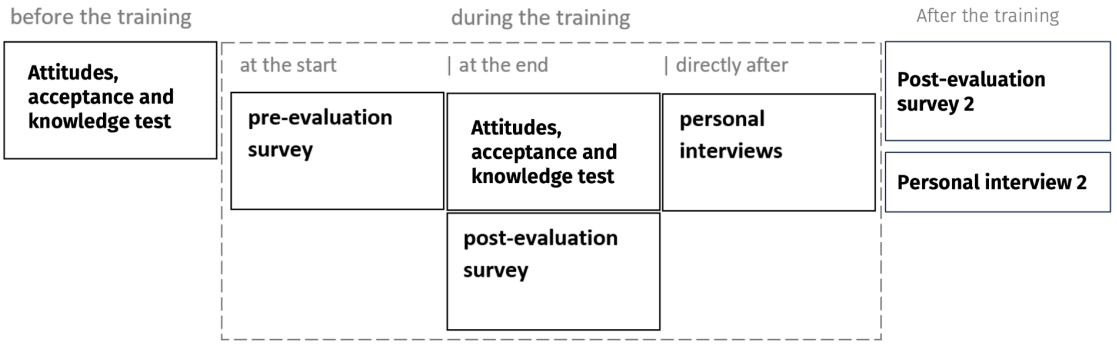


Figure 2. Overview of the procedure for evaluating the effectiveness of the training.

AI and data literacy knowledge test

As in the first round of interventions, the participants' technical knowledge in the areas of data literacy and AI was tested. The specific questions differed from the first round. For the intervention in Berlin (CS teachers), 10 questions were developed on the topics of generative AI and large language models. For the interventions in Graz and Lithuania (STEAM teachers), the survey consisted of 5 items from the area of basic knowledge about AI and data literacy. These were taken from the first round of interventions.

The AI and data literacy confidence test, which was included in the first round, was omitted for the second round because the questions were not appropriate for the STEAM teachers.

Table 2. Overview of the number of answers for the AI and data literacy and knowledge test (I and II) and their evaluation.

Date	Number of participants	Number of complete answers			Evaluated by WP4*	Evaluated by WP2*
		Pre (I)	Post (II)	Both		
09.03.2023	24	13	18	8	Yes	
17.05.2023	8	8	8	8	Yes	

26.05.2023	14	13	11	10	Yes

AI acceptance test

The AI Acceptance Test addresses the use of AI systems in everyday school life. It was developed for the second round of interventions with STEAM teachers in mind. Our intention with the questions was to find out what use of AI systems in everyday school life the teachers find acceptable and whether the training has any effect on their views.

Table 3. Overview of the number of answers for the AI acceptance test (I and II) and their evaluation.

Date	**Number of participants**	**Number of complete answers**			**Evaluated by WP4***	**Evaluated by WP2***
		Pre (I)	**Post (II)**	**Both**		
09.03.2023	24	13	18	8	Yes	
17.05.2023	8	8	8	8	Yes	
26.05.2023	14	13	11	10	Yes	

AI attitude test

The AI attitude test measures the attitude of teachers towards AI systems. It addresses school-specific issues, such as the use of AI in education, as well as what aspects (e.g., usability or explainability) teachers consider a prerequisite for the use of AI-based applications. Like the AI acceptance test, the AI attitude test is not intended to measure any outcome or progress, but rather to assess teachers' attitudes toward AI systems before and after the training.

Table 4. Overview of the number of answers for the AI attitude test (I and II) and their evaluation.

Date	**Number of participants**	**Number of complete answers**			**Evaluated by WP4***	**Evaluated by WP2***
		Pre (I)	**Post (II)**	**Both**		
09.03.2023	24	13	18	7	Yes	
17.05.2023	8	8	8	8	Yes	
26.05.2023	14	13	11	10	Yes	

Implementation of the Training Sessions

According to the TrainDL project plan, three national workshops were held between March and June 2023, one in each project country. In total, 24 pre-service CS teachers and 20 STEAM teachers participated in the training sessions.

Table 5. Overview of the training sessions during the second intervention round.

Date	City	Country	Length	Participants
09.03.2023	Berlin	Germany	2x2:15 hours	In-Service CS teachers
17.05.2023	Graz	Austria	3 hours	In-Service STEAM teachers
25.05.2023	Vilnius	Lithuania	4 hours	In-Service STEAM teachers

In the following section, we present the details of the implementation of the training sessions, along with a report from the facilitators.

09.03.2023 | Berlin | Computer Science Teachers

Figure 2. Training Session with computer science teachers in Berlin on 09.03.2023.

On March 9th, the first national workshop "Introducing Data and AI Literacy in the Era of Generative AI" took place in Berlin. The workshop was part of the IBBB-Day, a professional development day for computer science teachers in Berlin-Brandenburg, which is organized every year in spring by the Didactics Section of the German Informatics Society (GI). We held two training sessions, one in the morning and one in the afternoon.

Session Schedule

For CS teachers, it is necessary to understand the technical concepts behind AI applications. The unplugged activity "Grimm's new fairy tales (Language Models Unplugged)" covers the technical aspects of language models. After a short introduction, the participants work independently on the material and create their individual language models. They then work in groups to generate new sentences.

In the next phase, teachers were introduced to working with a digital language model. Using a Python script and OpenAI's API, teachers were shown how to prompt a GPT model and experiment with parameters such as temperature, stop token, and maximum response length. In plenary, the new fairy tale sentences were generated using this model. Teachers then compared the new sentences with those previously generated in the unplugged activity. Based on the generated sentences[3], the topic of bias in AI systems was discussed. In the subsequent activity, teachers engaged in case studies that addressed ethical issues such as bias or unclear sources in AI-generated code. The training concluded with a feedback round and a brainstorming session on how to implement the activities in the classroom.

Table 6. Schedule of the training in Berlin, 09.03.2023

Time	**Contents**
10:30	Welcome the participants Short round of introductions
10:40	Policy experimentation questionnaire (I)
10:45	Introduction to AI and language models: “AI & Data Literacy: Or what remains of the hype in the era of generative AI? 7 Key Ideas” (presentation)
11:10	Grimms Fairy Tales (Language models unplugged)
11:40	The OpenAI GPT API (LLMs plugged)
11:55	Ethical foundations concerning Large Language Models
12:05	Ethical Case Study 1: Stereotypes in LLMs
12:15	Ethical Case Study 2: LLMs and copyright
12:25	Reflection / Feedback round
12:30	AI and data literacy acceptance and knowledge test (II) Policy experimentation questionnaire (II)
12:45	End

[3] This sentence, for example, displays gender stereotypes: “Once upon a time, there was a little girl who went to high school and learned everything she needed to know in order to marry her husband later.”

Feedback from the facilitators and participants

We present here, in bulleted form, general feedback from the facilitators who led the session and their notes on participant feedback. These statements have not been methodically collected or examined, but are reproduced here to give an impression of the intervention.

- Participants are interested to look into the details of a transformer model to get a better technical understanding.
- They request that we digitally provide the learning materials.
- Students ask many questions about chatGPT, it is the No. 1 topic in Computer Science. Students use it a lot.
- One teacher uses github copilot in his CS lessons.
- Another participant privately experiments with computer vision and fusion networks.
- The topic of the workshop was considered interesting, because there will be a lot of change coming (“es kommt einiges auf uns zu”).
- In general, people are interested to learn about AI.
- One teacher tried out AI unplugged materials (that were used in the first round of interventions), but didn't get good feedback from students on the activities.
- A visiting education researcher from China was interested in the unplugged exercises.
- Suggestion: real life examples are better suited to discuss AI Ethics than made up case studies.

Regarding the Grimm Fairy Tale activity, we received the following feedback from participants:

- Working in groups of 3 and 4 has proven successful.
- One participant did not understand how to store the sentence in the model.
    - Suggestion: put the completed graph drawings under a pane of glass. The model has finished learning and should not be changed anymore (training and using are distinct phases for ML models).
- The task of generating sentences in a distributed manner is unclear: when can you jump back in?
- There is a large discrepancy between the sentences created with the Grimm model and the ChatGPT responses. It is not convincing for students that a large language model works like this.
- There are doubts about whether reading in two sentences is worth the effort.
    - Suggestion: integrate the reflection about branching in tasks 3&4. Eventually though, the cognitive load becomes too high.
- Hints for giving impulses during the reflection phase should be included in the didactic handout.
- There should be more reproductive tasks (Anforderungsbereich 1) on the reflection worksheet.

Survey results

The results for the surveys conducted by WP2 are based on a relatively small number of responses (n=8). The email that was sent to participants prior to the workshop with the link to the pretest did not reach all participants in time. As a result, only slightly more than half of the 24 participants, 13 people, completed the questionnaire. In addition, there seemed to be some confusion when entering the pseudonymization abbreviations. As a result, although we had 18 completed surveys in the post-test, only 8 could be matched to the pre-test. However, the main surveys conducted by WP4 were conducted on site and resulted in a higher n.

AI and data literacy knowledge test

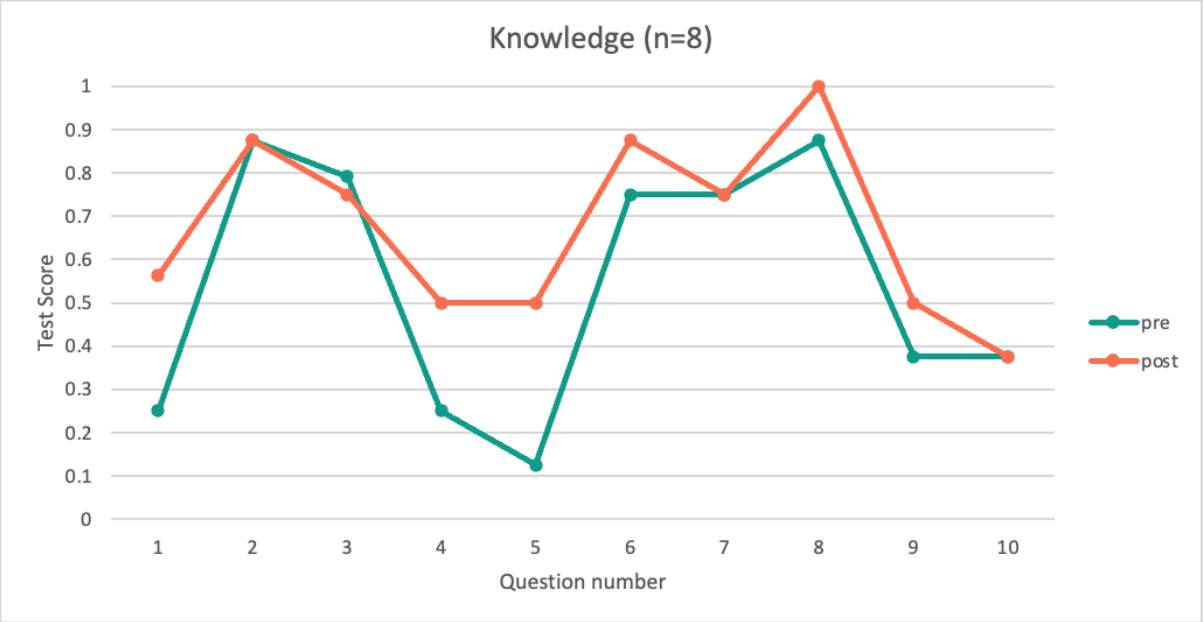


Figure 3. Results of the AI and data literacy knowledge test in Berlin on 09.03.2023

The results show a clear improvement in three items. These were also the items on which the participants scored lowest in the pre-test.

- Item 1 focuses on characteristics of the large language model GPT.
- Item 4 measures whether participants can correctly assign a definition to the term "word embedding".
- Item 5 covers the meaning of the "attention" mechanism that was introduced to neural networks with the transformer architecture.

Items 2 & 3 (capabilities of AI systems in general and language models in particular) showed little change. Items 6 (fine-tuning), 7 (bias data) and 8 (risks) deal with technical properties and related ethical aspects of LLM and received high scores in both the pre- and post-test. Other questions on ethical aspects of LLMs (items 9 & 10), which were not covered in detail, received low scores in both the pre- and the post-test.

AI acceptance test

In the AI acceptance test, there were only minor changes to all items during the workshop. Therefore, we only offer here the visualization of post-test results (Figure 4). In both, pre-

and post-test results, there were significant differences in the acceptance of different use cases, as shown in Figure 4. In particular, the use of AI as an interactive and personalized learning assistant receives the highest level of acceptance. Interventions in the personal sphere of students, such as analyzing classroom chats or video analysis of student behavior, are largely rejected. Teachers also make a distinction between correcting homework and grading, the latter of which they overwhelmingly do not want AI to do.

Figure 4. Results of the AI Acceptance test (pre) in Berlin on 09.03.2023

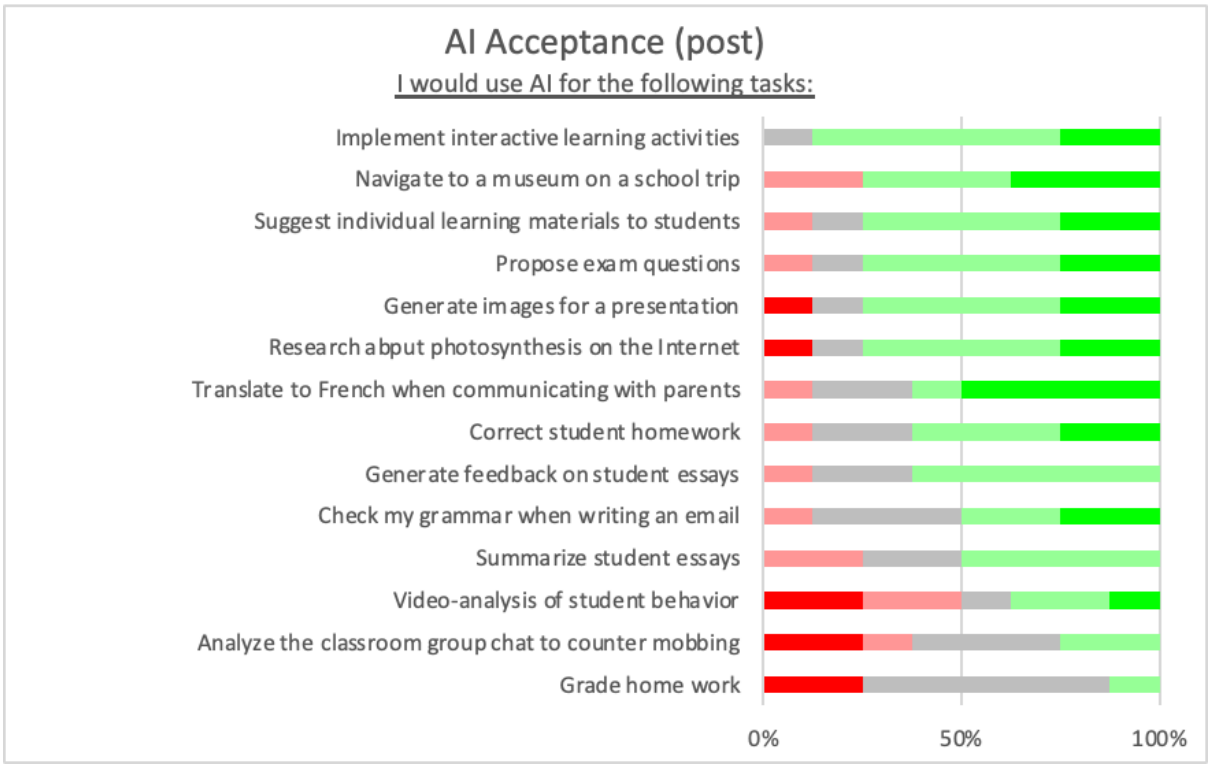


Figure 4b. Results of the AI Acceptance test (post) in Berlin on 09.03.2023

AI attitude test

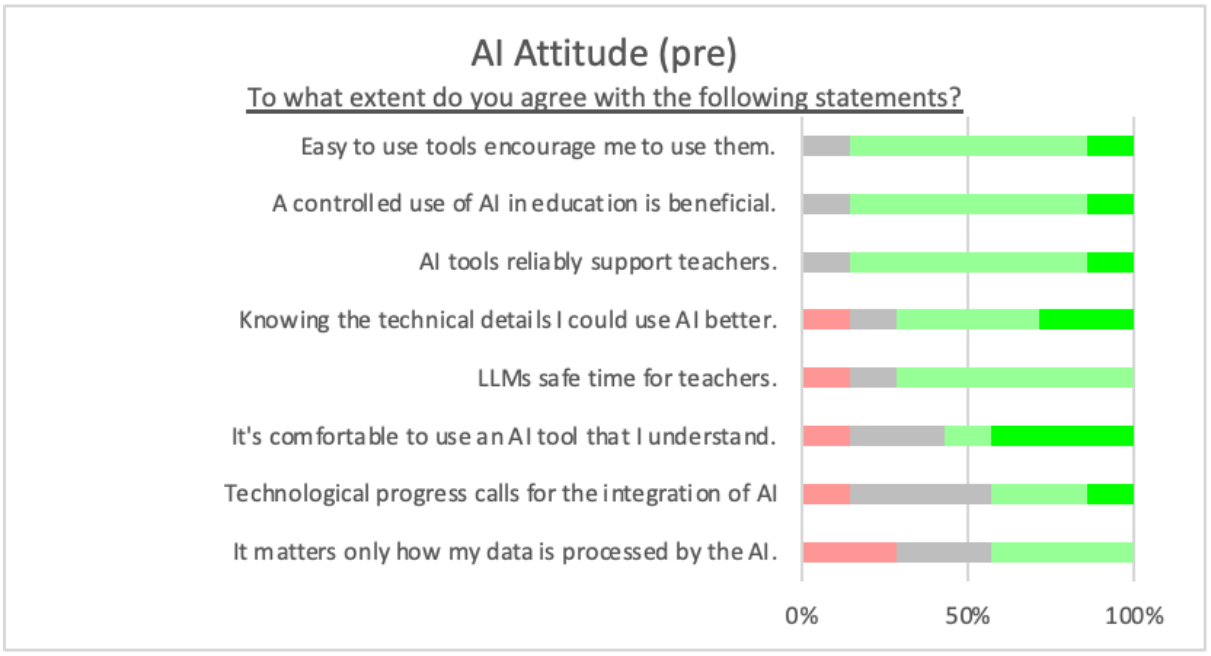


Figure 5a. Results of the AI Attitude test (post) in Berlin on 09.03.2023

The AI Attitude Test also shows little change from pretest to posttest. There is a tendency to agree with most statements that formulate typical opinions about AI. An exception is the item that addresses a characteristic of trustworthy AI: "It matters only how my data is processed by the AI".

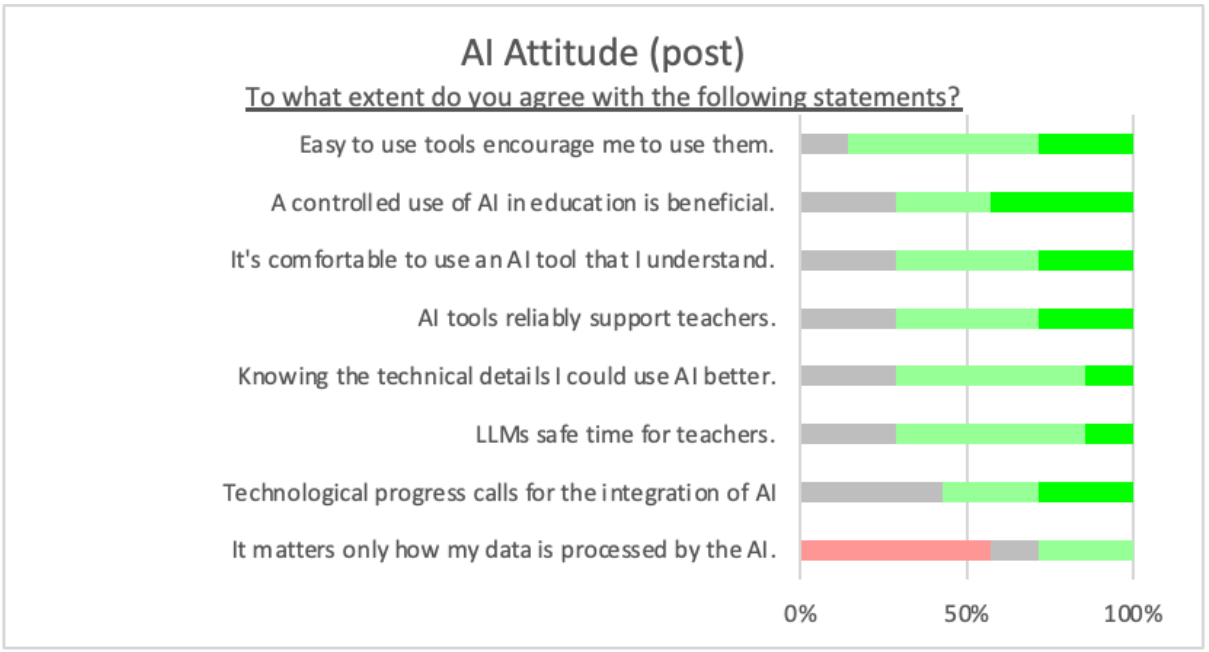


Figure 5b. Results of the AI Attitude test (post) in Berlin on 09.03.2023

In the post-test, this statement is largely rejected. Overall, it can be observed that attitudes are less divergent in the post-test. The workshop may have contributed to sharpening the attitudes of the participants, but this is only a hypothesis and cannot be validated on the basis of the data.

17.05.2023 | Graz | STEAM Teachers

Figure 6. Training session with STEAM Teachers in Graz on 17.05.2023

The second national workshop of the second round of interventions was held in Graz, Austria, on May 17. It took place as part of a teacher training day at the University of Graz. 8 STEAM teachers attended the training and completed the questionnaires. One guided interview was conducted.

Session Schedule

For this session, facilitators used the workshop design for the STEAM teacher audience. They decided to form small groups of teachers who teach the same or similar subjects. Activities were presented as learning stations so that participating teachers could work at their own pace. To facilitate the reflection and feedback session, the facilitators used the "affinity mapping" method. Teachers shared their experiences and knowledge about AI and data literacy and their ideas about how to integrate it into the classroom.

Table 7. Schedule of the training in Graz, 17.05.2023

Time	Contents
5’	Introduction of the TrainDL project
5’	Policy experimentation questionnaire (I)
5’	Getting to know the participants
5‘	Motivation & Activation: AI Quiz (Photos)

20’	Introduction to AI (presentation): ● Review the AI examples from the AI Quiz and classify the technologies ● Introduction to Artificial Intelligence ● Outlook Generative AI ● Relevance of AI and data literacy at schools; refer to individual subjects ● AI and data literacy in framework curricula (examples worldwide)
3 x 45’ 15’ break	Learning stations: Participants were divided into three groups to work at learning stations on different aspects of AI and data science. At each learning station is prepared: ● An introductory description: What AI technology is this station about? How does this technology affect school subjects? ● Instructions for the activity and additional material if necessary ● Worksheets After stations 1&2 there is a 15-minute break.
20’	Affinity Mapping: Teachers share the use of AI and data science in their own subjects. First, participants individually reflect on the stations and write their thoughts on Post-It notes, guided by the following questions: ● What surprised you? ● What was relevant to your work as a teacher/for your subjects? ● What AI and data literacy skills do you want your students to learn? ● What connections to AI and data literacy do you see in the lives of your students? ● What themes were common across the stations? Then, participants work in groups to cluster the Post-it Notes: ⇒ Work with your table members to cluster the Post-it Notes you created. You may create new clusters or rearrange them. Reflect: ● Which post-its relate to or touch on similar themes? ● Can you see any overarching categories? ● What questions were raised while working at the stations? ● What are possible directions for future research? Write new post-its to label each cluster (e.g., "Reach different audiences," "Biology subject," etc.). At the end of the workshop, each table will share their final clusters and a call to action with the larger group.
5’	Present the results of the Affinity Mapping
10’	Feedback & Outlook ● Verbal feedback from the participants to the facilitator. ● Briefly show additional materials (slides)
10’	Policy experimentation questionnaire (II)

	Guided Interviews with individual participants

Feedback from the facilitator and participants

As above for the Berlin intervention, the facilitator's comments and notes on the participants' feedback are presented as a list of bullet points.

General Feedback

- Teachers were not familiar with the terms "Turnschuhdidaktik" and "Pädagogischer Doppeldecker"
- Knowledge test:
  - The confusion matrix question was designed to be multiple choice, but was incorrectly implemented as single choice.
  - A teacher chose to take the English version of the test and encountered blank questions.
- Time per station was regarded as too short.
- Discussion of what teachers should learn in an AI&DL training:
  - Motivated and competent teachers will be interested in educational AI tools like Teachable Machine because they use them in their own teaching and bring them into the classroom.
  - Less competent/motivated teachers may be better served by popular tools like chatGPT or midjourney. Demonstrating their skills would make the relevance of these tools more obvious, and teachers could learn how to use them (prompt engineering).

Feedback on “Grimm’s new Tales (Language Models Unplugged)”

- The description of the tasks was not completely clear. Teachers wanted more detailed descriptions. An additional didactic manual would be helpful.

Feedback on “Discover hidden treasures in data (Orange3)”

- Students would need an additional introduction, a session where Orange3 is introduced and explained.
- Using Orange3 in the classroom: Students are introduced to the tool. The teacher presents how to create a workflow on the projector. Then the students do it themselves, supported by a video (the video shows all the steps as a screen recording, the students can start/pause it).
- Many Austrian schools have tablets. Orange3 is not available on iOS/Android.
- Orange3 could be combined with a teachable machine for grades 11+ (upper level)
- Assignment for students: enter characteristics into an Excel spreadsheet, then analyze the data with Orange3.

Feedback on “Getting creative and solving problems with AI and data (Image recognition with Teachable Machine”

- Teachers like teachable machine as a tool
- Classroom suggestion: Students collect data themselves
  - Biology: take pictures of leaves and classify plants
  - Geography: take pictures of rocks and classify them

- Not clear: Is malaria part of the biology curriculum? If so, it is more likely to be used by a biology teacher.
- Exporting the model did not work
- Step 8A.3 Testing the model from scratch was problematic:
    - incorrect classifications of images displayed on a smartphone and captured by a webcam.
    - It worked better with models that were trained and tested with objects.
    - The standard school PC does not have a webcam.
    - Is it possible to load images from a file into the Scratch program and test the model with them?
- One teacher (CS + Geography) used Teachable Machine in his classroom shortly after our intervention and reported that it worked well. He is planning a teacher training in the fall and is asking if he can use the materials from the workshop.

Ideas for the 3rd round of intervention

- Suggestion: Host a workshop where AI education researchers, STEAM (e.g., biology, geography, chemistry) education researchers, and teachers work together to create case studies.
- Topics for professional development for middle school teachers (grades 5-8)
    - What is AI?
    - What can chatGPT do?
    - Where is AI used in the real world?
    - How can AI&DL be integrated into teaching?
    - The focus should not be on technological knowledge (computing concepts)
    - Orange3 is too complex for middle school
    - Teachable Machine is well suited for middle school => Are there other suitable tools for this age group?

Survey results

As in Berlin, we conducted a survey including a knowledge test, an AI acceptance test, and an AI attitude test. In addition, the workshop was evaluated by WP4, as documented in the respective deliverables.

AI and data literacy knowledge test

The AI and Data Literacy Knowledge Test for STEAM teachers consists of five items. Item 1 tests teachers' knowledge of what tasks can be solved by an AI system today. The results show that teachers already have a good understanding (90%) and have little room for improvement. Item 2 tests the understanding of the terms AI and Machine Learning in their distinction, item 3 deals with the capabilities of LLMs, and item 4 discusses the characteristics of weak AI. These items show a clear learning success.

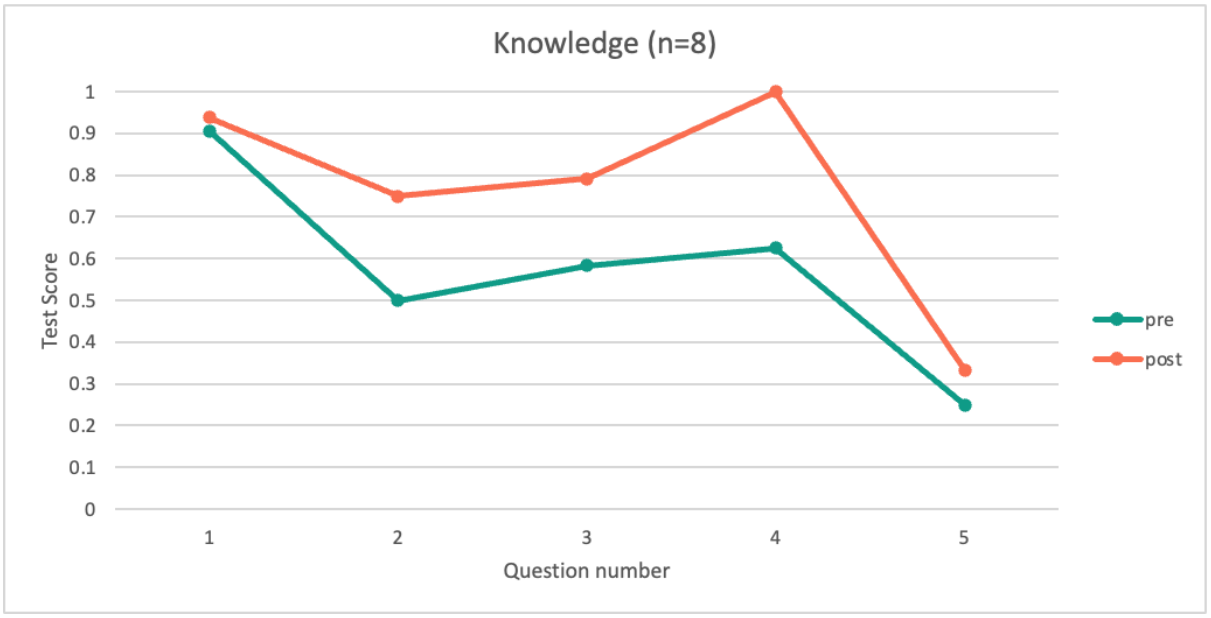


Figure 7. Results of the AI and data literacy knowledge test in Graz on 17.05.2023

Item 5 is a task where students evaluate a confusion matrix. Little progress was made here. A possible explanation for this is the limited time available to the participants to work through the three stations. The Teachable Machine exercise, which ends with a test of the self-created classifier and an accuracy calculation, may not have been completed by the students.

Overall, however, there is an improvement in each item among the participants, although the significance is limited by the small number of participants (n=8).

AI acceptance test

In the AI acceptance test, as in Berlin, there are only minor changes over the course of the workshop. Acceptance increases for 9 items, stays the same for 4 items, and decreases for 3 teams. Overall, there is a tendency towards higher acceptance of AI applications.

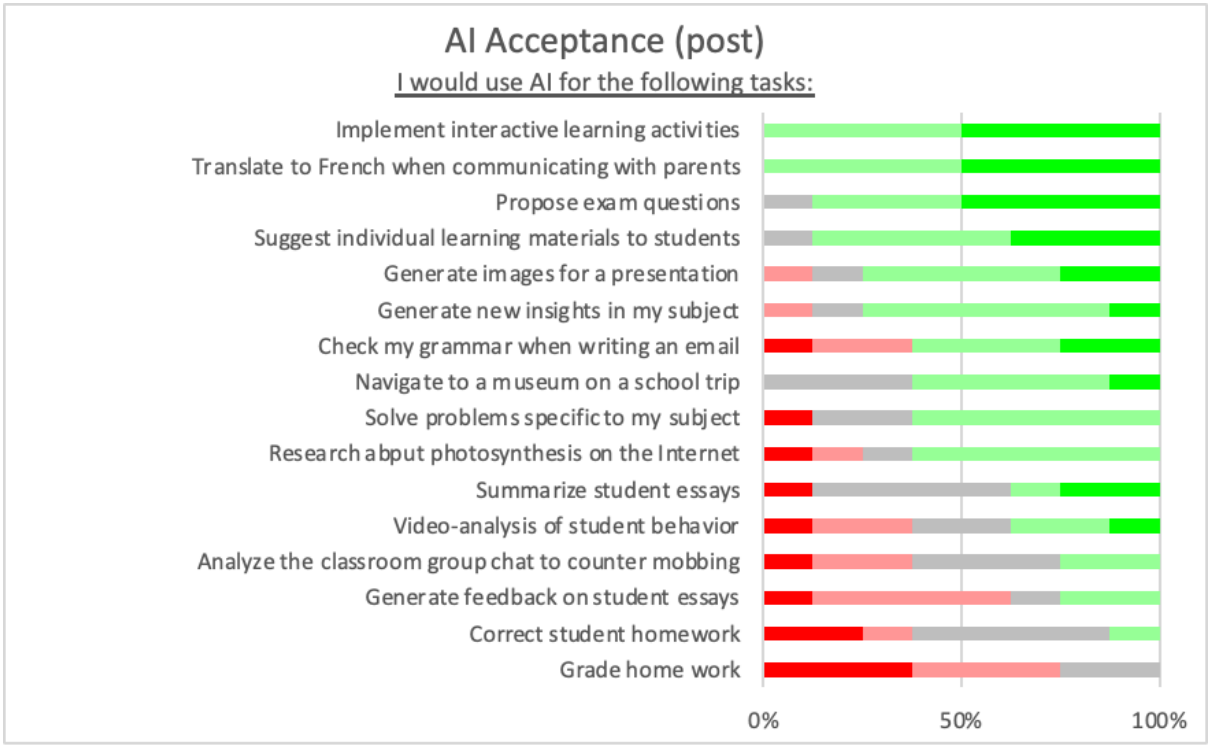


Figure 8. Results of the AI acceptance test in Graz on 17.05.2023

As in Berlin, interactive learning activities (along with AI translation) are the most accepted, while homework assessment is the most rejected. In contrast to Berlin, homework correction by Austrian STEAM teachers is viewed critically and largely rejected, as is AI-generated feedback on student essays. Overall, however, the results are comparable, even though the workshops had different audiences and were conducted in different countries.

AI attitude test

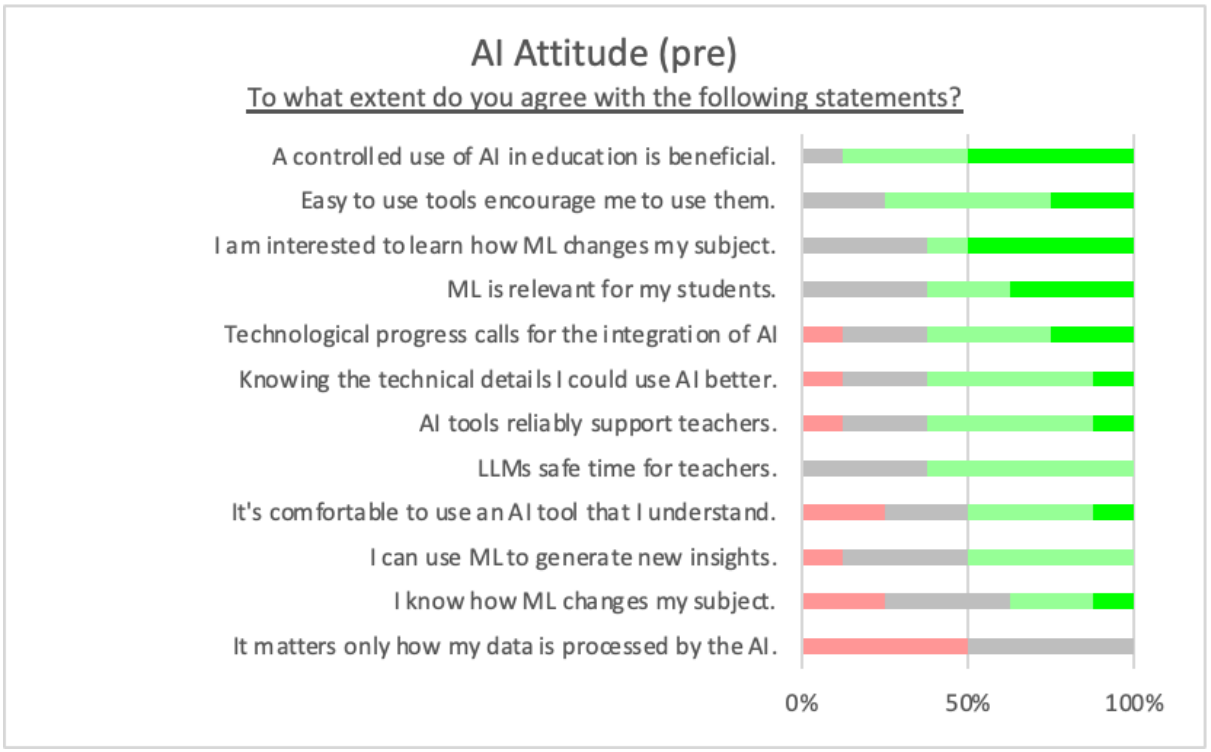


Figure 9a. Results of the AI attitude pre-test in Graz on 17.05.2023. The items are sorted by acceptance level.

Attitudes towards AI systems become slightly more positive as a result of the workshop. As already observed in Berlin, it is noticeable that the attitudes of the participants become more similar in the post-test than in the pre-test.

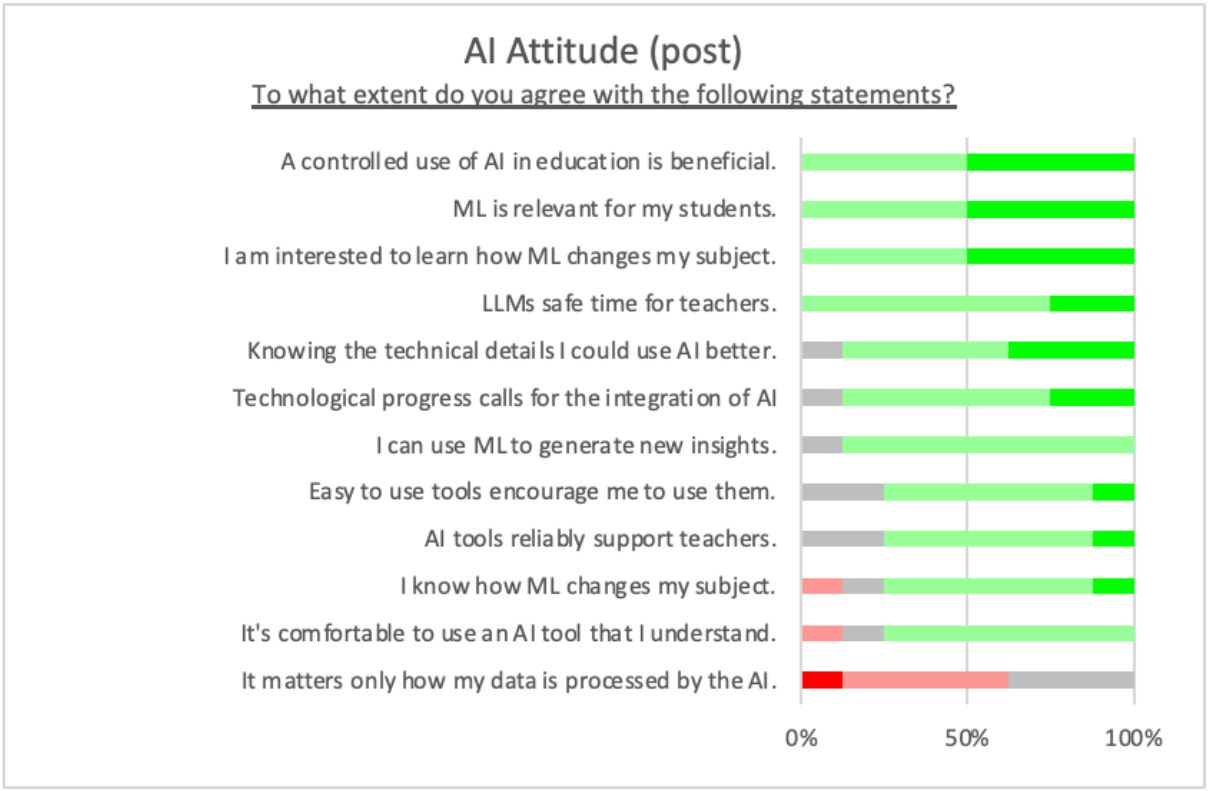


Figure 9b. Results of the AI attitude post-test in Graz on 17.05.2023. The items are sorted by acceptance level.

26.05.2023 | Vilnius | STEAM Teachers

Figure 10. Training Session with STEAM teachers in Vilnius on 26.05.2023. Photo by Alvida Lozdienė

The third workshop in the second round of interventions was conducted on May 26th in Vilnius. The session was conducted by a facilitator and two assistants. 12 STEAM teachers from a middle school (progimnazija, covering grades 5-8) participated. Represented were the following subjects: technology, biology, Lithuanian language, Russian language, math and history. The intervention was hosted in a school's computer room that was prepared by the school's administrator (e.g. installing software).

Originally, the same schedule as in Graz was planned for the workshop. But as the topic of AI & DL was new to the participants, the self-taught way of working at the stations seemed to be too difficult. So the facilitators switched to a moderator-centered teaching approach and first introduced each activity to the participants, in the same order as the stations are numbered (1 Grimm, 2 Orange3, 3 Teachable Machine). The participants formed 4 groups of 3 participants each and after each explanation performed the respective activity, followed by a round of feedback in a plenary setting.

Feedback from the facilitator and participants

As with the interventions already reported, we include here the facilitator's report and notes on the participants' feedback.

Feedback on "Grimm's new Tales (Language Models Unplugged)"

- Teachers for Lithuanian language especially liked the activity.

- Through the translation to Lithuanian language, the sentences had a more diverse vocabulary as in English or German. So several sentences did not have any words in common. It depended on the sentences that were used in a group, how many interconnections emerged.
- The facilitator used this as an opportunity to discuss automated translation and properties of Lithuanian language (different endings that do not exist in English).
- One group with a well-connected graph could generate five new sentences.
- Another group did not have any connections (no overlapping words) at all.
- Teachers judged the activity as "easy and understandable".
- One of the insights discussed was "what you can generate depends on the training data".

Feedback on “Discover hidden treasures in data (Orange3)”

- the participants understood the main idea, but won't be able to repeat the analysis with other data.
- In general, the digital skills of the teachers are low. For example, some participants did not know how to drag-and-drop an icon.
- The facilitator told them "you can study after the seminar and try by yourself".
- A better sequence of exercises would have been to conduct Teachable Machine before Orange3.

Feedback on “Getting creative and solving problems with AI and data (Image recognition with Teachable Machine)”

- They solved the malaria task.
- The facilitator also demonstrated how to train a model with the camera that recognizes if a person smiles or does not smile.
- An idea by a history teacher: take the class on a field trip around the city and let them make pictures of buildings. Then train a model to classify architectural styles like gothic, romanic, barrock, modern etc.
- Teachers asked: "How exactly can we relate the case study to our subject?" They saw an example for biology (Malaria). "How can I get an example for my subject?""
- We see and understand the example from biology. How can we get more examples for other subjects? Like math, history e.g.
- Compared to Orange3, most participants preferred Teachable Machine.

Survey results

Because the training took place on a Friday afternoon, at the end of the school week and after classes, many teachers were tired and wanted to go home after the training. Not everyone completed the entire post-test. In addition, the survey data show that some individuals answered a number of questions with the same option. This casts doubt on whether they had read the questions and is consistent with the impatience observed by the trainers at the end of the workshop. Some items had to be explained orally by the instructor.

AI and data literacy knowledge test

A notable difference from the Graz test is the apparently lower prior knowledge of the participants. In all items, the field of participants improved over the course of the training. In addition, there might be a pre-selection bias due to the fact that in Graz the participants deliberately chose this workshop, while in Vilnius the training was conducted at the school and the teachers only had the choice to attend or not.

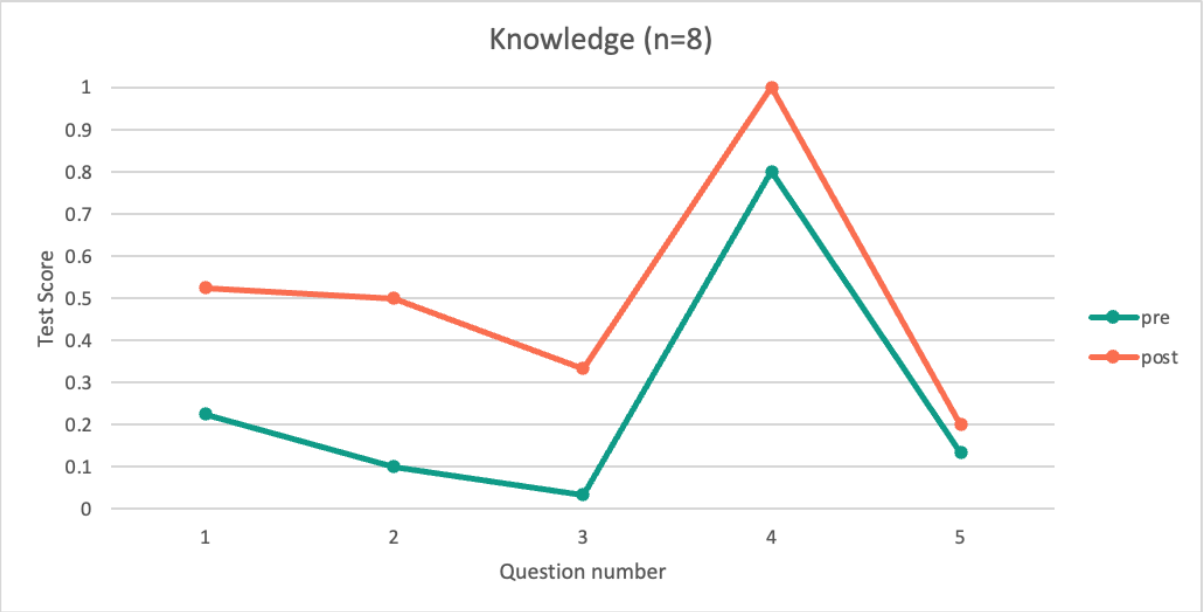


Figure 11. Results of the AI and data literacy knowledge test in Lithuania on 26.05.2023

AI acceptance test

In contrast to the tests in Berlin and Graz, it is noticeable that the measured AI acceptance changed significantly during the workshop. However, as mentioned above, it was observed that the post-questionnaire was filled out hastily by the participants.

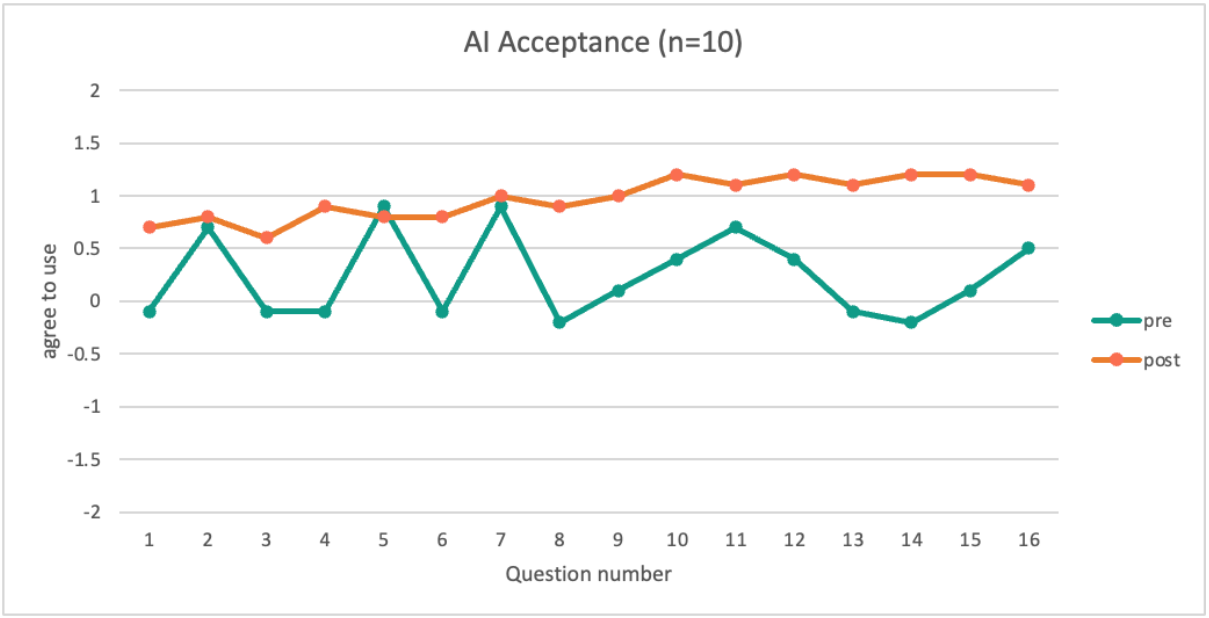


Figure 12. Results of the AI acceptance test in Lithuania on 26.05.2023

There are also some peculiarities in the pre-test results. For example, the item "correct student homework" was rejected less often than "grade homework".

AI attitude test

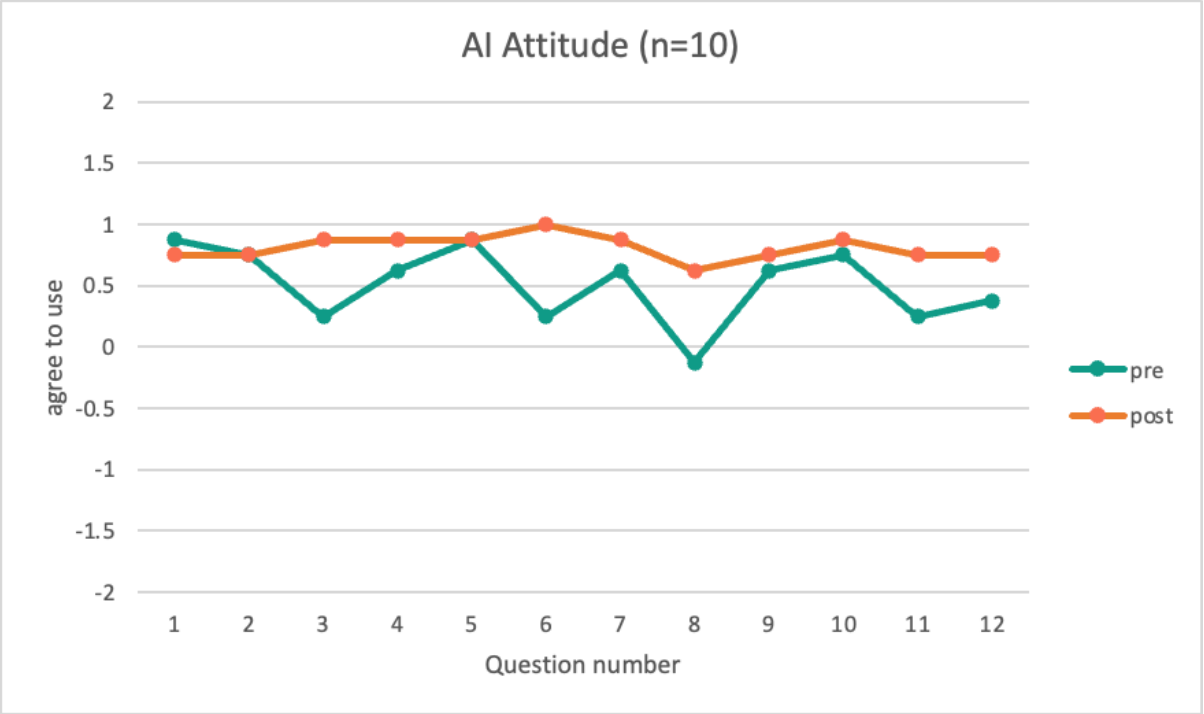


Figure 13. Results of the AI attitude test in Lithuania on 26.05.2023

A similar picture emerges from the AI Attitude test. As can be seen in Figure 13, the differentiated attitudes in the pre-test have leveled off in the post-test.

Focus Group Sessions

In cooperation with the TrainDL project partner Landesinstitut für Schule und Medien Berlin-Brandenburg (LISUM), we conducted two focus group sessions with a total of three training experts from LISUM and four STEAM teachers from two secondary schools in Brandenburg. During these videoconferences, we presented our training concept, the stations and the teaching materials, and engaged in feedback to better understand the STEAM teachers' opinions about our pedagogical approach and design of teaching materials. The results are presented below.

30.05.2023 | Jahngymnasium Rathenow

Two teachers from Jahngymnasium Rathenow, two experts from LISUM and one staff member from the TrainDL project (WP2) participated in the first exchange. The session lasted about 1½ hours. The feedback is given as bullet points, grouped according to the materials.

Feedback on “Grimm’s new Tales (Language Models Unplugged)”

- Additional learning objective: Explain the "temperature" parameter that can be used to control the randomness of generating new sentences.
- For a plugged-in version, the privacy requirements are unclear. In fact, no applications should be used that collect data from students outside the EU (Google, Microsoft). This also applies to chatGPT. De facto chatGPT can be used anonymously via the school server. Idea: Use fobizz tools

Feedback on “Discover hidden treasures in data (Orange3)”

- The motivation for including data literacy in high schools could be science propaedeutics.
- Learning goal for non-CS subjects: Use AI and data analysis as tools
- Where does data literacy fit into the curriculum? There is no place. So if we want to integrate it into the current curriculum, data literacy should not cost time.
- Especially in biology and chemistry, time is scarce because these are popular subjects that students choose for their exams, and teachers need to make sure that students are well prepared. Not only is time in the classroom limited, but so is the time teachers need to prepare.
- Is there a data set suitable for physics?
- Objective: to find topics that can be taught well with data analysis.
- In the past, there was an attempt to use Excel across subjects. However, this failed because it was not accepted by the staff. Orange3 is more complex than Excel, so it is likely to fail.
- Monkey and decision tree: One teacher knows the material and has used it unplugged. A three-step decision tree was very difficult for students to understand.

Feedback on “Getting creative and solving problems with AI and data (Image recognition with Teachable Machine)”

- Easy to use. A teacher has experience with Teachable Machine and the images provided by Google.
- Idea: Have students take pictures of coins and train a model with those pictures.
- For biology, plant recognition could be a good example, e.g. collect leaves and use them to train the Teachable Machine.
- In Brandenburg, WAT (Wirtschaft, Arbeit & Technik) is a compulsory subject up to grade 9 (Gymnasium) or 10 (Oberschule). This subject is often taught as computer science for all students and includes physical computing (e.g. using Calliope microcontroller) and block-based programming. Teachable Machine would fit in well here.
- For WAT, the question "How do machines work?" is important, but also the socio-technical perspective "How does AI change our society?
- The unit should include that students collect their own data, as this is an essential part of the data life cycle, or that students search the WWW for data. On the other hand, this requires additional time and teacher training.
- The material would be well suited for a project day. However, the concept of a project day is not sustainable because it stands and falls with one motivated teacher and therefore cannot be rolled out on a large scale. Elaborate concepts from one year are often not used in subsequent years.
- As long as the students do not have their own devices, the lesson with Teachable Machine has to be taught in the computer room. The danger is that the lesson will be perceived as "not so important" because it is not part of the core curriculum and the knowledge will not eventually be part of an exam.

08.06.2023 | Otfried Preußler Oberschule Großbeeren

The second focus group meeting was attended by two teachers from the Otfried Preußler Oberschule in Großbeeren, one expert from LISUM, two staff members from the TrainDL project (WP2), and one staff member from the Computing Education Research Group at the Freie Universität Berlin. The session lasted a little more than two hours. In addition to feedback on individual stations, there were also general comments and suggestions.

General Feedback

- LISUM runs the portal Jetzt Wirklich Digital, which covers topics related to digitization in schools.
- Teachers report that students use AI tools for their homework, and entire essays have been written by ChatGPT.
- There is a consensus that AI should be included in the curriculum. How can teachers be trained for this?

- Licensing: LISUM recommends releasing educational materials under the CC-by-sa license. The attribute "non-commercial" would prohibit the use at adult education centers (Volkshochschule) and private schools.
- Idea: analyze sound waves of music tracks and extract features, thereby clustering similar tracks.
- Suggestion: Include information about the materials, grade levels, and skill areas they cover. Refer to the curriculum frameworks that exist in all states.[4]
- Suggestion: contact the iMint Academy[5] for collaboration on subject-specific activities (activities using AI & DL for a non-CS subject).
    - Clarify whether our material (Teachable Machine / Orange3) can be used in biology.
    - In the iMint Academy, teachers from different school levels create OER teaching materials.
    - Recommendation: Subject Set Biology.[6]

Feedback on "Grimm's new Tales (Language Models Unplugged)"

- According to the teachers, the material is suitable for students from 7th grade. For younger students, the material should be supplemented with intermediate steps.
- For elementary school, the material could be made more playful, e.g. with a game board.
- To be clarified (for a didactic handout):
    - What background knowledge should the teacher have?
    - What previous knowledge do the students need?
- See if there is a link to German lessons: check the grammar, what does the generated sentence mean. Possibly the material could be adapted for foreign language lessons, e.g. with English sentences.
- To generate new sentences, randomness can be simulated by the teacher preparing numbered cards with the possible continuations (all nodes with more than one incoming edge).
- A sample solution should be provided for a completed graph (Task 2).
- Report of an exercise in a media class:
    - The teacher asks the students to open WhatsApp on their phones and use the autocomplete feature to create sentences.
    - As a result, all the sentences were different.
    - Students were able to recognize vocabulary from messages they had recently written.
    - They compared it to autocomplete in the Google search form.
- Variations and other exercises that can be used to teach automatic language processing:

[4] Example for biology: https://bildungsserver.berlin-brandenburg.de/rlp-online/c-faecher/biologie/kompetenzen-und-standards
[5] https://bildungsserver.berlin-brandenburg.de/i-mint-akademie
[6] https://bildungsserver.berlin-brandenburg.de/unterricht/faecher/mathematik-naturwissenschaften/mint/i-mint-akademie/weiterfuehrende-schulen/fachset-biologie

  - Word Cards: Students write a word on an index card, which is then used to create sentences..
  - Design strips of paper with gaps in sentences. The students' task is to write the correct words in the gaps and compare them with complete sentences from a document.
  - There is an analogy between a fill-in-the-blank exercise and the way a large language model learns. Exercise: Students guess the words that belong in the gap. Could be done digitally with the H5P “Fill in the Blank” module.
  - Students cut out the word strips and then count the frequency of the words.
  - Mentimeter word cloud: the most frequent words are displayed largest.
  - Fridge Magnets[7]
- Idea: design different stations on language models.
- How can a machine that doesn't know what it's saying pass the Bavarian Abitur? Learning objective: the machine doesn't know what it is saying (Stochastic Parrot).

Feedback on “Discover hidden treasures in data (Orange3)”

- The title "Discover hidden treasures" (Verborgene Schätze in den Daten entdecken) sparks interest.
- The tool is definitively too complex for primary school students.
- Framing would be useful for classroom use, e.g: I go for a walk and find several iris flowers. How can I determine which type of iris they are? I am training an AI for this task.
- The activity can be integrated into a field trip. Take the students for a walk and take pictures of the plants. Question: How can the AI recognize the species in the photo?
- For teachers, the effort must be worth the result. For many teachers, learning how to use Orange3 is not worth the effort.
- The data sets (and the research questions for them) are critical to integrating the material into STEAM subjects.
- For older students, other domains such as music (genre, artist, album, and song) might be of interest as a data set. (Link to Shazaam and Spotify).

Conclusions from the Focus Group Sessions

Unplugged activities offer a significant advantage over computer-based activities because they remove any technical barriers. This makes them particularly suitable for STEAM teachers. The Grimm activity could even be evaluated in an elementary classroom. Acceptance of browser-based tools tends to be higher than for applications that need to be installed on PCs.

Challenges include limited classroom time and teacher preparation time. Curriculum reform could support more effective use of limited classroom time and help teachers signal the importance of the topic.

[7] https://www.supermagnete.de/kuehlschrankmagnete/magnetwoerter_LIV-47

The intrinsic motivation of STEAM teachers to integrate AI & DL into their teaching appears to be low. One approach to make it easier for them is to develop activities and materials tailored to specific classroom topics that teachers can master without significant time investment. However, the creation of such materials require collaboration with the relevant subject didactics. Currently, a researcher from WP2, in collaboration with the History Education Research Group at the Freie Universität Berlin, is conducting research on an example subject (history) to demonstrate what such an activity might look like.

Challenges

In terms of content and activities, we received very mixed feedback on learning station "Discover hidden treasures in data". Orange3 was challenging for teachers without a strong background in computer science. It should be kept in mind that it takes a considerable amount of time and effort for teachers to become familiar enough with an application to be able to use it to teach their students. In addition, many teachers are not familiar with data-driven approaches that are already widely used in science. Here, we see the responsibility of subject didactics to integrate it into teacher training and to select appropriate problems and data sets for each subject.

We also received feedback from computer science teachers that the AI ethics case studies were too simple for the grade levels at which computer science is taught in Berlin (9th grade and up). Computer science teachers would need training in ethics and pedagogical experience in discursive methods in order to integrate teaching ethics into computer science classes.

In terms of the teacher training format, most training sessions for teachers are organized by educational institutions or associations, to which the facilitators are invited. The content of the training courses could be proposed by the facilitator, but the format (duration, number of participants) was given by the framework of the event. Since such events usually have a limited duration, it is difficult to provide an in-depth introduction to AI in schools, answer all questions of teachers, practice teaching materials and discuss the concerns. Future projects could create formats for introducing teachers to AI and data literacy over a longer period of time, such as six months or a year.

The intervention in Vilnius was organized in cooperation with a school. A challenge here was that the teachers were not released for the training and the event took place after school hours, on a Friday afternoon, which affected the receptiveness and concentration of the participants. We experienced that the format of self-learning stations was not accepted by the participants.

Conclusion and Future Work

In contrast to the first round, in which we based our activities on existing materials, for the second round we developed new materials that address new developments in the field of generative AI, perform data analysis with image recognition, and discuss ethical aspects of AI. These materials were adapted for the target audience of STEAM teachers. We evaluated the learning success of the training courses in the form of a competency test and also evaluated how the attitudes of the participants towards artificial intelligence changed.

Through focus group sessions, we received valuable feedback from STEAM teachers on the learning materials and activities regarding their suitability for the classroom.

The evaluation of the policy experiment questionnaires and guided interviews by WP4 is still ongoing. Therefore, the following conclusions, which will inform the design of the third round of interventions, are preliminary.

- When teaching AI & DL, a distinction can be made between the technical, application, and societal perspectives. While computer science teachers regularly address the technical perspective and focus on computer science concepts, STEAM teachers lack the technical background, so the application and societal perspectives should be addressed first.
- STEAM teachers expect materials and activities tailored to their discipline and curriculum. Such activities need to be created in collaboration with the subject didactics.
- The Grimm materials and the Teachable Machine activity received positive feedback. Following an iterative approach, they should be further refined so that they can be used in language and science classes, respectively.
- The workshop format should be flexible and modular so that it can be adapted to the time frame given by the training organizer.
- The individual modules and activities should be designed so that they can be used for independent learning at stations as well as for teacher-led plenary sessions. In this way, different groups of participants can be addressed and, in the case of longer training courses, a variety of methods can be used.

List of Abbreviations

AI	Artificial intelligence
CS	Computer science
DL	Data Literacy
FU	Freie Universität Berlin
LLM	Large Language Model
OCG	Austrian Computer Society
STEAM	Science, Technology, Engineering, the Arts, and Math
TrainDL	Train Data Literacy
UP	University of Potsdam
WP2	Work package 2
WP4	Work package 4

Appendix

The appendix contains sample materials we used during the training sessions. Photos are included, as well as excerpts from the skills test and the AI attitudes questionnaire. The materials are included for illustrative purposes only and do not purport to be complete documentation.

Learning stations

Station 1

Computing Education

Mit KI Texte erschaffen

Neue Märchen, Rezepte oder Prüfungsaufgaben generieren - das sind nur einige der Aufgaben, die mit KI in Sekundenschnelle erledigt werden können. Was steckt eigentlich dahinter?

In dieser Station lernst du, wie man aus einem Textdatensatz ein einfaches Sprachmodell erstellt, das neue Texte generieren kann.

Und so geht's

2-4

4 **Spielfelder**
1 **Spielstein**
Briefumschlag mit ausgeschnittenen Grimms Beispielsätzen
Arbeitsblatt 1
Sprachmodell trainieren
Arbeitsblatt 2
Mit dem Sprachmodell neue Märchen erstellen
Arbeitsblatt 3
Aufgaben zur Vertiefung und Sicherung

Du wirst in der Lage sein...
... zu erklären, wie einfache Sprachmodelle entstehen.
... ein einfaches Sprachmodell zu erstellen, das neue Texte generieren kann.
... unterscheiden zwischen dem Training eines Sprachmodells und dem Benutzen eines Sprachmodells.
... die Grenzen von Sprachmodellen zu benennen.

Ablauf

1 Nimm dir eines der Spielfelder.

2 Ziehe mit geschlossenen Augen zwei Textauszüge aus dem Briefumschlag.

3 Trainiere dein Sprachmodell. Die Anleitung dazu findest du im **Arbeitsblatt 1**.

4 Das Training ist abgeschlossen und du kannst dein Sprachmodell verwenden, um neue Sätze zu generieren. Die Anleitung dazu findest du im **Arbeitsblatt 2**.

5 Siehst du die Grenzen deines Sprachmodells? Kooperiere nun mit den anderen Spielern, um die Texte noch informativer und kreativer zu gestalten. Reflektiere anschließen zur Übung im **Arbeitsblatt 3**.

Station 2

Computing Education

Verborgene Schätze in den Daten entdecken

In Geschichte wird viel mit Textdaten gearbeitet, in Biologie mit Bilddaten, in Mathematik und Physik mit Zahlen. Texte, Bilder, Zahlen können mit den Methoden der Datenanalyse und der künstlichen Intelligenz neue Erkenntnisse über die Welt liefern und uns helfen, verschiedene Probleme zu lösen. Nicht umsonst gilt die Fähigkeit, mit Daten planvoll und kritisch zu arbeiten, heute als viertes Standbein der Wissenschaft.

In dieser Station lernst du grundlegende Arbeitsabläufe kennen, du analysierst Daten und entdeckst darin verborgene Schätze.

Und so geht's

1-2

Computer mit dem installierten Programm **Orange3**

Aufgabenblatt „Entscheidungsbäume in Orange3"

Anleitungen Orange3 "Erste Schritte in und "Blüten mit maschinellem Lernen klassifizieren"

Hilfekarten Orange3

Datenlebenszyklus an einem Beispiel

Weitere **Datensätze** zum Thema Fast Food

Du wirst in der Lage sein...

... Datenanalyse-Workflows nachzuvollziehen.

... durch die Analyse von numerischen Daten fachlich relevante Fragen zu beantworten und neue Erkenntnisse zu gewinnen.

... verschiedene maschinelle Lernmodelle mit Orange3 zu erstellen, zu trainieren und zu testen.

Ablauf

1 Starte **Orange3** und blättere das **Aufgabenblatt** sowie **Anleitungen** durch.

2A Folge der **Anleitung** "Erste Schritte in Orange3". Du lernst, einen Datensatz zu laden, zu visualisieren und zu explorieren. Arbeite anschließend die **Anleitung** "Blüten mit maschinellem Lernen klassifizieren" durch.

2B Zusätzlich oder alternativ arbeite das **Aufgabenblatt** „Entscheidungsbäume in Orange3".

3 Experimentiere mit einem weiteren Datensatz in Orange3. Lege dafür ein neues Projekt an.

4 Analysiere den Datensatz. Benutze zur Orientierung die **Anleitung** sowie die **Hilfekarten**.

Station 3

Computing Education

Mit KI und Daten kreativ werden und Probleme lösen

Gesunde von kranken Blutzellen unterscheiden, Briefumschläge in Millisekunden sortieren, Schäden in der Produktion erkennen - all das und noch viel mehr ist heute dank maschinellem Lernen möglich. Von der Zahnmedizin bis zur Numismatik - maschinelles Lernen revolutioniert viele Disziplinen und eröffnet neue Erkenntnisse und Perspektiven.

An dieser Station kannst auch du mit maschinellem Lernen experimentieren und kreativ werden. Finde heraus, wie maschinelles Lernen auch dein Fachgebiet verändert.

Und so geht's

1-2

Rechner mit Internetverbindung und einer Kamera

Teachable Machine

AI Playground

Anleitung „Malaria erkennen mit Teachable Machines"

Datensätze

- Malaria erkennen
- Analyse des Bestäuberverhaltens
- Pflanzenkrankheiten vorhersagen
- Münzen analysieren

Du wirst in der Lage sein...

... den Zweck der Trainings- und Testdaten bei der Erstellung eines Klassifikationsmodells zu erklären.

... ein Klassifizierungsmodell zu trainieren, um eine fachspezifische Frage zu beantworten.

... die photographische Artefakte auf deren Einfluss auf die Güte des Modells zu bewerten.

... das Klassifikationsmodell mithilfe der zurückgelegten Testdaten zu testen.

... das Modell zu benutzen, um eine fachspezifische Frage zu beantworten.

Ablauf

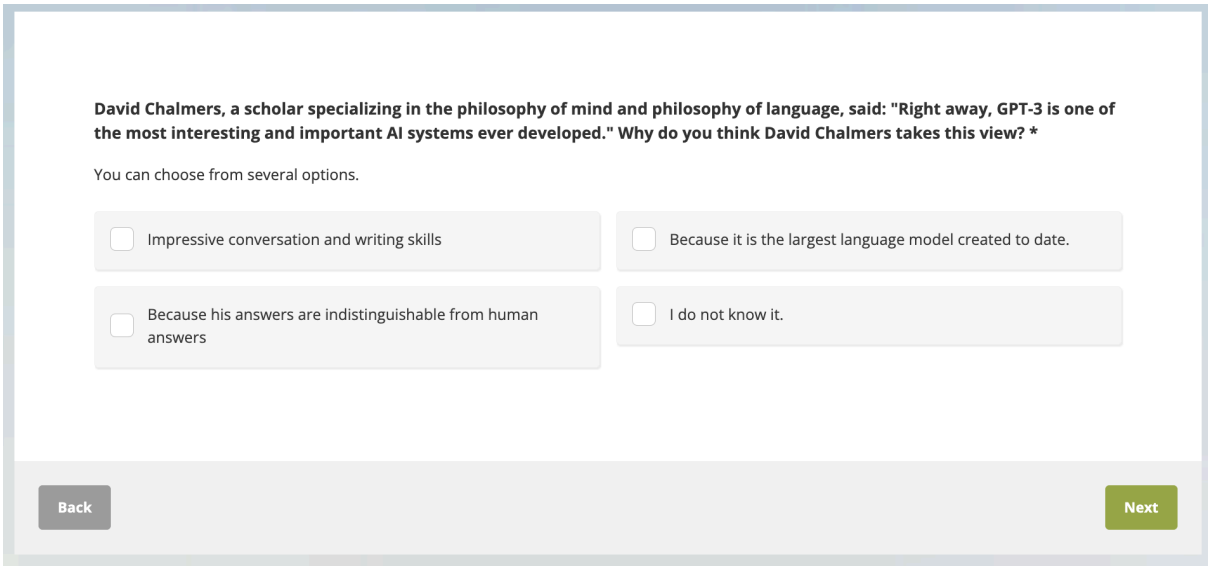
1 Mache dich mit der Problemstellung in der **Anletiung** „Malaria erkennen mit Teachable Machines" vertraut.

2 Bearbeite Schritte 1-3 aus der Anleitung. Starte **Teachable Machine** und bearbeite Schritte 4-8.

3 Exportiere das Modell und arbeite damit weiter im **AI Playground**.

Online Survey Questions (sample screenshots)

AI acceptance questionnaire



AI & DL competency test

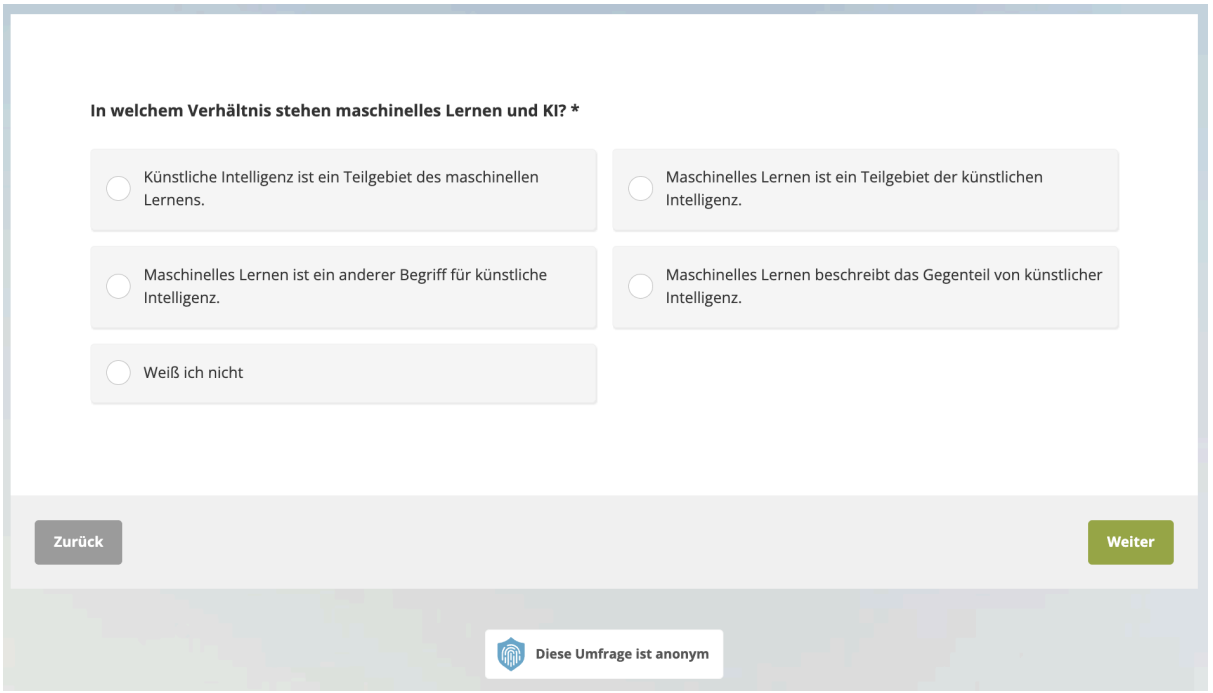
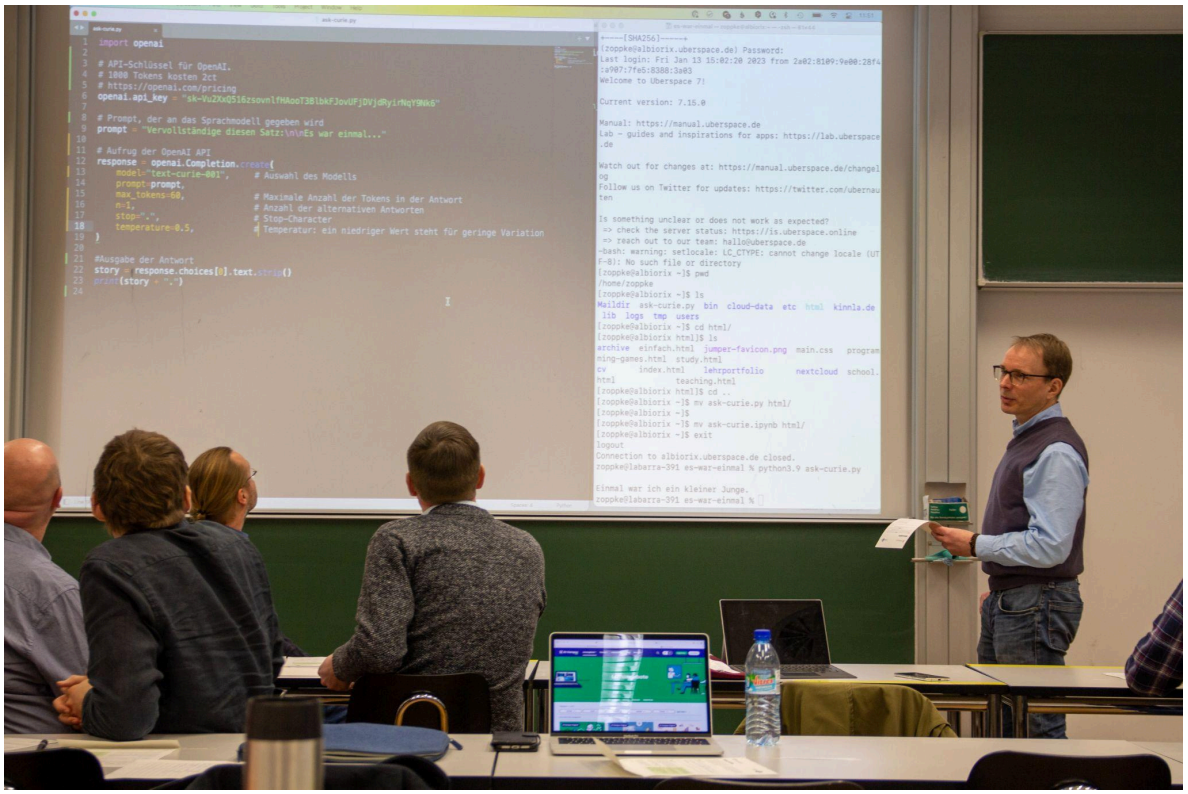


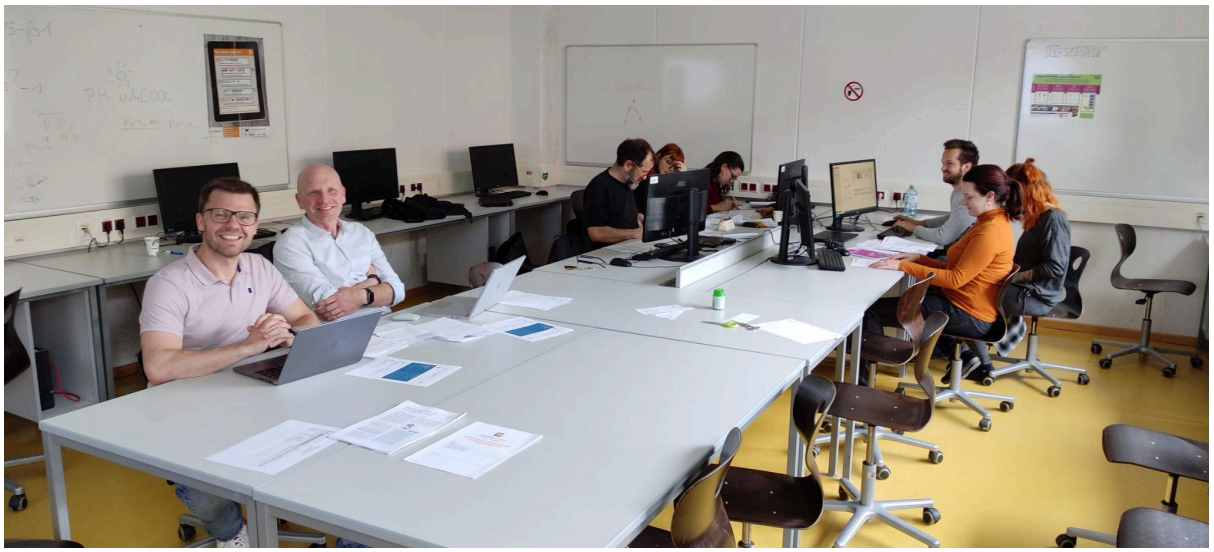
Photo documentation

09.03.2023 | Berlin



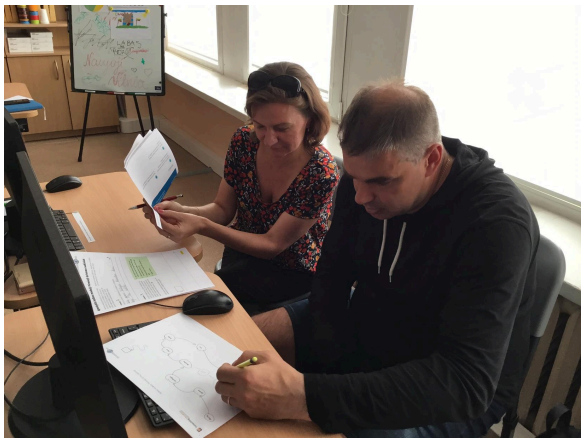
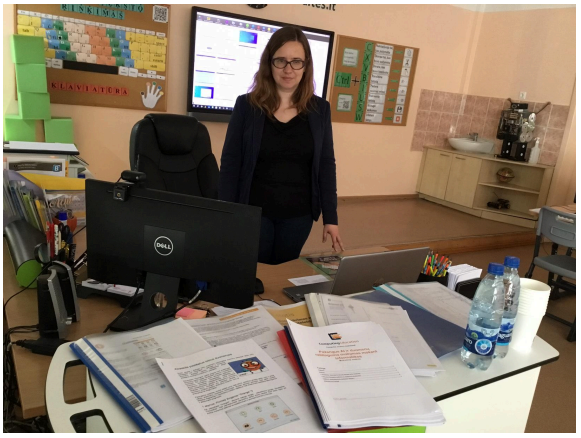
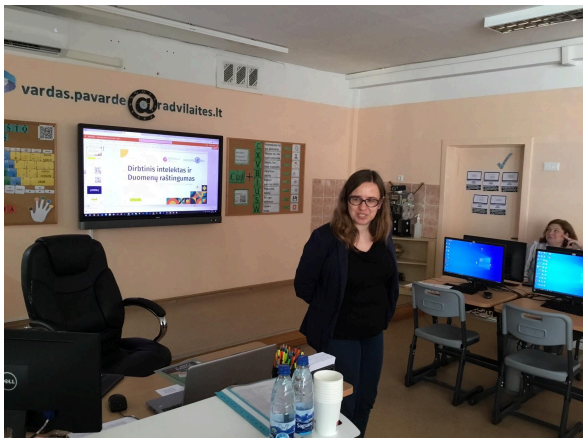
Photos by Viktoriya Olari

17.05.2023 | Graz



Photos by Austrian Computer Society (OCG)

26.05.2023 | Vilnius



Photos by Alvida Lozdienė