>TrainDL

# >Teacher training for Data Literacy & Computer Science competences // Report Round 2 // **Deliverable 4.5: Report on Second Evaluation Phase**

train-dl.eu

*Martin Reger, Evgenia Samoilova, Ulrike Lucke*

*University Potsdam*

# Content

# List of Figures

Figure 16 Boxplot of post results for the survey items on teachers' perceived competences and willingness to invest time and effort to incorporate DL and AI into their teaching, Berlin/Graz/Vilnius, post-evaluation survey, n=35 .....30

# List of Tables

# Executive Summary

## Introduction

The TrainDL project aims to provide policy recommendations for integrating the subjects of data literacy (DL) and artificial intelligence (AI) into teacher university education, as well as professional development programmes for teachers. To achieve this, the project adopts an iterative approach to design, deliver and evaluate teacher trainings. The project comprises three intervention rounds targeting computer science (CS) teachers as well as teachers from other subjects at the primary and secondary levels. This report presents the evaluation results of the second round of interventions, focusing on the implemented training concepts targeting in-service computer science (CS) and STEAM teachers[1] at the secondary level. The evaluated trainings included four trainings conducted between March and May 2023 in Berlin, Germany (2 hours); Vilnius, Lithuania (4 hours); and Graz, Austria (4 hours). Within the trainings, in-service teachers were primarily introduced to the topic of generative AI, including a theoretical introduction to large language models (LLM), an unplugged exercise, an introduction to GPT API, as well as case studies on ethical aspects of LLMs. While the theoretical foundations and the unplugged exercise were the essential part of all the trainings, some sessions had to skip the API and/or ethical foundations due to the time shortage.

The evaluation of the all three rounds of interventions focuses on teachers' capacity to integrate the acquired content on DL and AI into their teaching, as measured immediately after the trainings. The first intervention round included 8-hour sessions. The second round adapted to shorter training durations, including 44 in-service teachers: 2 hours 15 minutes for CS teachers and 4 hours for STEAM teachers. This

[1] STEAM teacher interventions refer to the interventions targeting teachers who teach non-CS subjects (specifically Science, Technology, Engineering, Arts, and Mathematics)

reduction in hours was primarily due to logistical challenges in scheduling 8-hour slots for training, as well as difficulties in finding suitable platforms for reaching teachers. Consequently, existing formats were utilized, such as for example, IBBB conference[2] in Berlin for CS teachers. In the second round, we also examined teachers' awareness of the significance that DL and AI have for their subjects. The results of the follow-up data collection, conducted approximately six months after the training session, will be reported and discussed in the final report. While the CS training was uniquely conducted in Germany and differed from the initial rounds in terms of both hours and content, the STEAM trainings took place in Austria and Lithuania, but not in Germany. This report will respectfully acknowledge these variations, focusing on providing a descriptive overview of each training. We aim to present the unique characteristics of each session without emphasizing comparisons, appreciating the distinct contexts of each training.

To evaluate the training sessions, the following instruments were used:

- The evaluation survey administered before and immediately after the training
- The DL and AI self-assessment and knowledge test, which includes both self-assessment and knowledge questions on DL and AI, administered before and immediately after the training;
- The semi-structured personal and online interviews administered right after the training.

Table 1 presents an overview of the evaluated training sessions and the evaluation instruments used.

| Date | Location | Target group | Duration | Evaluation instruments used | Number of participants |
|---|---|---|---|---|---|

[2] https://fg-ibbb.gi.de/veranstaltung/19-gi-tagung-zur-schulinformatik-in-berlin-und-brandenburg

| | | | | | |
|---|---|---|---|---|---|
| 09.03.2023 | Berlin, Germany | In-service CS teachers at the secondary level | 2 hours 15 minutes | • (Pre- and post-) evaluation survey<br>• (Pre- and post-) DL and AI self-assessment and knowledge test<br>Semi-structured personal interviews | 12 |
| 09.03.2023 | Berlin, Germany | In-service CS teachers at the secondary level | 2 hours 15 minutes | • (Pre- and post-) evaluation survey<br>• (Pre- and post-) DL and AI self-assessment and knowledge test<br>• Semi-structured personal interviews | 12 |
| 17.05.2023 | Graz, Austria | In-service teachers teaching STEAM subjects at the secondary level | 4 hours | • (Pre- and post-) evaluation survey<br>• (Pre- and post-) DL and AI self-assessment and knowledge test<br>• Semi-structured personal interviews | 8 |
| 26.05.2023 | Vilnius, Lithuania | In-service CS teachers teaching STEAM subjects at the secondary level | 4 hours | • (Pre- and post-) evaluation survey<br>• (Pre- and post-) DL and AI self-assessment and knowledge test<br>• Semi-structured personal interviews | 12 |

**Table 1** Overview of the evaluated trainings of the second intervention round: date, location, target group, duration, evaluation instruments used, number of participants

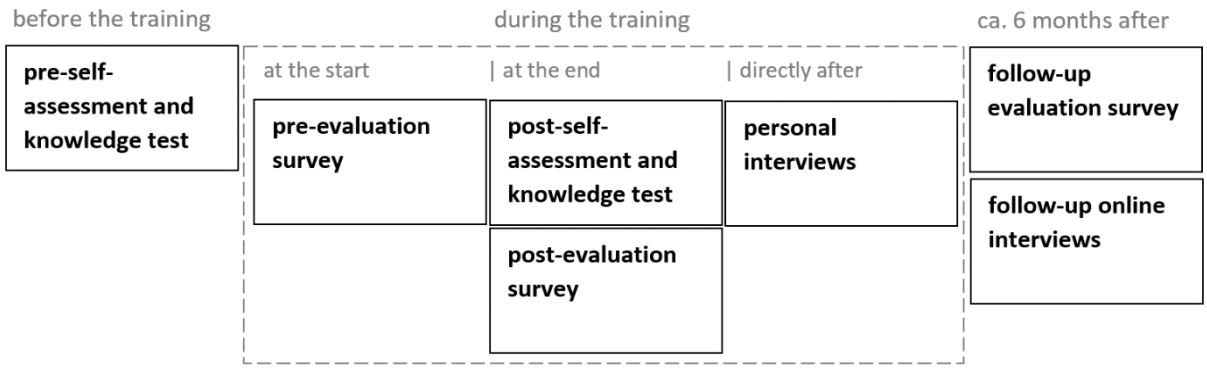
This report is structured into six sections. The first section outlines our research questions and hypotheses, and evaluation methodology. The quantitative and qualitative findings are described in the sections 2 and 3 respectively. Sections 4, 5, and 6

wrap up the report, presenting a comprehensive summary of the salient findings across all trainings and discussing limitations.

# 1. Methodology

Deliverable 4.3 includes a detailed description of the methodology and research questions and hypotheses used for all three intervention cycles. The project uses an action research methodology (Baskerville and Wood-Harper 1996; Burns 2010), characterized by its iterative nature, involving multiple rounds of designing, implementation, observation, feedback, and reflection. This report specifically addresses the research questions related to the second round of interventions designed for secondary CS and non-CS (STEAM) in-service teachers.

To evaluate the trainings, we followed the procedure outlined in Figure 1. To gain a more nuanced understanding of the trainings' impact, we employed a mixed methods approach following a concurrent nested design suggested by Creswell and Plano Clark (2018). This design allowed us to enrich and clarify our quantitative findings using qualitative data. The quantitative data were primarily used for the examination of the participants' characteristics and changes or lack thereof in teachers' perceived competences on how to use DL and AI in class, as well as their understanding of these concepts introduced during the training. Additionally, qualitative interviews with the teachers provided additional insights into their experiences and perspectives on the training effectiveness. They also highlighted the teachers' expectations for future training content and identified barriers to DL and AI integration into the classroom.



Additionally, participants offered suggestions for policy changes that could better facilitate the inclusion of DL and AI into their teaching. The follow-up data collection (that will be reported in the final report) captured the ability to integrate DL and AI into the teaching.

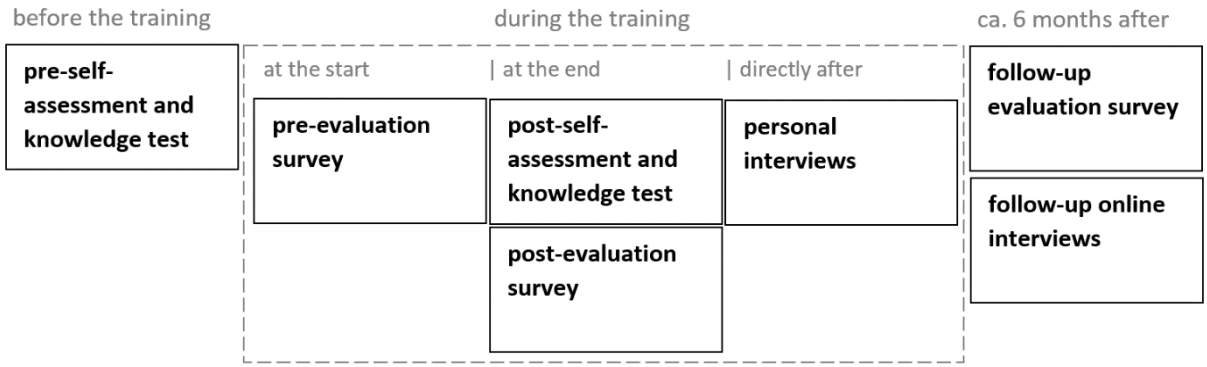



**Figure 1** Overview of the evaluation process for each training in the second intervention round

To ensure the privacy of participants while still enabling the linkage of pre-, post- and follow-up datasets, participants were requested to create a unique pseudonymisation code (see Appendices 1-5 for translated versions of the surveys for Germany[3]), which they were required to enter or recreate during each subsequent round of data collection. This approach allowed for the protection of participants' privacy while maintaining the ability to connect and analyse the various datasets. Both the survey data and interview data were collected following informed consent, which included comprehensive information about anonymization, data storage, retention period, potential publication of anonymized data, and the option for participants to withdraw their consent and have their data deleted. The project did not require any personal data from the teachers, so no questions pertaining to personal information were included. Any personal information present in the interview transcript (such as place of work or names) was removed. Contacting potential respondents for the follow-up data collection is being done via local partners, eliminating the need to collect and store contact information.

[3] Questionnaires and interview guides in Germany, Lithuania, and Austria were almost identical with exception of the country-specific questions (e.g., types of schools, states, subject names).

The following definitions of DL and AI were used in the project and shared with the training participants, particularly in the evaluation surveys:

- DL is the ability to systematically handle data and consciously utilize and question them in the respective context. This includes the competences to collect, explore, manage, analyse, visualize, interpret, contextualize, evaluate, and apply data (Ridsdale et al. 2015).
- AI encompasses various technologies and methods that deal with the automation of intelligent behaviour such as decision-making, problem-solving and machine learning.

In the training, the DL content primarily revolved around the concept of the data lifecycle and was taught to the STEAM teachers only.

### 1.1 Quantitative data: research questions, instruments, and analysis

The quantitative data of the second intervention cycle was primarily used to address the hypothesis 1 and 2[4], derived from the two research questions:

1). *What is the effect of the designed DL and AI training on the ability of in-service CS and STEAM teachers to integrate DL and AI into their classes?*

2). *What is the effect of the designed DL and AI training on teachers' awareness of the significance that DL and AI have for their subjects?*

Table 2 describes the two guiding hypotheses used for the second intervention round. The first hypothesis for in-service CS and STEAM trainings, deals with teachers'

[4] Numbers of the hypotheses refer to the Deliverable 4.3, where each developed hypothesis was assigned a number.

ability to incorporate DL and AI into their teaching. Integrating learned content after the training into teaching is a process that takes time. Therefore, the most suitable measures for such integration are the ones over an extended period. However, given the ongoing nature of the follow-up data collection and the anticipated low response rate inherent in contacting training participants six months post-training, the quantitative part of this report primarily focuses on perceived ability to integrate DL and AI into the classroom measured immediately after the training. Specifically, we looked at the following aspects with a focus on both pedagogical content knowledge and content knowledge:

- (pedagogical content knowledge) **teachers' perceived competences** on how to use DL and AI content in class, and
- (content knowledge) **teachers' understanding of DL and AI concepts** introduced in the trainings.

| Hypothesis # as assigned in Deliverable 4.3 | Hypothesis text |
|---|---|
| Hypothesis # 1 | If in-service CS/STEAM teachers with a solid background in CS participate in a DL and AI teacher training workshop, they then are able to integrate DL and AI in their CS classes. |
| Hypothesis # 2 | If in-service STEAM teachers without any prior knowledge in CS receive a teacher training, covering basics of DL and AI, they then are aware what significance DL and AI have for their subject and they are able to integrate selected AI & DL topics into their regular teaching. |

**Table 2** Guiding project hypothesis used for the first round of interventions for the pre- and in-service CS teachers in Berlin, Germany; Vilnius, Lithuania, and Vienna, Austria

Additionally, for the CS and STEAM in-service teachers we have looked at the following aspects that can clarify and complement the main findings:

- teachers' feedback on the learned content and format of the training.

The quantitative data were also used to collect information on the socio-demographic characteristics of the participants. As random assignment to trainings was not possible, understanding the participants' characteristics was crucial for interpreting the findings and addressing potential selection bias. Each country's partner was tasked with internally advertising the training sessions, aided by local partners.

To collect quantitative data, we have used two instruments:

- the pre- and post-evaluation survey developed by the University of Potsdam, and
- the DL and AI self-assessment and knowledge test, developed by the Freie Universität Berlin

The data for the CS trainings conducted in Berlin, characterized by almost identical content, the same instructors, and similar dates and target groups, were merged for a statistical analysis. To analyse the pre- and post-data, we employ descriptive statistics and (for the CS sample) the Wilcoxon signed-rank test. This test uses mean ranks to assess, whether there is a statistically significant difference between two related samples: the pre- and post-measures, taken from the same individuals. If the Wilcoxon signed-rank test is statistically significant, this supports the conclusion that there is a difference between the pre- and post-measures. However, the test statistic and p-value from the Wilcoxon signed-rank test do not tell us the direction of the difference (i.e., which group has higher values). Therefore, to interpret the differences, we will look at the descriptive statistics with a focus on measures of variability for non-parametric data - median and interquartile range (IQR) that describes where the middle 50% of the data falls. To visualize the data we mostly use boxplots (that display the median, IQR, and possible outliers), which are very useful in comparing distributions between groups (i.e., pre- and post-measures as well as differences between the countries).

It is important to note that since samples are very small, results should be interpreted with caution. While small sample sizes notably limit the generalizability of the study, they provide indications of trends within the sample and can be valuable when combined with the qualitative results of the interviews.

The evaluation survey:

The questionnaires (see Appendices 1-2) included information on demographics (e.g., sex and age), educational background, type of the school where in-service teachers are employed, teaching hours and subjects, experience with DL and AI as well as

attitudes towards these topics, expectations and the perceived ability to generate student interest. Also included are: engagement for the topics of DL, teachers' feedback on the learned content and format of the training, and most importantly teachers' perceived competences on how to use DL and AI in class. The follow-up survey for the in-service teachers focused on the actual integration of the content into the classroom.

The perceived competences on how to use DL and AI in class were measured via the following pre- and post-survey items. For each item, teachers were asked, "How much do you agree with the following statements?" and were given a scale from 1 ("not at all") to 6 ("definitely"):

- [measured in the pre- and post-survey] "I know how to use content about DL in the classroom."
- [measured in the post-survey] "I know how to use content about AI in the classroom."

In the post-evaluation survey the following items were included to provide additional insights on the aspect of pedagogical content knowledge and potential for DL and AI integration:

- [measured in the post-survey] "After the training, I have gathered enough competences to teach the learned content in class."
- [measured in the post-survey] "I am willing to invest time and effort to incorporate AI into my teaching."

The post-evaluation survey included a series of questions to assess participants' reactions to the topics and materials/exercises covered in the training:

- [measured in the post-survey] “How suitable did you find the choice of the topics on introducing DL and AI for your teaching?” (1-not suitable at all – 6 very well suited):
    - Technological Perspective: the Functioning o Large Language Models
    - Sociocultural perspective: Risks of Language Models in Everyday Life
    - Introducing to handling data
    - Introduction to AI
- [measured in the post-survey] How suitable did you find the practical examples from the workshop for your teaching? (1-not suitable at all – 6 very well suited)
    - Grimm’s New Fairy Tales (unplugged)
    - Grimm’s New Fairy Tales (Jupyter/plugged)
    - Risks of LLM in everyday life through case studies
    - Data mining – Orange3
    - Teachable Machine: AI

For the hypothesis 2, were asked, "How much do you agree with the following statements?" and were given a scale from 1 ("not at all") to 6 ("definitely")

- [measured in pre- and post-survey] "I believe that the integration and teaching of DL/AI in the [respective subject] is very important” (1-not suitable at all – 6 very well suited):

The web-based surveys were programmed in QUAMP survey software (versions 4.4.4-4.4.5). Table 3 provides an overview of the response rates for the pre-, post-, and the follow-up versions of the evaluation survey for each training sessions:

| Training | Number of participants | Number of participants completed the survey (%) | | |
|---|---|---|---|---|
| | | Pre | Post | Both |
| 09.03.2023, CS Berlin | 24 | 22 | 17 | |
| 17.05.2023, STEAM Graz | 8 | 8 | 8 | 8 |
| 26.05.2023, STEAM Vilnius | 12 | 12 | 12 | 12 |

**Table 3** Response rates for the evaluation survey for each of the trainings: number and % of participants, who completed the test prior to the training, after the training, and both the pre- and post-versions.

### The DL and AI self-assessment and knowledge test:

The AI and DL self-assessment and knowledge test includes 10 knowledge questions for CS students and 5 knowledge questions for STEAM students. Each question in the objective knowledge test carried a potential score ranging from 0 to 1. Participants had the opportunity to select multiple answers, with scores being deducted for selecting incorrect options. This test was administered before and directly after the training. Table 4 summarises response rates for the pre- and post-test for each of the trainings.

| Training | Number of participants | Number of participants completed the test (%) | | Number of participants completed both pre- and post-tests |
|---|---|---|---|---|
| | | Pre | Post | |
| 09.03.2023, CS Berlin | 24 | 13 | 18 | 8 |
| 17.05.2023, STEAM Graz | 8 | 8 | 8 | 8 |
| 26.05.2023, STEAM Vilnius | 12 | 10 | 10 | 10 |

**Table 4** Response rates for the AI and DL knowledge test for each of the trainings: number and % of participants, who completed the test prior to the training, after the training, and both the pre- and post-test versions.

## 1.2 Qualitative data: research questions, instruments, and analysis

The research questions for the qualitative part of the evaluation included:

1). *How did participants perceive the training, and what suggestions do they have for enhancing the efficiency and effectiveness of future sessions?*

2). *How has the training influenced the integration of DL and AI into teaching, if at all?*

3). *How do participants evaluate the difficulties of conveying DL and AI concepts to students?*

4). *How can DL and AI be effectively integrated into the classroom, and what potential challenges could hinder this integration?*

The trainings were followed up by the two rounds of qualitative interviews right after the training using a semi-structured interview guide developed by the University of Potsdam was used. The interviews were conducted in person by the evaluators or/and by the instructors with the help of the evaluators.

### The interview guide:

The questions in the interview guide administered right after the training (see Appendix 5) focused on teachers' perception of the respective workshop, experiences in integrating the DL and AI content into their classroom and barriers for such integration. Teachers were also asked about the importance of both topics for teacher training and for framework curricula, as well as their wishes for policymakers. The follow-up interview guide roughly six months after the training (see Appendix 6) is primarily focused on the integration of the training content into the classroom. Table 5 includes an overview of the number of interviewed participants for each training.

| Training | Number of participants | Number of participants who took part in the qualitative interviews (%) |
|---|---|---|
| 09.03.2023, CS Berlin | 24 | 4 |
| 17.05.2023, STEAM Graz | 8 | 1 |
| 26.05.2023, STEAM Vilnius | 12 | 2 |

**Table 5** Number and % of participants, who took part in the qualitative interviews right after the training.

While the qualitative interviews in Germany and Austria were conducted in German, the interviews in Lithuania were conducted in English. All the interviews were transcribed and analysed.

The interviews were analysed with the help of the focused interview analysis approach (Kuckartz and Rädiker 2020). We have used both deductive and inductive coding. While the former codes were developed based on the interview guide and applied to all the interviews, within them, an inductive code captured new information that emerged directly from the data. For all personal interviews, we analysed participants' familiarity (consisting of prior knowledge and previous experience) with the topics of

DL and AI using the standardized scale approach developed by Maying (2010), employing a 3-point scale (no familiarity, moderate/average familiarity, high familiarity).

The qualitative results of the analysis in this report are presented according to the structure of the interview guide (see Appendix 5). The primary questions from the interview guide serve as the main themes under which inductively generated categories are presented. The related sub-questions align with the respective sub-themes.

Themes used in the qualitative analysis:

Below is a brief description of both deductive top-level- and sub-themes derived from the research questions and interview guide:

A. Themes for interviews immediately after the training:

- Training: In this category, some fundamental aspects related to the training are clarified. Firstly, we explore the participants' expectations and the factors that led to their participation in the training. Secondly, we delve into their personal perspectives on the difficulty level of the topics. Furthermore, we examine the alignment between the training content and participants' prior familiarity with the concepts of DL and AI, which includes their previous knowledge and experience.
- Teaching DL and AI & difficulties conveying DL and AI concepts: In this category, our focus lies on exploring the integration of DL and AI topics into teaching practices prior and after the training. When it comes to the aspect of integration after the training, our objective is to assess participants' readiness and confidence in effectively incorporating DL and AI into their teaching. Furthermore, our interest extends to evaluating the challenges associated with conveying knowledge about DL and AI to students.
- Establishment and steps to integrate DL and AI: This category is dedicated to exploring the integration of DL and AI topics into teaching, specifically within the context of school classrooms. It encompasses the following key aspects: participants' perspectives on the integration of DL and AI topics within the

framework curriculum; identification of effective steps or strategies for embedding both topics into classroom and school curriculum; teachers' opinions regarding the integration of DL and AI topics in teacher education programs; lastly the exploration of any anticipated barriers that may hinder the successful integration of DL and AI.

- (Possible) changes through the integration of DL and AI in the framework curricula:
  This category centers on potential changes resulting from the incorporation of DL and AI into framework curricula. Specifically, we query teachers about anticipated changes concerning students, school authorities, and other school-related aspects, as well as the broader societal implications.
- Training feedback and potential for improvement: This category primarily focuses on training feedback. Specifically, we focus on participants' feedback on the length, content (topics and exercises), format (in-person event), and participant interaction (such as the balance between frontal and interactive parts). Alongside positive feedback, we are particularly interested in suggestions or criticisms that can be used to inform improvements for future trainings.
- Wishes for education policy: This category describes teachers' wishes or suggestions for education policies concerning the topic of DL and AI in school education.

B. Themes for follow-up interviews:

- Training: This category focuses on how the training is retrospectively perceived with some time elapsed and to what extent (if at all) the participants have benefited from it.
- Integration after the training: The category describes if there has been any integration of DL or AI into teaching CS after the training. In addition, in the case of integration, we are interested in the experiences during the process

and details such as the duration of implementation, grade level, topics covered, etc. In the case of non-integration, we are interested in the reasons behind it or what would facilitate future integration.

- **Training feedback and potential for improvement:** With this category, we identify possible improvements that we could implement in the future. Additionally, we are interested in general suggestions related to the DL and AI in the context of teacher training.

# 2. Quantitative results

## 2.1 Description of the sample

Out of 44 participants, 35 participants took part in the pre-evaluation survey. Figures 2-4 describes socio-demographic data of the participants as reported in the pre-survey for each intervention. Most participants in Berlin reported being between 30 and 59 years old, while most participants in Austria were between 20 and 39 years old. STEAM teachers in Vilnius seem to be on average older than their colleagues in Berlin and GRAZ, with most teachers being between 30 and 59 and no participants in the category under 30 years old. The share of women in the CS training in Berlin constituted ca. 27% (six out of 22 participants). The gender composition in the STEAM training in Graz was balanced with half of participants reporting female and half male gender identity. In Vilnius only two STEAM teacher reported male gender identity.

While 19 participants in Berlin reported being fully trained teachers working in Berlin (one participant reported being employed in Brandenburg), three participants reported being at the final practical stage of their training ("Referendariat"). In the STEAM trainings, all of the participants were in-service teachers (in Steiermark for the Graz training and Vilnius for the Vilnius training). Most CS teachers in Berlin indicated that they teach students ranging from grade seven to grade 12. STEAM teachers in Austria reported teaching the grades five through 12. Interestingly, the STEAM teachers in Lithuania mostly were mostly teaching from grade five to grade 8.

The socio-demographic data of the participants is consistent with the 2022 data on CS teaching in Germany. Schröder et al. (2022) report that depending on the federal state, the percentage of female computer science teachers ranges from 24 to 45%. Based on the 2020 indicators released by the OECD, the share of women for all subjects at the lower and upper secondary level in both Germany and Austria is about 65% (OECD 2023c). The overall share of women teaching at the lower and upper secondary levels in Lithuania, is one of the highest among the OECD countries, at 82.4% and 78.4% respectively (OECD 2023b). In addition, Lithuania stands out for a relatively older average age of teachers compared to other OECD countries (OECD 2023b).

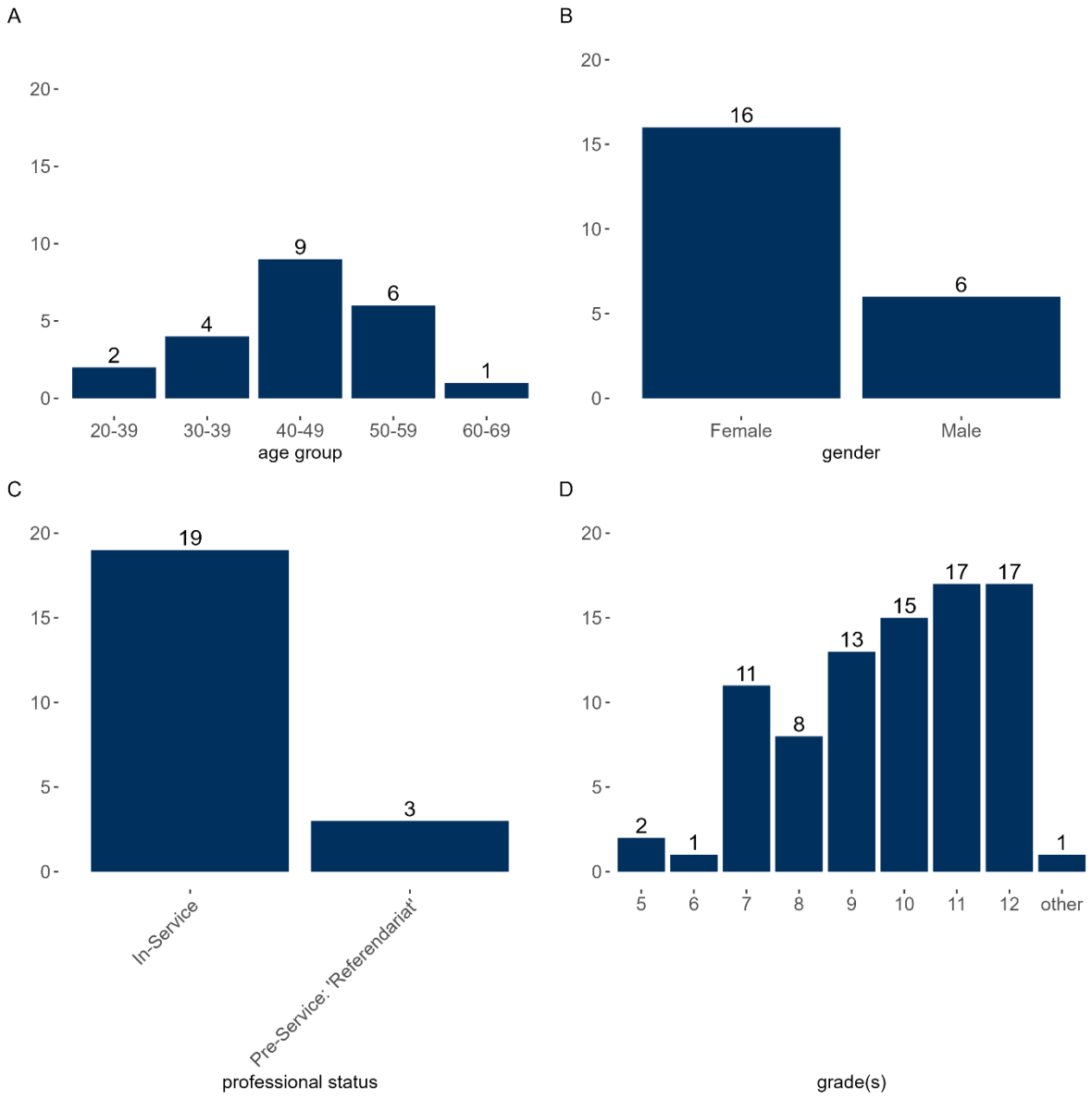



**Figure 2** Socio-demographic data of the participants, 09.03.2023 Berlin, pre-evaluation survey, n=22

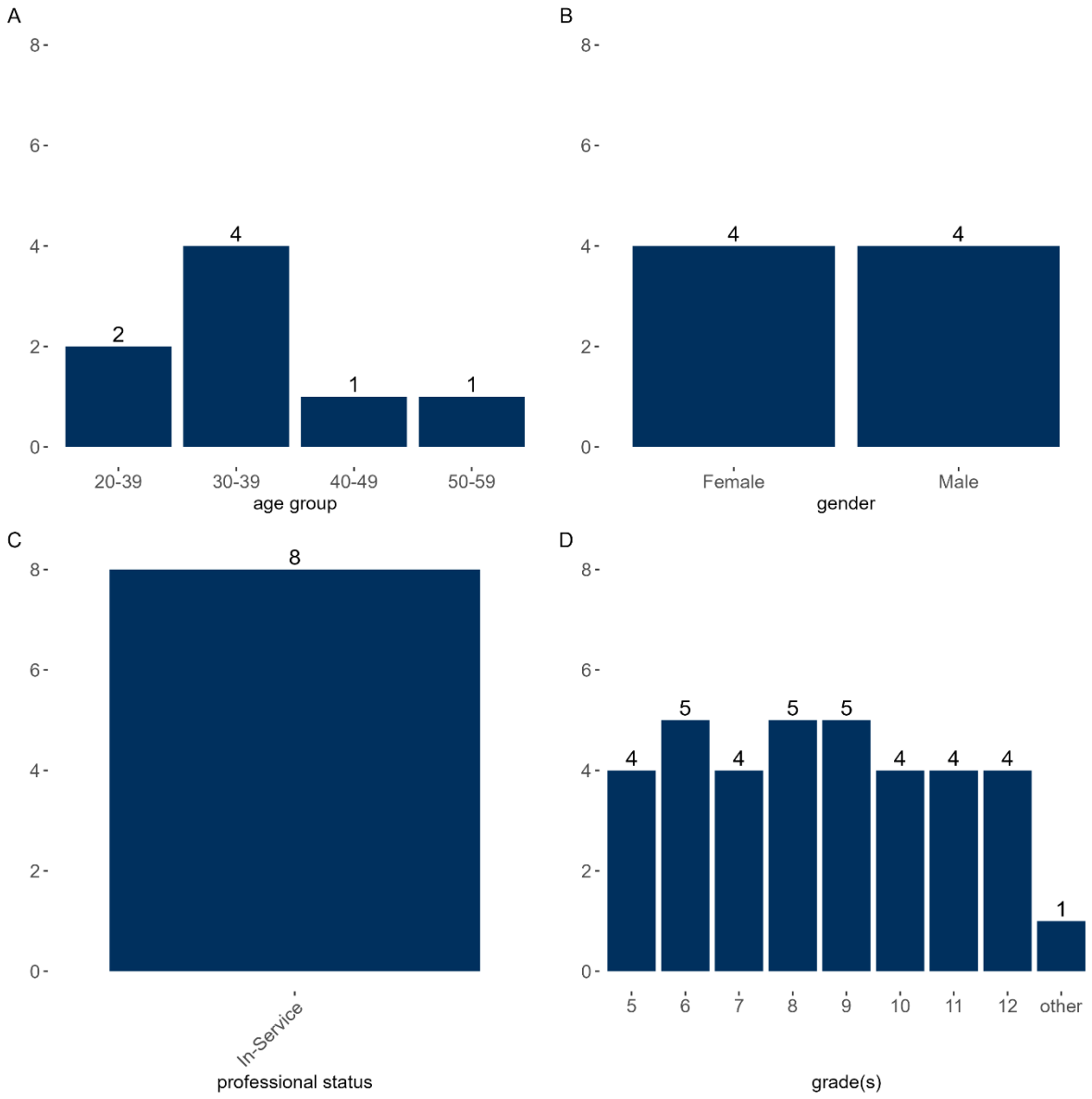



**Figure 3** Socio-demographic data of the participants, 17.05.2023 Graz, pre-evaluation survey, n=8

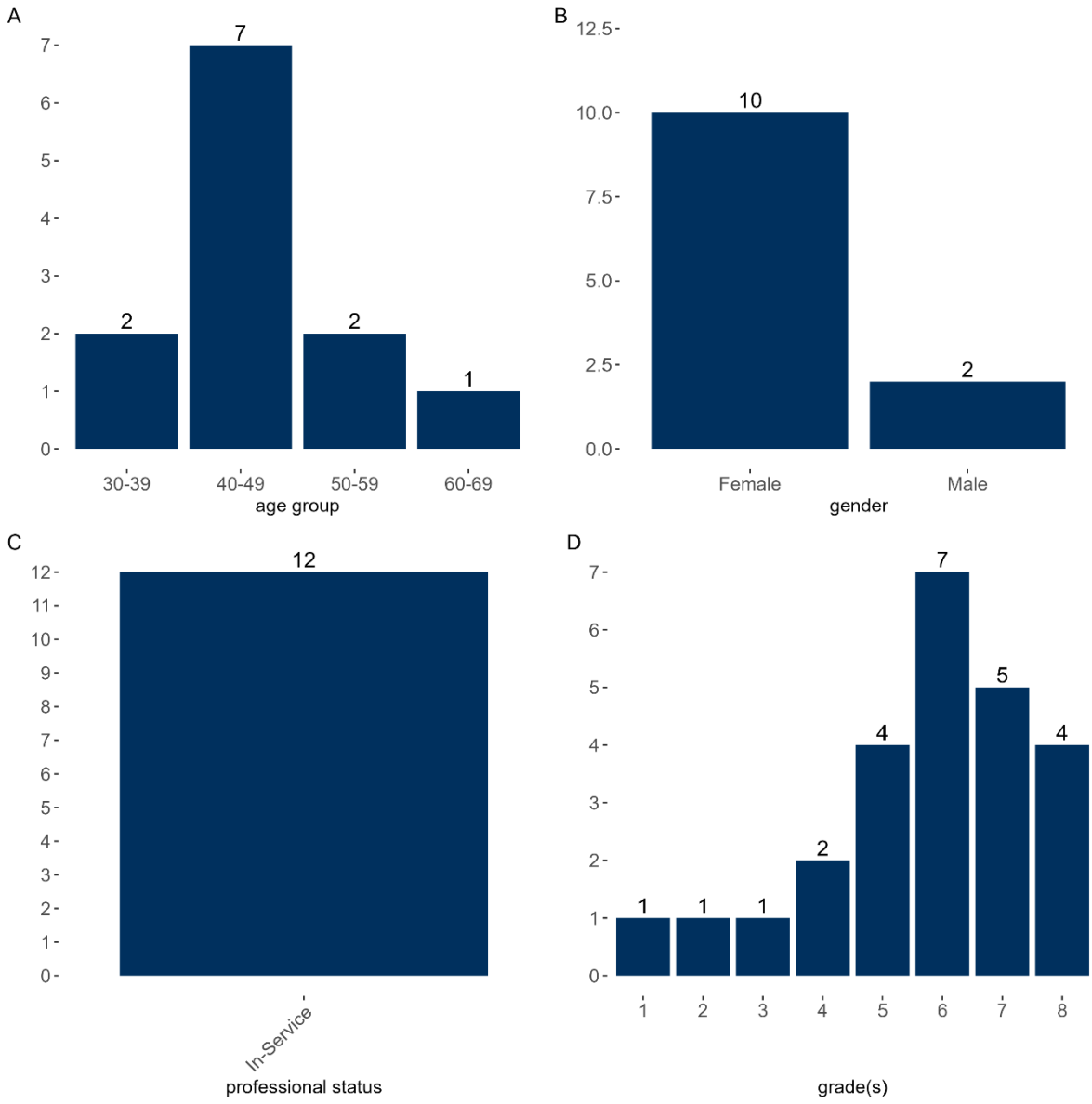



**Figure 4** Socio-demographic data of the participants, 26.05.2023 Vilnius, pre-evaluation survey, n=12

In the CS training in Berlin, 20 out of the 22 respondents reported teaching or studying computer science as a subject, with the following additional/second subjects: mathematics, physics, biology, economics, chemistry, biology, economics, sports, music, and information technology basic education, ethics, philosophy, and geography. According to the pre-survey, mathematics and physics were the most common second subjects among the participants. Two participants were not teaching CS-subject, but instead reported teaching physics, mathematics, ethics, and German language. In the STEAM training in Graz, five out of eight teachers reported teaching CS and or digital

foundations (digitale Grundbildung), so that only one person had no background in CS or digital literacy. Other subjects included: foreign languages, music, and sports. In Vilnius, two teachers reported teaching CS. Other subjects included: technology, literature, history, languages, and sports.

As Figure 5 shows, participants in the three interventions differed in the level of their experience with DL and AI. CS teachers in the Berlin training reported having more experience in teaching DL compared to AI. STEAM teachers in Lithuania with a high ratio of CS teachers reported having more experience with AI compared to DL. STEAM teachers from Lithuania with only two CS teachers among 10 respondents had little experience with both DL and AI.

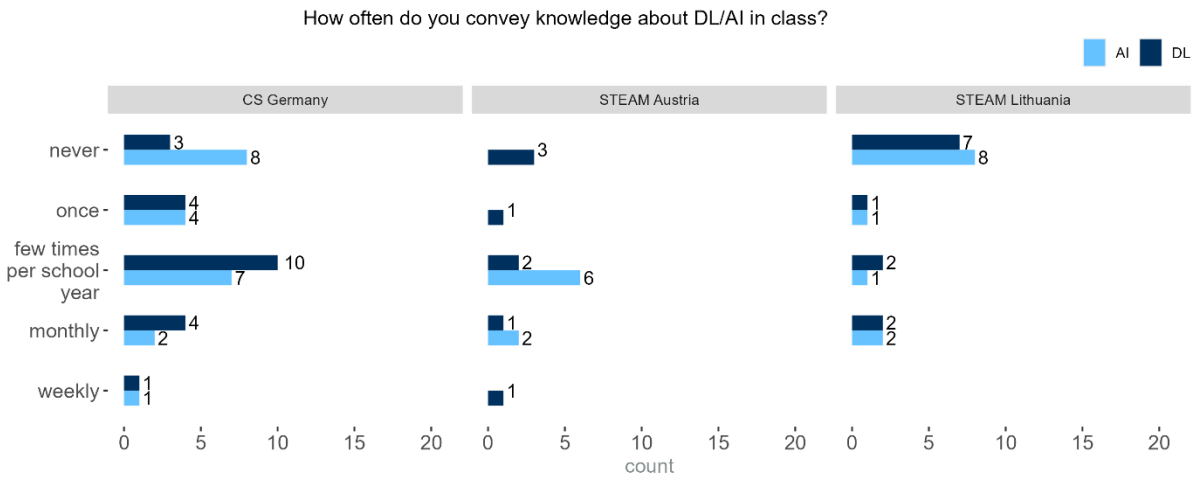



**Figure 5** Experience with DL/AI, Berlin/Graz/Vilnius, pre-evaluation survey, n=42

## 2.2 Feedback on the learned content and format

The post-survey included a series of questions to assess participants' reactions to the suitability of the topics, practical materials and exercises presented in the training. Figure 6, 9, and 12 show that participants reported a very high level of perceived suitability of the topics. At least 80% of the participants selected category "4" or higher on a 6-point scale. The evaluation of the exercises corresponding to those topics varied more, which indicated various degrees of success in the designing and implementing exercises. In the Berlin CS training, the LLM analogue exercise using texts of fairy tails was reported the highest (88% of the respondents selected category

"4" or higher on a 6-point scale). The digital version of this exercises as well as case studies on ethics were rated lower, which might be due to the fact that these exercises did not receive much time within the training. In Graz, all exercises and the lecture were evaluated as being quite suitable for teaching, although Orange3 received the lowest rating. In Lithuania, the feedback for the same set of exercises and lecture was somewhat less positive, possibly due to the Lithuanian teachers' lesser experience with CS. Figures 8, 11, 14 show additional items for teachers' post-training feedback, including training length and format. With respect to the length of the training, participants tend to agree that the training should have lasted longer. At least 80% of the respondents selected the score of "4" or higher.

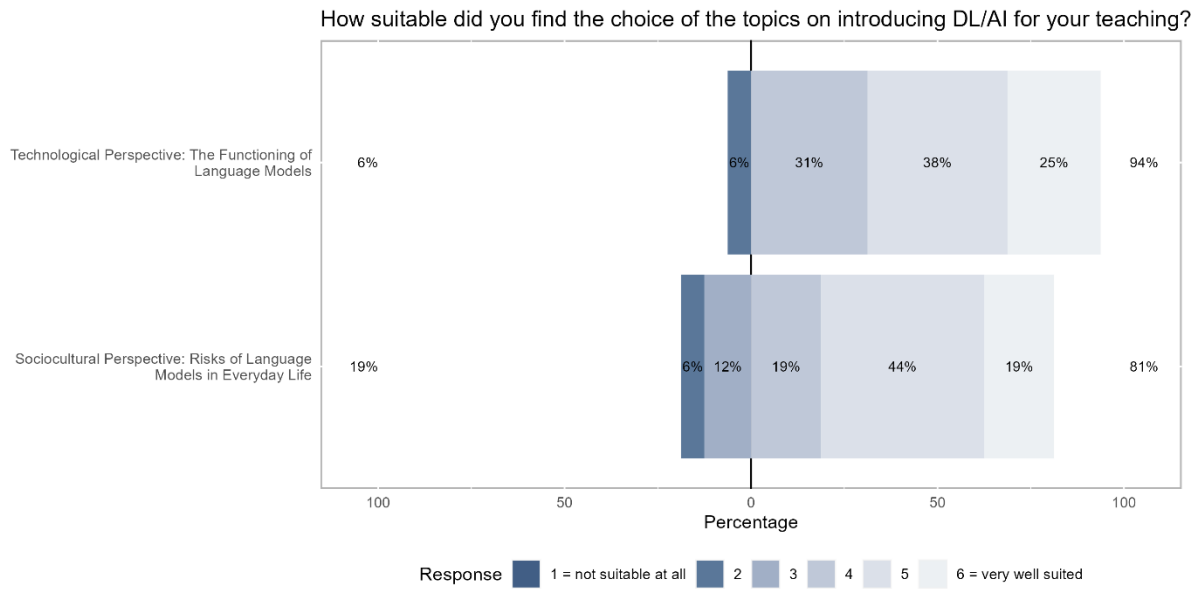



**Figure 6** Distribution of the post results for the assessment of the chosen themes, 09.03.2023 Berlin, pre-evaluation survey, n=22

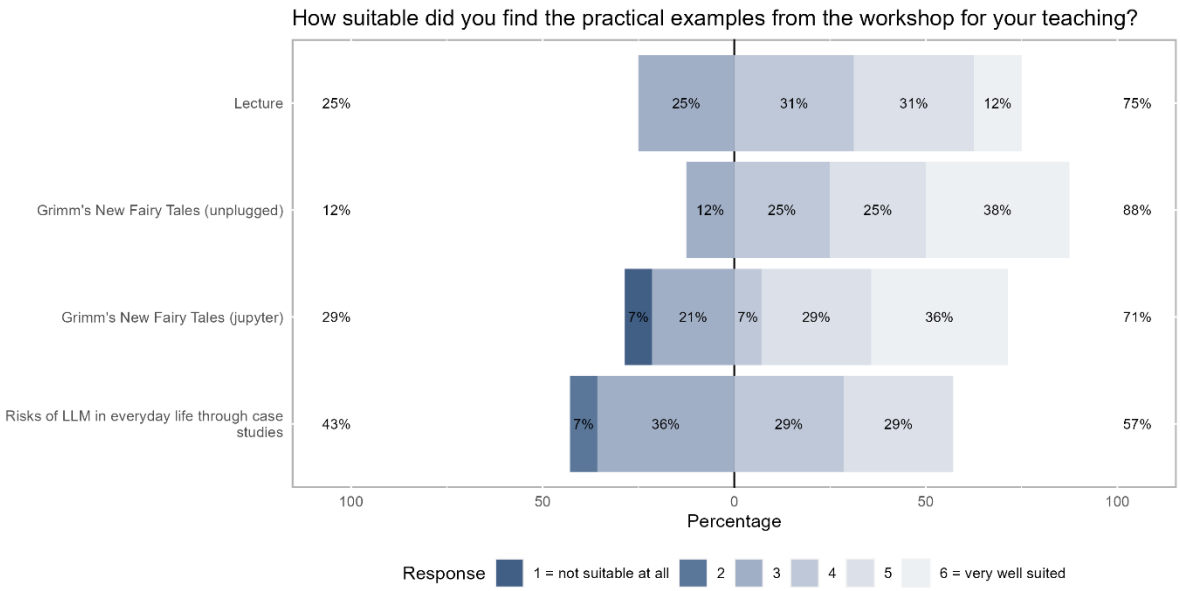



**Figure 7** Distribution of the post results for the assessment of the training material/exercises, 09.03.2023 Berlin, pre-evaluation survey, n=22

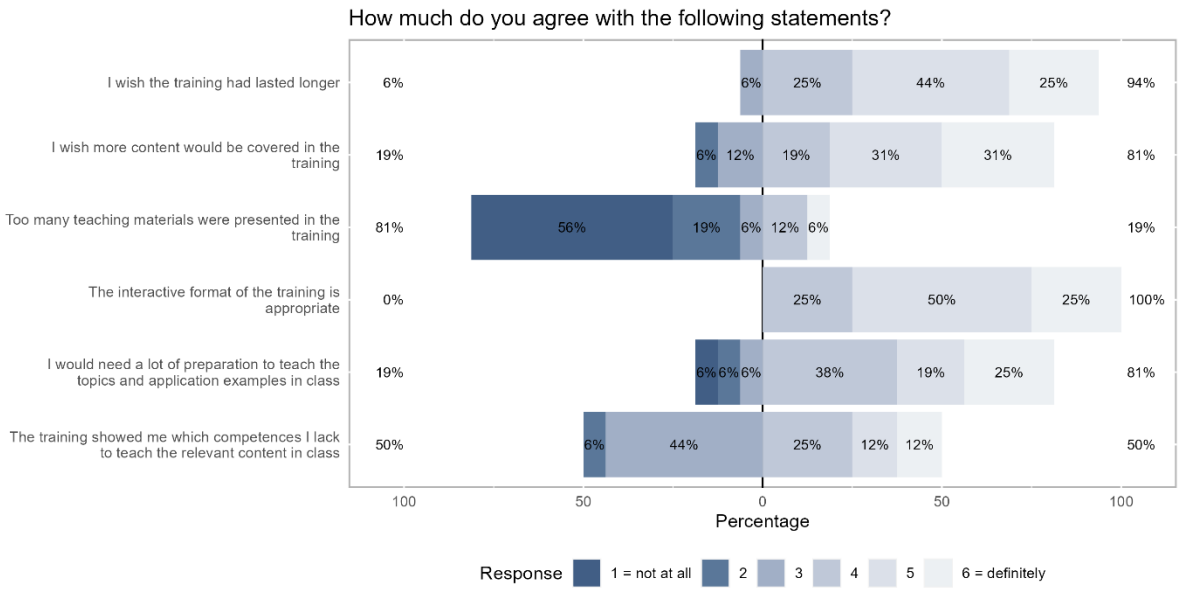



**Figure 8** Distribution of the post results for the assessment of the training format and outcomes, 09.03.2023 Berlin, pre-evaluation survey, n=22

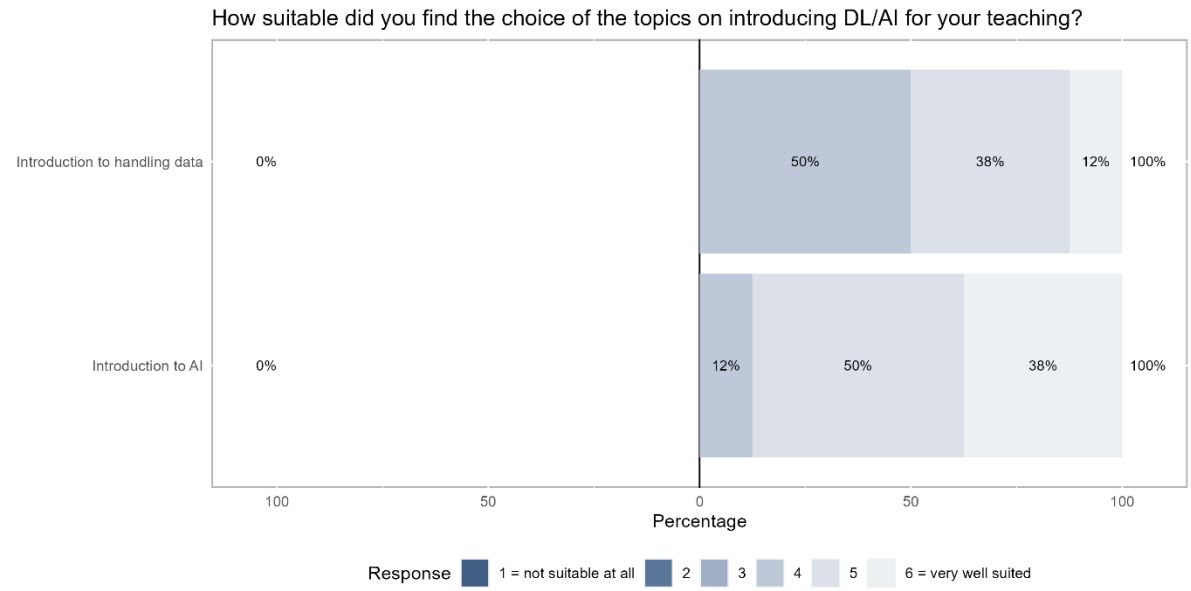



**Figure 9** Distribution of the post results for the assessment of the chosen themes, 17.05.2023 Graz, pre-evaluation survey, n=8

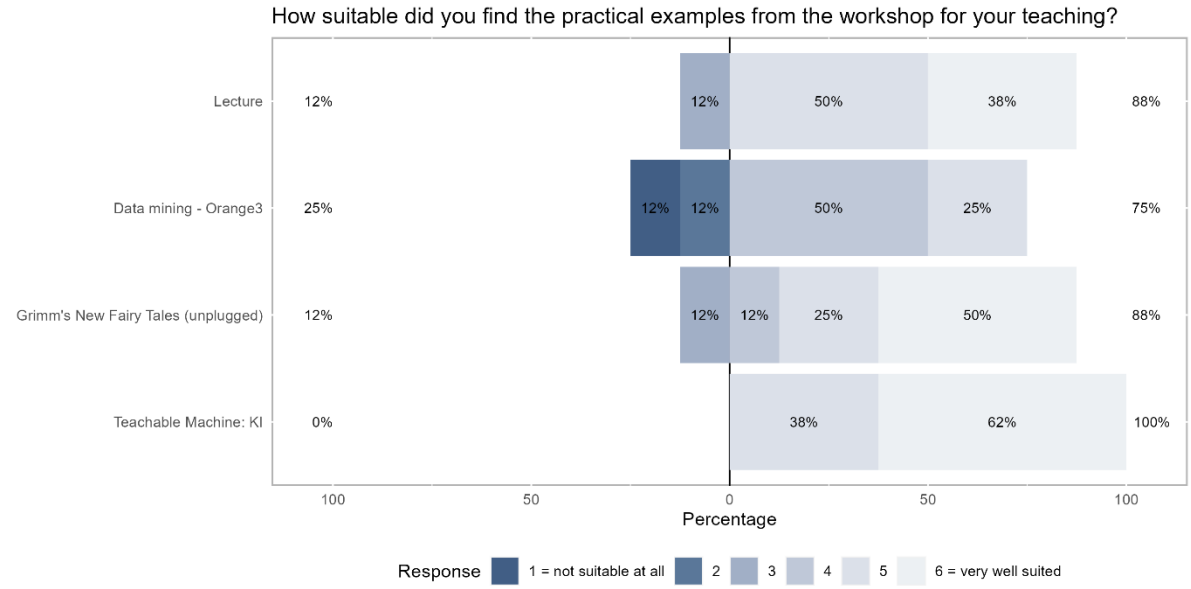



**Figure 10** Distribution of the post results for the assessment of the training material/exercises, 17.05.2023 Graz, pre-evaluation survey, n=8

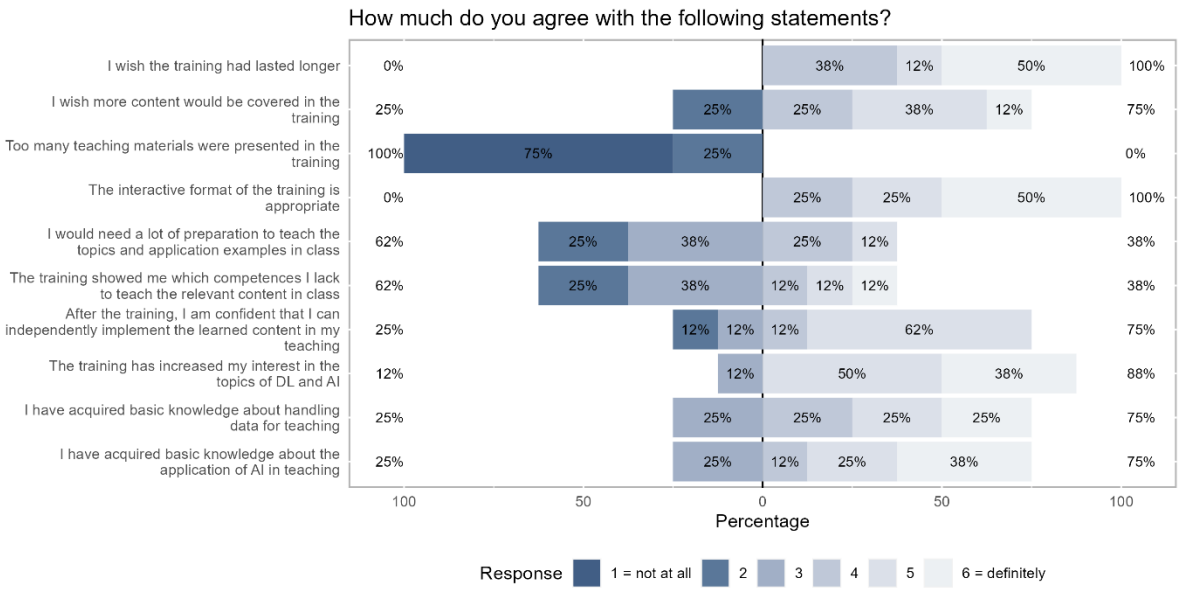



**Figure 11** Distribution of the post results for the assessment of the training format and outcomes, 17.05.2023 Graz, pre-evaluation survey, n=8

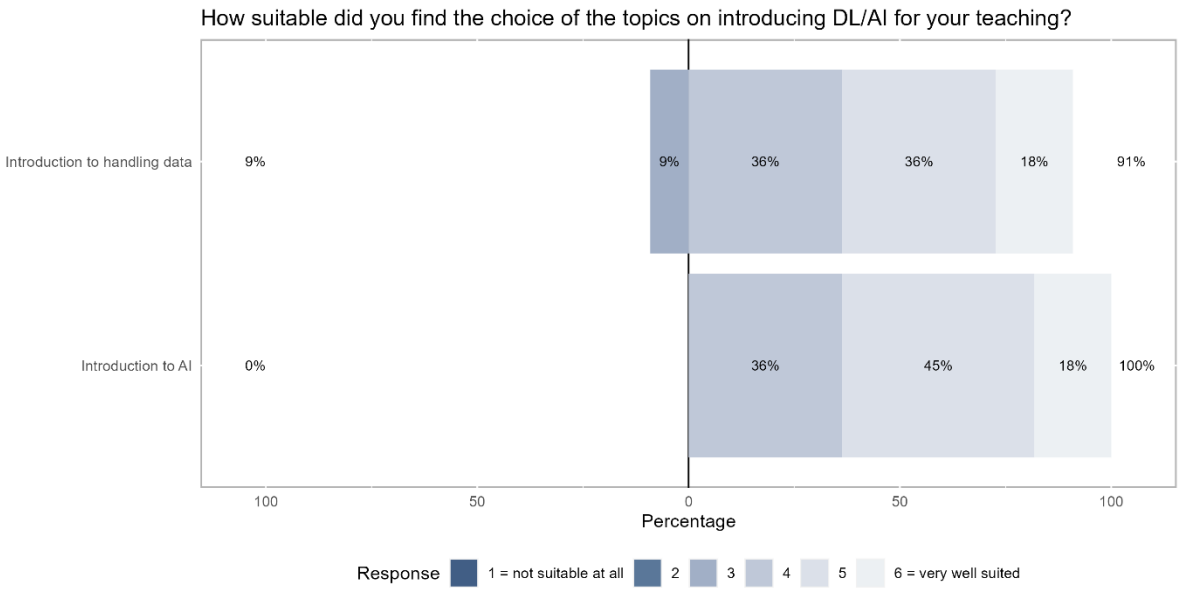



**Figure 12** Distribution of the post results for the assessment of the chosen themes, 26.05.2023 Vilnius, pre-evaluation survey, n=12

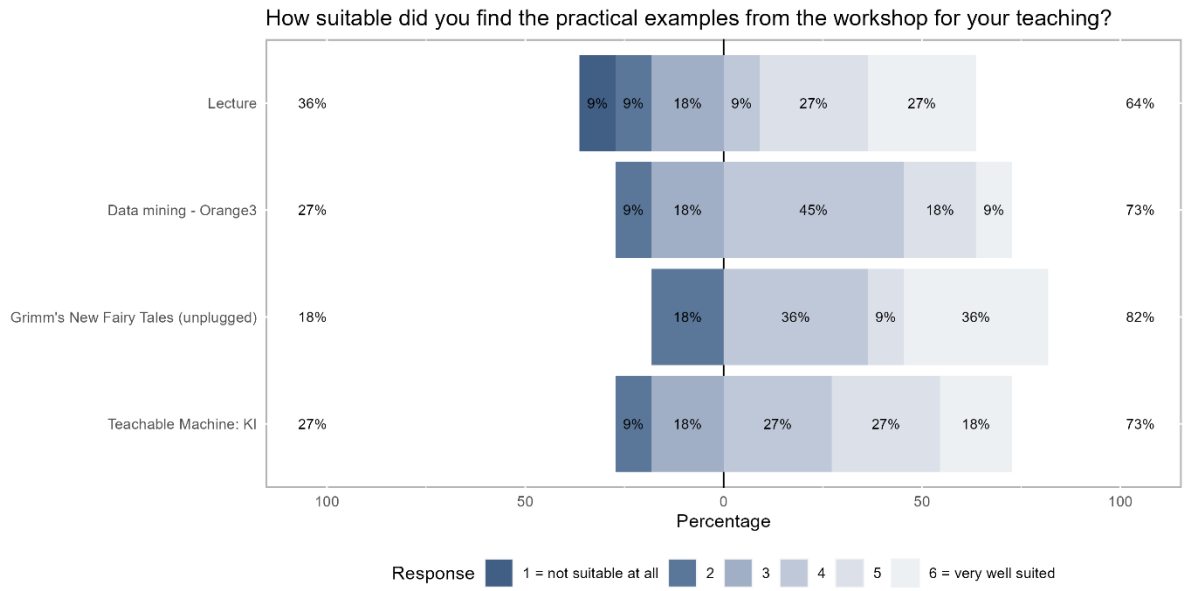



**Figure 13** Distribution of the post results for the assessment of the training material/exercises, 26.05.2023 Vilnius, pre-evaluation survey, n=12

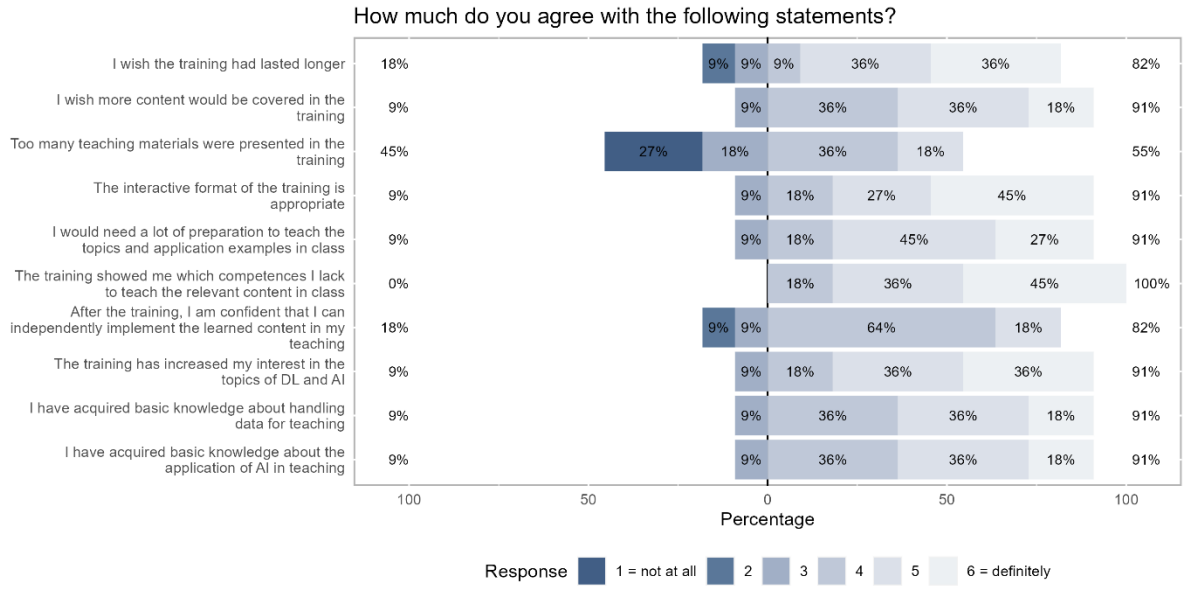



**Figure 14** Distribution of the post results for the assessment of the training format and outcomes, 26.05.2023 Vilnius, pre-evaluation survey, n=12

## 2.3 Perceived competences on how to use DL/AI in class

As demonstrated in Figure 15, respondents reported an increase in the average self-reported level of competences in using DL and AI content through all the trainings. Notably, the interquartile ranges (IQRs) – the boxes representing the middle 50% of responses – appear narrower for the post-training data, particularly for the AI content. This narrowing suggests a decrease in variability among participants' responses after the training, indicating a more consistent level of perceived competence across the respondents. It's noteworthy that the pre-training self-assessed competence level for CS teachers in Germany is lower than that of non-CS teachers, which may not be intuitive given the technical nature of CS education. The latter could suggest potential issues with the scaling or interpretation of the survey (e.g., understanding of the item by participants with different depth of knowledge).

While the Wilcoxon signed-rank test showed a statistically significant difference between the pre- and post-results for the DL item for the Berlin intervention, it was not significant for the AI item. Given a very small sample size for the other two trainings, we do not discuss the statistical significance for the Austria and Lithuania data.

| Question | Median | | Q1 | | Q3 | | Min. | | Max. | |
|---|---|---|---|---|---|---|---|---|---|---|
| | Pre | Post | Pre | Post | Pre | Post | Pre | Post | Pre | Post |
| CS Berlin: I know how to use content about DL in the classroom | 2.0 | 4.0 | 1.0 | 3.0 | 2.8 | 4.0 | 1.0 | 1.0 | 5.0 | 6.0 |
| STEAM Graz: I know how to use content about DL in the classroom | 3.0 | 4.5 | 1.5 | 5.0 | 4.0 | 5.0 | 1.0 | 3.0 | 5.0 | 5.0 |
| STEAM Vilnius: I know how to use content about DL in the classroom | 3.5 | 4.0 | 2.0 | 4.0 | 4.0 | 5.0 | 1.0 | 3.0 | 5.0 | 5.0 |
| CS Berlin: I know how to use content about AI in the classroom | 3.0 | 4.0 | 2.0 | 3.75 | 4.0 | 5.0 | 1.0 | 2.0 | 6.0 | 5.0 |
| STEAM Graz: I know how to use content about AI in the classroom | 3.5 | 5.0 | 2.8 | 5.3 | 4.3 | 4.0 | 2.0 | 3.0 | 5.0 | 6.0 |

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| STEAM Vilnius: I know how to use content about AI in the classroom | 3.0 | 4.0 | 1.8 | 4.0 | 4.0 | 4.5 | 1.0 | 3.0 | 4.0 | 5.0 |
| CS Berlin I am willing to invest time and effort to incorporate DL into my teaching | - | 5.0 | - | 4.0 | - | 5.0 | - | 2.0 | - | 6.0 |
| STEAM Graz: I am willing to invest time and effort to incorporate DL into my teaching | - | 5.0 | - | 4.8 | - | 5.25 | - | 4.0 | - | 6.0 |
| STEAM Vilnius: I am willing to invest time and effort to incorporate DL into my teaching | - | 5.0 | - | 4.0 | - | 5.0 | - | 3.0 | - | 6.0 |
| CS Berlin: I am willing to invest time and effort to incorporate AI into my teaching | - | 4.0 | - | 5.0 | - | 5.3 | - | 2.0 | - | 6.0 |
| STEAM Graz: I am willing to invest time and effort to incorporate AI into my teaching | - | 5.0 | - | 5.0 | - | 5.3 | - | 4.0 | - | 6.0 |
| STEAM Vilnius: I am willing to invest time and effort to incorporate AI into my teaching | - | 5.0 | - | 4.0 | - | 5.0 | - | 3.0 | - | 6.0 |
| CS Berlin: After the training I have gathered enough competences to teach the learned content in class | - | 4.0 | - | 3.8 | - | 4.0 | - | 2.0 | - | 6.0 |
| STEAM Graz: After the training I have gathered enough competences to teach the learned content in class | - | 5.0 | - | 3.0 | - | 5.0 | - | 2.0 | - | 5.0 |
| STEAM Vilnius: After the training I have gathered enough competences to teach the learned content in class | - | 4.0 | - | 4.0 | - | 4.0 | - | 2.0 | - | 5.0 |

**Table 6** Summary statistics of pre- and post-results for the survey items on teachers' perceived competences to use DL and AI content in class and post-survey results for additional items, Berlin/Graz/Vilnius, pre- and post-evaluation survey, n=42

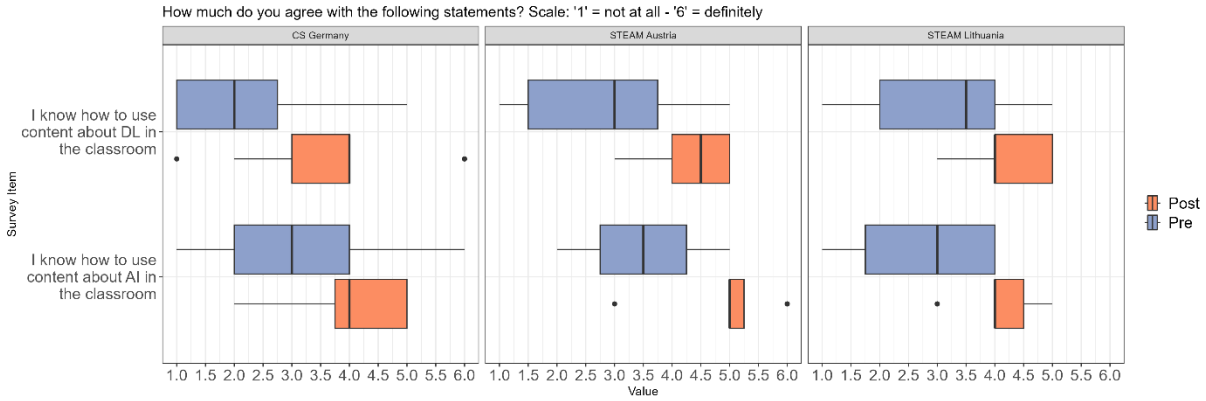



**Figure 15** Boxplot comparison of pre- and post-results for the survey items on teachers' perceived competences to use DL and AI content in class, Berlin/Graz/Vilnius, pre- and post-evaluation survey, n=42

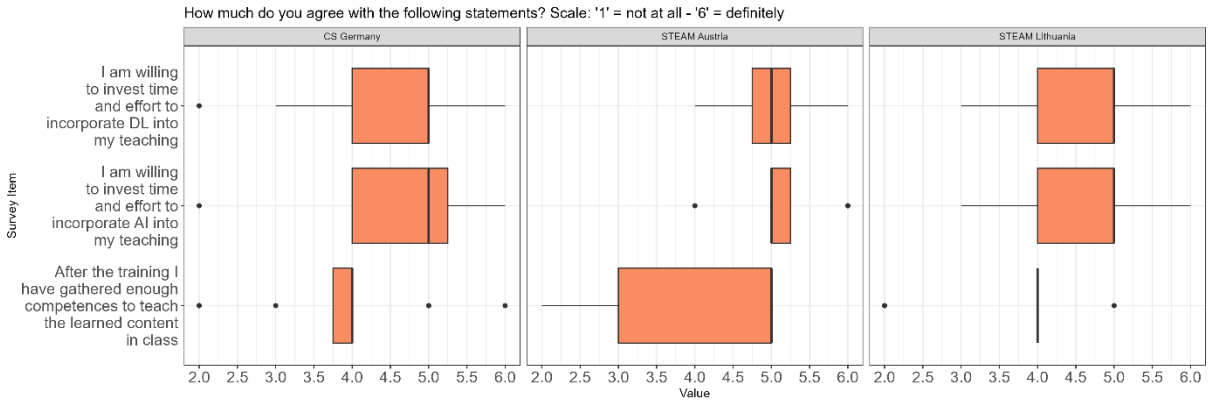



**Figure 16** Boxplot of post results for the survey items on teachers' perceived competences and willingness to invest time and effort to incorporate DL and AI into their teaching, Berlin/Graz/Vilnius, post-evaluation survey, n=35

Figure 16 describe further survey items that were measured only in the post-version of the survey. Participants were asked to what extent they agree with the statements: "I am willing to invest time and effort to incorporate DL/AI into my teaching." Results suggest that on average, participants expressed a strong willingness to invest time and effort to incorporate both DL and AI into their teaching. After the training, teachers were also asked to what extent they agree with the statement: "After the training, I have gathered enough competences to teach the learned content in class". Overall, the statistics for this post-survey item indicate a slight leaning towards agreement that participants have gathered enough competences with the highest spread for the training in Graz. However, for each item there are outliers, indicating that a few respondents felt differently from the majority.

## 2.4 Understanding of DL/AI concepts introduced in the training

Table 7 provides summary statistics of pre- and post-training results for the knowledge tests on DL and AI concepts introduced during the training sessions. For all the trainings, the median scores improved and the interquartile range reduced. These statistics suggest an improvement in the understanding of DL/AI concepts post-training, with all medians increasing. Yet, as no participant achieved the maximum score on the knowledge test, it might suggest a very challenging set of questions or room for improvement in the learnt content.

| Question | Median | | Q1 | | Q3 | | Min. | | Max. | |
|---|---|---|---|---|---|---|---|---|---|---|
| | Pre | Post | Pre | Post | Pre | Post | Pre | Post | Pre | Post |
| CS Berlin: Knowledge test DL/AI, n=8 (10 questions, max 10.0) | 5.6 | 6.9 | 4.4 | 6.2 | 6.2 | 7.3 | 4.0 | 5.2 | 6.8 | 8.1 |
| STEAM Graz: Knowledge test DL/AI, n=8 (5 questions, max 5.0) | 2.7 | 4.3 | 2.3 | 3.8 | 3.2 | 4.1 | 2.1 | 2.3 | 4.3 | 4.3 |
| STEAM Graz: Knowledge test DL/AI, n=10 (5 questions, max 5) | 1.2 | 2.8 | 1.0 | 2.4 | 1.8 | 3.1 | 0.0 | 1.0 | 3.0 | 3.4 |

**Table 7** Summary statistics of pre- and post-results for the knowledge test, Berlin/Graz/Vilnius, pre- and post-self-assessment and knowledge test, n=26

## 2.5 Attitudes towards DL/AI

Table 8 shows the distribution for the survey items on teachers' perception of the importance of DL and AI, which were measured via both pre- and post-policy experimentation surveys. For both DL and AI items, participants express a tendency to believe that DL/AI content is lacking in the current CS framework curriculum. Similarly, respondents reported a relatively high score for the pre- and post-results of the items: "In the future, teaching DL/AI will provide added value to students." However, for both items, respondents reported a high score for the pre- as well as post-results, suggesting a high level of perceived importance of DL and AI for teaching already

prior to the trainings. The sample size was too low to compare the perceived significance of DL and AI for non-CS teachers compared to the CS teachers. Participants reported the following subjects (taught by the respective respondents), where they gave a high score to the significance of DL and AI for the subject: mathematics, physics, biology, chemistry, digital education (in Austria).

| Question | Median | | Q1 | | Q3 | | Min. | | Max. | |
|---|---|---|---|---|---|---|---|---|---|---|
| | Pre | Post | Pre | Post | Pre | Post | Pre | Post | Pre | Post |
| CS Berlin: In the future, teaching DL will provide added value to students | 5.0 | 5.0 | 4.5 | 4.0 | 6.0 | 5.5 | 2.0 | 4.0 | 6.0 | 6.0 |
| STEAM Graz: In the future, teaching DL will provide added value to students | 5.0 | 5.0 | 4.8 | 4.8 | 6.0 | 5.3 | 4.0 | 4.0 | 6.0 | 6.0 |
| STEAM Vilnius: In the future, teaching DL will provide added value to students | 4.0 | 5.0 | 3.0 | 5.0 | 5.0 | 5.5 | 3.0 | 3.0 | 6.0 | 6.0 |
| CS Berlin: In the future, teaching AI will provide added value to students | 5.0 | 5.0 | 4.5 | 5.0 | 6.0 | 6.0 | 2.0 | 3.0 | 6.0 | 6.0 |
| STEAM Graz: In the future, teaching AI will provide added value to students | 5.5 | 5.5 | 5.0 | 5.0 | 6.0 | 6.0 | 4.0 | 4.0 | 6.0 | 6.0 |
| STEAM Vilnius: In the future, teaching AI will provide added value to students | 4.0 | 5.0 | 3.8 | 5.0 | 5.0 | 5.0 | 3.0 | 4.0 | 6.0 | 6.0 |
| CS Berlin: I think the content of DL is missing in the current framework curriculum of computer science | 4.0 | 5.0 | 3.5 | 4.0 | 5.5 | 5.0 | 3.0 | 2.0 | 6.0 | 6.0 |
| STEAM Graz: I think the content of DL is missing in the current framework curriculum of computer science | 5.0 | 5.0 | 3.0 | 4.0 | 6.0 | 6.0 | 3.0 | 3.0 | 6.0 | 6.0 |

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| STEAM Vilnius: I think the content of DL is missing in the current framework curriculum of computer science | 4.5 | 5.5 | 4.3 | 5.3 | 4.8 | 5.8 | 4.0 | 5.0 | 5.0 | 6.0 |
| CS Berlin: I think the content of AI is missing in the current framework curriculum of computer science | 5.0 | 4.0 | 4.0 | 3.5 | 6.0 | 5.0 | 3.0 | 1.0 | 6.0 | 6.0 |
| STEAM Graz: I think the content of KI is missing in the current framework curriculum of computer science | 5.0 | 4.0 | 3.0 | 5.0 | 6.0 | 6.0 | 3.0 | 3.0 | 6.0 | 6.0 |
| STEAM Vilnius: I think the content of KI is missing in the current framework curriculum of computer science | 5.5 | 5.5 | 5.3 | 5.3 | 5.8 | 5.8 | 5.0 | 5.0 | 6.0 | 6.0 |

**Table 8** Summary statistics of pre- and post-results for a set of survey items on teachers' perception of importance of DL and AI, Berlin/Graz/Vilnius, pre- and post-evaluation survey, n=42

## 2.6 Summary of the quantitative results

Participants across three interventions varied in their professional backgrounds and showed varying levels of experience with DL and AI. CS teachers in Berlin had more experience with DL than AI, while STEAM teachers in Lithuania with a high share of CS teachers among participants were more experienced in AI. In contrast, STEAM teachers in Lithuania had equally limited experience with both DL and AI.

Post-training surveys revealed a high perceived suitability of the chosen training topics, with over 80% of participants rating them 4 or above on a 6-point scale. However, the evaluations of the exercises corresponding to these topics showed more variation, suggesting different levels of success in exercise design and implementation. The LLM analogue exercise in Berlin received high ratings, while the digital exercises and ethics case studies were rated lower, possibly due to limited time allocation in the training. In Graz, all exercises and lectures were deemed suitable for teaching, with Orange3 receiving the lowest rating. The feedback in Lithuania was slightly less positive, possibly due to the teachers' lower CS experience. Teachers' feedback also suggested a desire for longer training sessions, with 80% scoring the length as 4 or above.

Respondents reported an increase in the average self-reported level of competences in using DL and AI content though all the trainings. Post-training data indicated a narrowing of the interquartile ranges and a more consistent level of competence in using DL and AI content, with high willingness to incorporate these into teaching.

Knowledge tests showed improvement in DL/AI understanding, with higher median scores and a reduced interquartile range post-training. Participants also acknowledged the importance of including DL/AI in the curriculum and its added value for students, indicating a strong perception of the significance of DL and AI in teaching even before the training sessions.

# 3. Qualitative results

## 3.1 Personal interviews

The themes below describe key patters that repeatedly surfaced in the data, which was gathered based on the questions posed during the interview conducted immediately after the training. The labels on the right correspond to categories that were assigned inductively throughout the qualitative analysis process.

### 3.1.1 Training feedback

#### CS (Berlin):

Two out of four interviewees provided information on how the design of the training was tailored to their pre-existing knowledge/experience in the field of DL and AI. The training was well-designed; this is evident in the level of knowledge/experience in AI, which was indicated as average by both persons, as well as in DL, described as average once and non-existent once.

Design of training: Well aligned with prior knowledge / experience level (AI & DL: little / moderate / average, DL: no)

The training was considered good and as *"actually quite well-structured considering that the further education itself is still in development"* (Participant B - 09.03.23, Berlin).

Overall feedback: Positive – Good (general statement)

The (possibility of) practical tasks/exercises were very well received, as were the theoretical inputs which were perceived as visually appealing and professionally presented. Context-wise, the exercise "Grimm's Fairy Tales", as it was deemed very appealing. The content provided a good foundation for conveying knowledge to the students: *"I do think that the trainings now provide a good starting point. It helps me understand better how to begin and what options are available for handling it. Opening the black box, working with the black box, questioning the model construction, or simply using a*

Format - Content: Positive - Theoretical inputs and practical tasks well liked, good foundation as introduction for students, Grimm's Fairy Tales

*model. Creating our own models. These represent the range of different approaches."* (Participant C - 09.03.23, Berlin)

The ethics component received criticism. Planned exercises could not be carried out due to a lack of time or missing technical prerequisites; this concerns a task about GPT-API via Jupyter Notebooks.

Format - Content: Negative: Ethics did not work out, not enough time for planned content

The training was considered good and as *"actually quite well-structured considering that the further education itself is still in development"* (Participant B - 09.03.23, Berlin).

Format - Length: Positive - Good

However, one person perceived the training as somewhat too long, while others found it to be too short. There was a lack of time to complete or try out additional tasks (that were originally planned). The feeling of not receiving enough knowledge was present: *"I don't feel confident enough now to say that I can definitely implement this. I still need a bit more input for that"* (Participant B - 09.03.23, Berlin). The time needed for evaluation before and after the training felt like unused time, so the training itself should have been longer.

Format - Length: Negative: Too long, too short, evaluation time is 'dead time'

The structure and particularly the blend of theory and practice was praised, for example: *"I found the combination of introductory lecture and non-technically supported exercise to be good. Because it also corresponds quite well to the school reality."* (Participant C - 09.03.23, Berlin)

Format - Structure: Positive - Good, Good mixture between theory and exercises

It was unanimously praised that materials were made available to the teachers. The materials were novel approaches to some of the teachers. The materials provided good inspiration but might need to be modified for use in the classroom at times: *"But I found them, so*

Materials: Positive - Good / very good, novel approach / previously unknown, Good inspiration (e.g., for further modifying)

*they were good inspirations. And I can imagine using the ideas for creating my own materials. But, yes, exactly. To use them, they would need to go a bit deeper."* (Participant A - 09.03.23, Berlin)

Materials: Positive - Practical/implementable in classroom, good materials for AI and ethics

It was noted once, that AI ethics, had not been incorporated into the teaching until now: *"And now, I will definitely incorporate two or three things because it's actually really nice. You can also easily incorporate it into assignments [for students], actually."* (Participant A - 09.03.23, Berlin)

Materials - Negative: Too challenging, too basic, insufficient, more structure and didactics

Like already shown, the materials were sometimes considered too basic, but also sometimes too challenging: *"I just need materials that aren't at a university level because they were too complicated"* (Participant A - 09.03.23, Berlin). Additionally, the materials were found to be not extensive enough. Furthermore, it was wished that it should be better structured and more didactically elaborated, so that usability in the classroom is ensured.

Other factors: Interaction/communication: Positive - Pleasant, very interactive

The interaction of the presenters with the participants was perceived as very pleasant. It was praised that the event was highly interactive.

Other factors: Negative - Technical conditions were not adequately established

Due to the participants bringing their own devices, the technical prerequisites were not adequately met, resulting in additional effort. Moreover, for an exercise that was not feasible due to time constraints (GPT-API via Jupyter Notebooks), the brought devices would not have been sufficient.

It would be helpful if the systems/devices on-site were configured to suit the training's needs, allowing for a quick start and reducing the configuration efforts required.

Ideas for further training: General - Sufficient technical conditions should be established

If the training were to last longer, breaks should be included. Expanding the training into a series of sessions would have its advantages and disadvantages: On one hand, it would allow for more in-depth exploration, but on the other hand, scheduling participation for each session could be somewhat challenging.

Ideas for further training: Format - Length: More breaks if longer training, extension to as series of trainings (advantages/disadvantages)

The Large Language Model exercise (LLM) could be created as 'plugged' instead of 'unplugged' on a *"whiteboard-like online tool [...] or in such a graph, graph-drawing tool" [...], so that it is better visualised than having four different sheets with different things written on them"* (Participant D - 09.03.23, Berlin). For individuals who are already knowledgeable, the introductory presentation is not necessarily essential and could be omitted.

Ideas for further training: Format - Content - LLM exercise as plugged, introductory presentation could be removed

The materials should be adapted to various grade levels. Another suggestion is that they should go into more depth. Concrete, ready-to-use materials are preferable.

Ideas for further training: Materials - According to different grade levels, more in-depth, Ready-to-use materials

STEAM:

## Lithuania

One out of two interviewees provided information on how the design of the training was tailored to their pre-existing knowledge/experience in the field of DL and AI. The person had no experience or prior knowledge in DL and considered the training to be of medium

Design of training: Medium difficulty according to prior knowledge / experience level (AI & DL: no)

difficulty: *"It was not very easy, but also it was not impossible to understand what was going on and what was being told, and what task we are doing during the seminar. I can say that it was in the middle."* (Participant B - 09.03.23, Berlin)

Multiple trainings (as a series) should be conducted. It would be possible to structure them by topics or programs/applications. On each training day, there should be both theoretical and practical content, and, optimally, relevant materials should be provided.

Ideas for further training: Format - Length: Extension to weeks / months (as series) / multiple training

This makes it possible for teachers to build more knowledge, for example, by *"go[ing] more deeply with every program and to get more practice with that program"* (Participant A - 26.05.23, Vilnius). The goal is to have the ability to convey relevant topics to the students: *"As I mentioned before, the more a teacher knows, the better they can teach students"* (Participant B - 26.05.23, Vilnius).

Advantages of weeks / months (as series) / multiple training: More knowledge can be disseminated

## Austria

The person had average prior knowledge in both DL and AI. The training was well tailored to these levels of prior knowledge.

Design of training: Well aligned with prior knowledge / experience level (little / moderate / average in DL & AI)

The existence of research projects related to the organisation and conducting of such trainings is deemed to be good.

Overall feedback: Positive - Good

Although the research projects related to the training are perceived as positive, they only become visible to the participants when the evaluation takes place. This could be difficult for the participants to understand, especially since it creates a timing issue: *"So, I'm say-*

Format - Length: Negative - Too short and evaluation time is 'dead time'

*ing it a bit exaggerated, but it's [evaluation time] downtime for training, more like a good time for the project, and downtime for training."* (17.05.23, Graz)

The practical exercises were perceived as a good foundation and *"through one's own actions [in the training], immediately conveyed a sense of how one might possibly do it with students in the future."* (17.05.23, Graz)

Format - Content: Positive - Exercises as good foundation / introduction for students

The materials that are usable in class and provided to the participants were also rated as good.

Materials: Positive - Practical/implementable in classroom

Registration for these trainings should also be possible a few days in advance and not always based on booking months in advance, as it is difficult (for teachers) to plan months ahead. The training sessions should also be advertised more effectively.

Ideas for further training: General - Registration options should be more short-term, more public relations work

One idea is, that in the future, the time for evaluation should be taken into account more in the planning of the duration of the training.

Ideas for further training: Format - Length: Take into account evaluation time

In the future, instead of individual training sessions, it should be considered organizing training series, *"in the sense of basic events and follow-up events, building directly on the age groups of the students [...] or creating them for that purpose."* (17.05.23, Graz)

Ideas for further training: Format - Length: Extension to weeks / months (as series) / multiple trainings (basic and follow-up-events: different school/age levels)

For younger students, one can focus on playful approaches and also practical application of these topics to establish a fundamental understanding of them: *"[...] a bit of the underlying, so that it becomes tangible because, it feels like when you hear about AI again, nobody has any idea. The only thing you know is ChatGPT. But to*

Content for extension to weeks / months (as series) / multiple trainings - Easy & complex content, practical / real-world applications

*convey the basics of how it works, I believe, would be good and meaningful so that it can also be taught in schools."* (17.05.23, Graz) For older students, more complex content and approaches can be introduced.

It is also possible to create *"mini-training series"* (17.05.23, Graz) online for specific topics.

Ideas for further training: Format - Length: Small units (as online trainings)

Entry points and fundamental knowledge on topics like ChatGPT or AI can be conveyed. If the teachers want to delve deeper into the topics, they could do that potentially by addressing the topics in their classroom. Another advantage of online events is that *"a lot of effort is eliminated"* (17.05.23, Graz).

Advantages of small units (as online trainings): First introductions to topics, less effort

### 3.1.2 Teachability and difficulties of teaching DL and AI

CS (Berlin):

Although the topics are perceived as complex, the training has provided ideas on how to convey it: *"[...] I find the approach that has been chosen here good. That is to say: We start with something very simple to explain the basic principles. And then move on with the larger model [...]"* (Participant C - 09.03.23, Berlin). Unlike more complex content, the fundamental concepts of AI can be conveyed to students because this can be done manually with unplugged exercises. AI is *"easier to implement [than DL] in the sense of perhaps a series of lessons or lesson series spread over years"* (Participant B - 09.03.23, Berlin).

Teachability for AI: Feasible (and for basic concepts)

Bridging the gap from teaching the basics to application for bigger contexts (in tasks or exercises) is considered challenging.

Teachability for AI: Challenging / difficult

There is insufficient data obtainable for the incorporation of AI in the subject of biology.

In biology, insufficient data available to teach AI

Tasks should be designed in a way that students *"themselves recognize the problem and not where the problem is presented to them."* This could include, for example, *"how incorrect answers can occur or how copyright is violated"* (Participant B - 09.03.23, Berlin).

Ideas how AI could be taught: Miscellaneous, presentation from training (for 9th grade and above)

The teachability of DL can depend on the specific chosen topic, the other topics covered alongside it in the classroom, as well as the students.

Teachability of DL: Dependent on topics (and other topics in class) and students

One interviewee finds teaching DL more challenging than AI: *"[...] and correspondingly, its treatment in the classroom, I would say, Data Literacy is a very challenging topic. Because it's something that I would perceive as not being able to be covered in a lesson series, but rather it's a cross-cutting competence that should actually run through continuously. It's not enough if I say, 'I'll just do this alongside AI, and then they'll be able to do it.'"* (Participant D - 09.03.23, Berlin) It is also considered difficult to find a good (first) approach to convey DL.

Teachability of DL: Challenging / difficult

In the subject of biology, it is possible to collect data in natural environments (like a sea) using sensors and work with the data gathered. However, to make it relevant for AI, more data would be needed. Another idea regarding data protection in the context of DL is to use case studies from everyday life where issues arise, for example, *"where someone at a car dealership looks up a private phone number to arrange a date or something like that"* (Participant D –

Ideas how DL could be taught: Collect data in the natural environment (biology), Case studies (from everyday life), internet

09.03.23, Berlin). Such stories don't require much data and are considered very stimulating for the students. Furthermore, the topic of DL could be explored in the context of the of the internet.

Teachability of DL & AI: good and challenging

One person, when asked about the teachability of both DL and AI, perceives their own ability to convey the topics and the understanding on the students' part as highly promising. Another interviewee argues that it would be more challenging to familiarize the students with the topics, as they pose a high level of difficulty, stemming from the fact that they are *"indeed topics that are still relatively uncharted territory. And there aren't so many instructional examples readily available"* (Participant C - 09.03.23, Berlin). Another issue concerns the differentiation within the classroom, meaning the difficulty of finding an approach that works for all students, which is hardly realistic.

Teachability (unspecified): Ethical issues

One needs additional contextual knowledge that doesn't come from CS to address topics like the reproduction of stereotypes and discrimination.

Linkage between subject to DL & AI: Mathematics - yes and no, Languages - yes

On the one hand, in mathematics, few connections to DL or AI are seen, except when incorrect information is issued. In language subjects, on the other hand, more potential is perceived. Another opinion is that there is a high potential for application in mathematics, specifically in working with data and statistics.

Linkage between subject to DL & AI: Biology - Did not use in school: analysis of genome data with AI; used in school - work with data (from nature)

It would make sense to learn about the *"human genome or any other genome if you examine them with corresponding artificial intelligences or networks. That is a context that offers itself. But getting access to such data is completely impossible"* (Participant D –

09.03.23, Berlin). In the field of DL there is a connection with biology, as one can work with a lot of data, for example, the students can structure and visualise (potentially self-collected) data that stems, such as data from nature.

STEAM:

## Lithuania

The content of the training could be used in biology and adapted for the usage in mathematics.

Teachability: Feasible (biology) and adaptions (mathematics)

## Austria

It is possible to impart practical relevance and understanding, starting with a playful approach, the younger the students are. As they get older, it can probably become more complex.

Teachability: Feasible for lower grade levels, more complex for higher grade levels

In school tennis lessons, no connection to deep learning or artificial intelligence is recognized, but in professional training/sport, a lot of data is collected that can be used and worked with.

Linkage between school subject to DL & AI: Tennis - differentiated

In mathematics, there is a connection point in the *"data topic, of course, as well as data representation in trees and such concepts"* (17.05.23, Graz). Furthermore, basic education for DL and AI could take place within the framework of probability theory in the upper secondary level.

Linkage between school subject to DL & AI: Mathematics - Connection to data and probability theory (as an idea how to teach DL & AI)

### 3.1.3 Teachers' motivation to learn DL and AI & and sources used to learn about the training

CS (Berlin):

The professor as well as the department head alerted the interviewed person to the training.

Source of information: Personal contacts / colleagues

The interviewees were specifically interested in the topics of DL, AI, or both topics.

Motivation / expectation for participating: Interest in DL, AI or both topics

AI as a topic would have a special significance. For example, *"right now, through ChatGPT, the whole thing has gained a lot of momentum again, so I increasingly see it being used by students and had already had some contact with it. And because I'm not quite sure how to address it in any way, I decided to attend this training"* (Participant B - 09.03.23, Berlin).

Motivation / expectation for participating: AI - Importance of topics

The last quote already refers to another motivation for participating in the training, namely the desire to be able to address the topics in class. It is not only about (didactic) conveyability but also about the teachable subject matter, such as: *"My expectation was to get an understanding of the role that AI and data literacy should play in teaching and what examples one can work with"* (Participant C - 09.03.23, Berlin).

Motivation / expectation for participating: Application / usage in teaching

This is also related to the fact that the participants wanted to acquire knowledge about these topics e.g.: *"Yes, so I had the expectation that I would get to know various current artificial intelligences. And may be to discuss what dangers and things there are that should be directed towards the attention of the students"* (Participant D - 09.03.23, Berlin).

Motivation / expectation for participating: Obtaining (new) knowledge

Two participants also mentioned as motivation that they regularly attend the IBBB conferences (Informatiktag Berlin-Brandenburg).

Motivation / expectation for participating: Regular attendance at IBBB

There was also a case where a participant expressed that he had already worked with data in research and teaching (biology) and had also received questions about AI in teaching. Therefore, there was the wish for professional knowledge transfer through a dedicated training.

Motivation / expectation for participating: AI - Because of previous knowledge / experiences

Another participant *"actually thought that it was about data competence in the sense that students know how to process, handle, and analyse data. And not so much about AI, […] So that was the actual thought when I came. But it was AI."* (Participant A – 09.03.23, Berlin)

Motivation / expectation for participating: Expected to learn about DL not AI

The interviewee is new to the teaching profession and has been working at a school for one year.

Motivation / expectation for participating: (New) Responsibilities

The difficulty of the topic of DL was described as average (with the application in teaching being perceived as more challenging). However, there is also the perception that DL is a difficult topic that is not easily learnable.

Difficulty of the topics for teachers (after training): DL – Average and challenging / difficult

The topic of AI is also described as having an average level of difficulty. After the training, it appears to be less complex than before.

Difficulty of the topics for teachers (after training): AI – Average

On the one hand, colleagues' interest in further training is assessed as (very) high, as specifically mentioned for CS teachers. One

Interest of colleagues in DL & AI training: Positive: High

respondent describes their own interest and that of their CS colleagues as follows: *"Currently, we are [number] computer science teachers, and in [number] other cases, the focus is on addressing AI in terms of 'understanding it, implementing it ourselves, and considering the societal perspective on it."* (Participant B - 09.03.23, Berlin) CS teachers would also recognize the necessity to engage with these topics, unlike *"all colleagues from non-computer science backgrounds. So, I have not encountered anyone who says they're into it and would like to do it [gaining knowledge in these topics, attend trainings]."* (Participant B – 09.03.23, Berlin)

Interest of colleagues in DL & AI training: Neutral: Biology teachers not interested in technical questions/ teaching; for STEAM, LLMs could be of interest

Latter relates to the biology teaching staff, for whom it is estimated that there is no interest in furthering their education on these topics, as there is still a prevalent tendency to work very analogously in biology teaching. Considering the impact of LLMs (like ChatGPT) on many subjects, where students can use this tool, LLMs or relevant training about them might spark interest among STEAM teachers.

STEAM:

## Lithuania

Motivation / expectation for participating: Obtaining (new) knowledge, topics important, application / usage in teaching, (General) interest

The importance of teaching students about the topics was emphasized. Both the acquisition of new information and content that can be applied in the classroom were brought up. Therefore, the two interviewees expressed a general interest in the topics of the training, without specifically singling out DL or AI.

Interest of colleagues in DL & AI training: Positive: Topics could be interesting; Unclear: learning is compulsory

The contents of the training could be potentially interesting: *"I think for the other teachers, those topics are interesting because every teacher wants to renew their subject to generate more interest among students. So they are eager to refresh it. So that's why*

*those topics could be interesting for them."* (Participant B - 26.05.23, Vilnius) Another opinion is that learning is compulsory in order to be able to teach the students step by step.

## Austria

The motivation was to gain new insights in addition to those that were already present, and also to receive *"teaching materials or content relevant to lesson preparation, ideally"* (17.05.23, Graz). The interviewee was interested in both DL and AI.

Motivation / Expectation for participating: Obtaining (new) knowledge, because of previous knowledge / experience, application / usage in teaching, interest in DL & AI

The difficulty is generally manageable, but only up to a certain point, depending on topic depth. Specifically, what is needed for teaching digital education in the lower secondary level, is certainly achievable. However, for advanced topics like model creation and in-depth model evaluation in the upper secondary level, a more in-depth understanding is required.

Difficulty of the topics for the teachers (after training): Dependent on topic depth (from easy to complex)

Regarding the latter, one would need to do thorough research, but with the degree of this self-engagement being a question of *"how far do you go along the path with the students?"* (17.05.23, Graz).

Difficulty of the topics for the teachers (after training): Further engagement after training needed - Dependent on application in classroom

The acceptance of separate in-person training sessions among the teaching staff is viewed as *"difficult"* and *"not so trivial"* (17.05.23, Graz). Instead, suggestions for other training formats are proposed (see chapter 3.1, Training feedback).

Interest of colleagues in DL & AI training: Ambivalent

### 3.1.4 Integration experience before the training and ability to integrate D and AI after the training

CS (Berlin):

All individuals had so far integrated content from DL and/or AI into their teaching (one teacher integrated it not explicitly). First, the topics are mentioned, and then specific attention is paid to the content, the frequency of integration, and the experiences in each case.

In class, the basics of AI were covered, including types of artificial intelligence. Additionally, machine learning was integrated theoretically and practically, in the context of programming an algorithm. There was also work shown/done for GitHub Copilot (LLM - Large Language Models). Language assistants were covered (Machine Learning, possibly LLM). Theoretical aspects of image generation were also addressed. One teacher supervised the independent learning of students who were engaging with topics from AI.

Integration: Topics: AI

One teacher integrated content from DL (this was in the subject of biology). This involved various practical applications from everyday life and from the field, databases, and working with data.

Integration: Topics: DL

Participant A taught fundamental basics of AI in a workshop course (extracurricular activity) for three to four hours in the subject of CS. An algorithm was programmed for a chess program (machine learning). This course was done one time.

Participant A: Integration: AI, subject CS, frequency – once: Fundamental basics of AI, programming an algorithm (machine learning)

According to the teacher's perception, *"it just didn't go well. Because, in a sense, I came to school. That was my first lesson. And there were simply too many discrepancies between expectations and reality. So, to speak, nothing came of it, nothing special"* (Participant A - 09.03.23, Berlin). In retrospect, it was felt that teaching was

Only negative experiences: Not good, because no previous experience; difficulties in explaining / pedagogical difficulties; difficulty / complexity of topics; Money for the course was cut

too mathematical, and it was not easy to find the right level of difficulty, making it hard to explain things optimally to the students. Additionally, this couldn't be improved because the funding for the course was cut, and it no longer continued.

Participant B: Integration: AI, subject CS, frequency – more than once: Image generation, language assistents (machine learning, possibly LLM)

Participant B has already integrated AI multiple times in the subject CS. As part of a lecture series, theoretical aspects of image generation (functionality, opportunities, risks) were addressed as a student presentation. Furthermore, language assistants were discussed (machine learning, possibly LLM). Both topics were covered in the eleventh grade. The lecture series took place during the introductory phase in the basic CS course for students without prior experience.

Experiences: Positive – Students: well received, were motivated

The students embraced the topic(s) well and were motivated. However, *"[...] I don't know if, in the end, they will say they understood how it works"* (Participant B - 09.03.23 - Berlin).

Experiences: Negative - Teacher: Lack of knowledge / reaching limit of knowledge, difficulty / complexity of topics and in explaining

By assigning a presentation (topic of image image generation) to the students, the teacher came to the realization: *"As I mentioned, it wasn't very in-depth. Because I left it quite open. In hindsight, I also concluded: '"Hey, this might be a topic that is not so easy to work on by oneself [by the students]."* (Participant B - 09.03.23, Berlin) The lack of depth is also associated with the complexity of the topic, for which there are, for example, *„many imprecise explanations in numerous sources"* (Participant B - 09.3.23 - Berlin), making it difficult to convey the subject matter.

Participant C: Integration: AI, subject CS, frequency – more than once: Supervising independent learning of students

Participant C did not directly integrate the content into computer science classes, but supervised processes several times, such as *"multiple assignments [...], where students engaged with these top-*

*ics. Especially with the AI topics. There were often presentation exams or special learning achievements, such as seminar papers"* (Participant C - 09.03.23, Berlin).

Experiences: Students – Positive: Well received / understood; Negative – Not well received (when not guided by teachers)

An evolution in understanding has been evident, shifting from questions like *"'How does a neural network function? What is it, actually? What happens inside?' to inquiries such as 'Can I actually trust AI systems? What error rates should I expect?' Now, students come and say, 'I trained a net traffic signs and can make statements based on that'"* (Participant C - 09.03.23 – Berlin). If the students were not guided by the teacher, there was always a lack of insights into content as well as structuring (for example, in presentations).

Participant D: Integration (AI): subject CS, frequency – unclear: GitHub Copilot, Fundamentals of AI

In computer science, in a time frame of one lesson, Participant D demonstrated and also assigned simple programming tasks for GitHub Copilot (LLMs) to the students. Also, different kinds of artificial intelligences were discussed with them.

Experiences (AI): Positive - Content / teaching worked out fine, enjoyable (for teachers and/or students)

When working with GitHub CoPilot, even with simple programming tasks, very interesting solutions quickly come to light. In dealing with students, *"[...] the topic is actually very, very rewarding. Because it is socially super relevant. And it's also fun. You find a lot of problem-oriented approaches. And it's also socially fascinating"* (Participant D - 09.03.23, Berlin).

Integration (DL): subject biology, frequency – more than once: Various (example from everyday life and a field study), working with data, databases

DL was integrated into biology on multiple occasions. With a seventh-grade class, a breakfast was organized on the topic of nutrition for which students collected data which *"was evaluated and visualised using data that was previously created and compiled through research on the respective foods [...]. This immediately revealed results. Afterward, the students had to assess: Who ate the healthiest?"* (Participant D - 09.03.23, Berlin). In another described case, students collected various data at a lake using sensors. For example, *"[...] attempts were made to recognise patterns. Initially, this was*

*done without any specific methods, and then, using simple statistics and graphics to break it down a bit. These aids helped to better visualise the data"* (Participant D - 09.03.23, Berlin).

Experiences (DL): Positive – Successful (breakfast), enjoyable (for teachers and/or students); Negative (field study) - did not work well, difficulty / complexity, databases are for advanced courses ("Leistungskurs") in CS, lack of structured approach in STEAM subjects

The task with the breakfast was very successful. When measuring the data in the lake as a field study, the difficulty arose in pre-structuring the data/databases *"in such a way that the students knew why it was structured that way"* (Participant D - 09.03.23, Berlin). In this way, working with databases is simpler in CS, as for example in biology, as one must be clear about the structure of the database and it also needs to be normalised, which corresponds to the level of an advanced course ("Leistungskurs") in CS. This is something lacking in STEAM subjects: structured work, for example, like *"filling out a table in Excel or LibreOffice properly and distinguishing between, for example, long or short data formats, and transforming them if necessary. This is something you only learn, if at all, at university"* (Participant D - 09.03.23, Berlin). The concept of data is also not clear in STEAM subjects. That's why the field study did not work very well. But overall, students enjoy working with data.

Experiences (DL & AI): Negative – Lack of infrastructure / poor infrastructure

Participant D also describes a problem that applies to both AI and DL: There is a lack of technical infrastructure (for example, missing graphics cards or certificates to push or download things on GitHub) or it is deficient (packages need to be installed anew each time for each lesson): *"This sometimes makes it very strenuous for the students."* (Participant D - 09.03.23, Berlin) The problems with the infrastructure also take away time from the lessons.

Now follow mentioned reasons for not integrating content.

No Integration - DL: Lack of time, other priorities; AI: in biology, no / insufficient data for AI available; Unspecified: Because of

DL was not integrated because there was no time in the class for it, respectively, other priorities/topics existed that the framework curriculum prescribes. AI-exercises in the subject of biology could

not be conducted because there were not enough data available. In the subject of mathematics, the application context does not fit well for a central engagement in the class, *"because you can only address misinformation [...]. That means you can only somehow incorporate it on the side, right?"* (Participant B - 09.03.23, Berlin)

no link from subject to topic (mathematics)

The following statements address the extent to which the respondents assess their ability to (better) integrate the content into their teaching after the training.

Although it's possible to teach something with data and statistics in the math subject, there is no time due to other priorities in the curriculum (see above under 'No integration').

Ability to integrate (after training): Not (instantly) capable – DL: In STEAM subject no time/priority, focus on framework curriculum

There is the opinion that content on AI ethics could be immediately incorporated, for example, it *"wasn't the topic I had included in my teaching until now. And now I definitely plan to include two or three things"* (Participant A - 09.03.23, Berlin). This could be achievable in the eighth grade or perhaps even earlier. Content on text generation (LLM) would also be immediately integrable.

Ability to integrate (after training): Instantly capable – AI: Ethics of AI, LLM

There are also statements regarding the instant non-integrability of AI. In general, *"it's not like, if you come from one or two workshops, that you can immediately create a complete school instructional unit. It still requires preparation and training"* (Participant C - 09.03.23). Technical issues, for example, need to be addressed, and material must be prepared or modified. Furthermore, the material from the training was not sufficient; it needs to be more tailored to the students and better prepared from a didactic standpoint. The

Ability to integrate (after training): Not (instantly) capable – AI: Training as a good entry point, only basic / insufficient knowledge through training, Modify / adapt material, material not sufficient & won't be used in the presented way

materials need to go into more depth; they were good inspirations but couldn't be used in this form for teaching.

A participant feels unprepared in terms of knowledge and would need *"more input beforehand, with more background knowledge. […] partly it also requires background knowledge that doesn't exclusively come from computer science. So especially all the reproductions of stereotypes, discrimination. It will be difficult to address that beforehand in computer science class"* (Participant B - 09.03.23, Berlin).

Ability to integrate (after training): Not (instantly) capable – Unspecified: Only basic / insufficient knowledge through training, Ethics is problematic

STEAM:

## Lithuania

Both interviewees had not integrated content on DL and AI into their teaching before the training. One person with the subject mathematics stated that it did not happen due to a lack of knowledge. The interviewee with the subject technology did not know how to integrate the content into their subject.

No Experience: Integration in class - Lack of knowledge about topics (mathematic), lack of knowledge about integrability (technology subjects)

## Austria

With the students, *"[…] fundamental topics were discussed, but very, very, very, very superficially and briefly. But the future will bring more"* (17.05.23, Graz). It is unclear how often this was done.

Experience: Integration in class (before the training): DL / AI unspecified, subject unclear, frequency – unclear: Fundamental themes

Due to the practical exercises, *"where you can take away a lot, slash, incorporate into teaching"* (17.05.23, Graz), a better integration of DL and AI in the classroom would be possible after the training.

Ability to integrate (after training): Instantly capable - DL & AI

### 3.1.5 Steps to bring DL and AI into the classroom

#### CS (Berlin):

All statements refer to statements about content within the framework curriculum. No issues were addressed that pertain to content outside of the curriculum.

All categories: Within the framework curriculum

The topics should be integrated into the framework curriculum With the topics, *"it turns out that it now permeates so many areas of daily life that, independent of computer science classes, it actually has to come up in some form in all subjects, that one has to deal with it"* (Participant C - 09.03.23, Berlin). Accordingly, the subject of CS must reflect the increasingly digital society in line with its rising relevance. This can also establish a reference to the daily lives of the students, as they actually have *"contact with it every day [...] whether consciously or subconsciously"* (Participant B - 09.03.23, Berlin).

Framework curriculum integration: Pro arguments / important / necessary - reflects daily life / new technologies, topics are a contemporary societal relevant, connection to everyday life of students can be established

That's why It is important that the students *"not only understand these topics but also understand them critically. So, not just accepting, 'Okay, this is how it is,' but also questioning why it is so and what could be done differently?"* (Participant B - 09.03.23, Berlin).

Framework curriculum integration: Pro arguments / important / necessary: Students must understand / handle these topics, their critical / reflective thinking has to be promoted

Furthermore, the curriculum is not up-to-date and needs to be updated to include topics of DL and AI. Additionally, such an update aligns with the media literacy of the students.

Framework curriculum integration: Pro arguments / important / necessary: Because currently curriculum is not up-to-date

Additionally, such an update aligns with the media literacy of the students. The DL / AI topics can be integrated as elective subjects or as optional courses.

Framework curriculum: Details and subjects - As elective components / in elective subjects

AI should be more represented than it is currently, and as much DL as possible should be taught. The topics *"should actually come up in some form in all subjects, independent of computer science classes, so that one deals with them"* (Participant C - 09.03.23, Berlin).

Framework curriculum integration: Details for subjects and grade/school levels - More AI than currently, DL as much as possible

DL and AI can indeed be included in the curriculum, but the abundance of other topics already present in the curriculum should be considered: *"So, there are many topics, and if I had to say which topic would be removed for it, I wouldn't know. So, just in addition.'"* (Participant A - 09.03.23, Berlin)

Framework curriculum integration: Pro or against depending on conditions of framework - Curriculum already overloaded

In order to actually integrate the content, solutions must exist in both the software and hardware domains. For AI, data is also needed to train models, for example.

Framework curriculum integration: Prerequisites - Infrastructure and for AI, data is needed

There would be many possibilities to integrate DL, respectively data into biology (derived from Open Science and Open Data projects in biology and various contexts such as, for example, wild boar sightings). But it is difficult to integrate these themes, because the biology curriculum is well-structured and stringent. The computer science curriculum is much more open.

Framework curriculum integration: Contra arguments - Curriculum in biology is full and less open than CS

There is general agreement expressed to include the topics in teacher training. This can also be supported with details on how this can happen and/or arguments for why it should happen.

Teacher education: Pro arguments / specifics - General agreement

The AI lecture should be a mandatory lecture. In light of the fact that programming jobs are losing significance and, for example, the role of a prompt engineer will soon take a more central role, *"I still need the background knowledge about algorithms and so on, but I*

Teacher education: Pro arguments / specifics - Structure of integration in teacher education (AI), AI compulsory, AI for all teachers, not only CS teachers, AI increasingly important

*also need to know how to interact with these more.“* (Participant C – 09.03.23, Berlin) The optimal scenario for integration of AI into teacher education would be *“[…] a session once a week where you delve into it. In an ideal world, of course, both combined with a subject lecture so that you know how it works. And then a didactics seminar that deals with how to break it down. But whether that's realistic?"* (Participant B – 09.03.23, Berlin) Even non-CS teachers should know how to convey the topic of AI to students (didactically).

In biology, a lot of work is done with data, but as already expressed in the chapter 3.3 (Teachers' motivation to learn), the technical interest of biology teachers, as well as teachers from non-CS backgrounds, is considered low.

Teacher education: Pro arguments / specifics - Work with data is important in biology, but they are not motivated to learn about technology

The interviewees were asked about concrete steps on how to integrate DL and AI into the curriculum.

CS should be established as a mandatory subject, which would also allow for more time to address the topics of DL and AI. AI should be more prominently represented in the framework curriculum or be a mandatory part.

Steps for anchoring in the classroom: Framework curriculum - CS as subject and AI as topic mandatory

The focus should now be more on the 'new' topics: *“You have to, you still want to address the old topics, but now there is this new important topic added, which carries the risk of becoming a larger topic. That means you would have to somehow see, okay, can I maybe take something away from another topic?”* (Participant B – 09.03.23, Berlin) It must be clearly defined which competencies the students should acquire in DL and AI. Only then can sub-areas of the topics be determined and defined.

Steps for anchoring in the classroom: Framework curriculum - Clear focus on and definitions of AI and DL, clear definitions (of sub topics, competences etc.)

These clear definitions in the curriculum framework are the basis that allow the teacher to create teaching concept. Sufficient Materials that can be modified is desired, or preferably *“[…] barrier-free,*

Steps for anchoring in the classroom: Other – Teaching concept, provide sufficient and/or ready-to-go accessible

*accessible material that is free of charge. That is well-differentiated. That is ready to use as is. And, if necessary, also provide the corresponding source codes and installation instructions so that one can quickly get started and doesn't have to prepare much. I believe this is something that every teacher wishes for"* (Participant D – 09.03.23, Berlin). The necessary infrastructure at the school should also be in place.

materials and infrastructure

It is necessary to provide much more fundamental training for the digital requirements in non-CS subjects.

Steps for anchoring in the classroom: Other - In non-CS subjects one needs more than in CS (e.g., knowledge)

If there is no school curriculum for a subject, the content is discussed within the subject department, and the approach is then determined.

When there is no school curriculum

Otherwise, the interviews highlight a set of prerequires for integration into the school curriculum or ensuring that the stipulations in the school curriculum can be implemented.

The required software must be installed on the computers. Furthermore, appropriate teaching materials must be available.

The framework curriculum has to be followed, *"[...] considering what it requires. Then, one should consider at which points and in what form this is actually done in the school, in consultation and using which tools."* (Participant C – 09.03.23, Berlin) Furthermore, such relevant agreements would need to be documented in the school curriculum in writing.

School curriculum: Prerequisites - Installed software on computers, materials
School curriculum: Prerequisites - Alignment with the framework curriculum, written documentation in school curriculum

You need a subject faculty / subject department that wants to integrate the topics into the lessons and discusses them in the subject conference. Furthermore, there needs to be sufficient knowledge

School curriculum: Prerequisites - High commitment / motivation of teachers, convincing teachers to integrate topics, (Equal)

or a similar level of knowledge among the teaching staff, this means *"[...] educating colleagues who may not feel confident about it [the new topics]. But actually, I believe that most would do that, I would say."* (Participant A – 09.03.23, Berlin)

knowledge of teachers at school

Accordingly, among the potential obstacles It must be guaranteed *"[...] that essentially everyone is capable of having understood it to the extent that they can then reduce it didactically and teach it."* (Participant C – 09.03.23, Berlin) However, without appropriate training or the possibility of further education, this is not possible for the teaching staff.

Potential challenges: Teachers - Lack of knowledge or pedagogical implementability, take part in training opportunities / lack thereof

The framework curriculum is full, and if topics like DL and AI are added, its contents would need to be reprioritized.

Potential challenges: Framework curriculum - is full (repriotisation)

Regarding DL, the main obstacle lies in *"[...] providing reasonable datasets that can be used as examples. They should also come with interesting questions that can be explored because I believe handling a dataset is the best way to learn"* (Participant D – 09.03.23, Berlin). Furthermore, the terms DL and AI are not only very abstract and difficult to categorize on their own, but also, beyond that, they are suboptimally distinguishable from each other. Therefore, clearer definitions are needed.

Potential challenges: Other - Datasets are hard to find, DL and AI are abstract terms

STEAM:

## Lithuania

New technologies have already become a part of everyday life. Therefore, teachers need to be informed about these topics, respectively they must be integrated into teaching.

Framework curriculum integration: Pro - Reflects daily life / new technologies

Therefore, these topics have also become a part of the students' everyday lives. That's why it is necessary *"to show children how to use those technologies properly because they are already using them since they can access those technologies on the internet"* (Participant A - 26.05.23, Vilnius).

Framework curriculum integration: Pro -Students gain knowledge, must understand / handle these topics, their critical thinking about them must be promoted

It is beneficial to include these topics in the framework curriculum because in future professional life, certain simple tasks will be handled by artificial intelligence. Hence, it's essential to understand and work with AI to be well-prepared for these conditions.

Framework curriculum integration: Pro - AI will become important for students later occupation

The integration of these topics into teacher education is compulsory for both interviewees. One person argues this way, *"because teachers have to know more so that they would be able to prepare students for the future students"*(Participant A - 26.05.23, Vilnius).

Teacher education: Pro - Topics are compulsory, teachers must have knowledge

For the required steps for integration, it is also generally discussed that teachers should acquire more knowledge (in this case, AI). For AI for example, there is mention about the provision of training measures. For DL and AI, teachers need to be both knowledgeable and skilled in using them so that they can teach students in these topics.

Steps to integrate: Teach the teachers (knowledge gain)

There might be a shortage of seminars for teachers to become familiar with these topics. Also, there could be a lack of necessary resources for integrating topics of DL and AI. A big challenge is that students might not use AI appropriately; they may resort to using it for purposes like cheating or essay writing instead of using it for educational purposes.

Potential challenges: Lack of training opportunities and resources to integrate topics, students use AI for cheating

## Austria

The curriculum framework is already filled with a lot of content: *"So, at times, the volume is so high that it's difficult to add new content."* (17.05.23, Graz)

Framework curriculum integration: Depending on conditions of framework - Curriculum already overloaded

The interviewed person, therefore, advocates for the careful integration of new content, *"so that it doesn't become even more of a struggle, especially when it's already challenging to fit in what should be done."* (17.05.23, Graz)

Framework curriculum integration: Prerequisites: Careful integration

It is advocated to integrate the topics of the training for mathematics at the upper secondary level, as there the practical relevance for the topics is clearer, therefore the reason for addressing these topics is also more comprehensible.

Framework curriculum integration: Details for subjects and grade/school levels: Secondary level

Contents related to the training topics, respectively DL & AI, are partly already integrated into teacher training for the subject digital education. However, it is suggested to focus also on: *"More from an ethical perspective, from a media didactic, media pedagogical perspective, and a bit from the technical perspective."* (17.05.23, Graz)

Teacher education: Partly already integrated, but integrate new perspectives

### 3.1.6 Potential changes due to integration of DL and AI

CS (Berlin):

Students can make more conscious decisions in their everyday lives: *"That they simply know: 'Okay, where do I encounter AI at all, and what impact does AI have on me?' That they not only know: 'Okay, the algorithm suggests some videos or songs to me.' But also,*

Students: they can apply knowledge in everyday life and gain awareness / ability to reflect, regulate encounters with AI through treatment in school

*'What are the consequences of receiving suggestions from AI.'"* (Participant B - 09.03.23, Berlin) Through teaching, this otherwise unregulated encounters with AI can be regulated by informing about risks, enabling students to make more thoughtful decisions and actions. In this sense, students gain the ability to reflect in their thinking and actions, which also includes societal considerations: *"Do we want to let AI make so many decisions when it currently makes some difficult and questionable decisions?"* (Participant B – 09.03.23, Berlin)

This is made possible by imparting knowledge to the students, such as how to work with data.

Students: Gaining factual knowledge / skills

An interviewee states that there is no impact for the school authority.

School authorities: No impact

For the subject of CS, it is conceivable that *"[...] some societal issues that have been somewhat on the periphery could be brought more into focus. I see many connections to facial recognition in various countries worldwide. So, in general, many topics that concern us"* (Participant D - 09.03.23, Berlin). The addressing of these matters on the periphery currently concerns especially later grades of CS classes. Another opinion is that AI will have no impact on society.

Society: AI - Stimulation of societal discussion / discourse about AI, no impact

By incorporating more data and especially DL, biology instruction could be aligned *"much closer to the daily work of a real biologist"* (Participant D - 09.03.23, Berlin) thus orienting it more towards the activities in the corresponding profession, which revolves more around research. Furthermore, for another person, the use of AI for the assessment of student performances would be conceivable.

Other: Subject biology can become more research oriented, assessment of student performance with support of AI

STEAM

### Lithuania

The possible changes are only discussed by the interviewees in the context of students. On one hand, they will gain knowledge. Through the new topics it can be achieved that it *"opens wide capabilities for school children to research and learn"* (Participant A - 26.05.23, Vilnius) so that students can consequently expand their skill set.

Students: Gaining factual knowledge / skills

The integration of AI into subjects that may not be very interesting to students could pique their interest in these subjects, *"because children are using technologies a lot in our days [...]. For example, in chemistry there maybe not very big interest from the students, but if they start to learn using these technologies they may have to get more interest and then in the future, they may have to choose to study those topics"*(Participant B - 26.05.23, Vilnius).

Students: Increase interest in topics

### Austria

No statements regarding this research question can be found in the interview.

## 3.1.7 Wishes for educational policy

CS (Berlin):

Regarding (new) topics to be conveyed, such as AI, *"an open discussion about what else needs to be removed from the curriculum for it would be necessary. But that is a very difficult discussion. [...] Assessing the consequences there is always very difficult. One only*

Framework curriculum: careful re-prioritisation of topics in the curriculum, CS should be mandatory subject

*sees that ten, twenty years later. And that's why I don't envy anyone who has to make that decision"* (Participant C - 09.03.23, Berlin). Furthermore, there is an expressed desire not only to make AI, but to make the subject of CS mandatory.

Furthermore, digitalisation, along with its associated topics such as AI, should be prioritized as a subject in schools, which includes the necessary technical equipment, respectively infrastructure. Additionally, one interviewed person expressed the wish that *"[...] politics should not decide on my teaching"* (Participant A - 09.03.23, Berlin), emphasizing the importance of maximum autonomy as possible for teachers.

Other: Digitalisation should become a priority in school, desired high level of autonomy in teaching

STEAM:

## Lithuania

AI should be integrated into teacher education.

More resources could be provided *"which could be used for the teaching of artificial intelligence not only in computer science classes"* (Participant B - 26.05.23, Vilnius).

Teacher education: Implement AI
Other: More resources for teaching AI in non-CS-subjects

## Austria

The entire digital education should be expanded to all school levels. For example, in the ninth grade (upper secondary level), there are currently two hours of computer science, which should be more of. The mandatory subject of digital education in Austria should not only be retained, but ideally promoted.

Framework curriculum: Strengthen digital education comprehensively

Teacher education: Specialized training so that teachers remain in the profession

For this, it would be necessary to qualify teachers accordingly and to structure the training in a way that prospective teachers do not opt for other professions: *"so that they actually stay in the school and, as experts, further educate the children to such a high level that it makes sense for the future, society, and everything that goes along with the huge topic."*(17.05.23, Graz)

## 3.2 Summary of the qualitative results

The interview participants included both teaching CS and STEAM subjects (specifically math, technology (Lithuania), biology, digital basic education (Austria), and music). This diverse composition of interviewees, along with varied training formats and content, led to a broader range of perspectives compared to the first round, which exclusively involved CS teachers. Although there were overarching themes consistent across interviews from all three countries, specific views were more pronounced within each training group. Notably, the number of interviews conducted in the second round was less than in the first. Despite the diversity in their backgrounds, there was a unanimous recognition of the importance of integrating DL and AI into both teaching practices and the framework curriculum.

While the training was generally praised and served as a good introduction to DL and AI, both training formats (4 hours for STEAM as well as 2 hours and 15 minutes for CS) have increasingly been described as too short by both CS and STEAM teachers. This is also evident in the occasionally arising suggestions to focus on alternative training formats in the future, such as those that aim for a more sustainable knowledge transfer (e.g., via training series) or shorter modular learning (e.g., via online training). As in the first round of interventions, the difficulty lies in the fact that scheduling participation for each session within longer series could be somewhat challenging. Therefore, different and flexible training formats need to be developed to meet the diverse needs of teachers and the given challenges.

The provided materials were positively acknowledged but could benefit from enhancements in content depth, immediate applicability, and better tailoring to different grade levels and students' didactic needs. The inclusion of an ethics component in the CS training, a new addition compared to the first round, received mixed feedback: some praised it, while others saw it as challenging and in need of improvement. The inclusion of ChatGPT (LLM) as a topic boosted teacher motivation to participate in the TrainDL trainings, highlighting its relevance in educational discussions. In Germany, the CS training was seen as beneficial for a better understanding of LLM. However, participants expressed a preference for a digital version of the Grimm fairy tale exercise for more advanced students. The planned use of Jupyter Notebooks for hands-on learning (plugged) couldn't be executed because of time limitations, leading to disappointment among the participants. In Austria, the use of Orange3 for demonstrating decision trees was highlighted. This tool was suggested for integration into the STEAM mathematics curriculum.

In all three countries, the importance of inclusion in the framework curriculum is emphasized. In Germany (CS) and Lithuania (STEAM), the central argument revolves around the (increasing) relevance of the topics in society and everyday life and the associated necessity to prepare students for this circumstance. However, in Germany and Austria, both for CS and non-CS subjects, it is stated that framework curricula are already overloaded and thus in order to integrate new topics, ‘old’ topics likely need to be reprioritised. Similar to the first round of interventions, for the CS intervention in Berlin, it was proposed that CS should become a compulsory subject. CS teachers further suggested, that AI could be introduced at least as an optional component within the CS curriculum. Additionally, there's a consensus among both CS and STEAM educators on the importance of integrating AI into university education programs.

To prepare teachers for future teaching demands, there is a consensus in Lithuania and Germany that teachers, as part of their initial teacher education and/or professional training development opportunities, should enhance their DL and AI

knowledge accordingly. It was mentioned that STEAM teachers might initially require more basic knowledge than CS teachers on topics such as DL and AI for a successful integration into teaching. The reported ability to apply the learned topics post-training varied across the trainings and the participants.

While some see the training content as readily applicable, others currently lack adequate knowledge to effectively implement the content. Furthermore, it may be the case that in STEAM subjects, there is no time for non-curricular content, and the teachers do not see the direct link of DL and AI to their subjects. But the experiences and ideas described by teachers about the inclusion of DL and AI for STEAM subjects and the connections to these subjects show that there is a consensus on the possibility of the integration of DL and AI. Overall, DL and AI currently remain complex topics that are not always easy to convey to students.

In the context of Germany and Austria, the integration of DL and AI is primarily challenged by a shortage of DL and AI expertise among teachers, alongside the issue of lacking or insufficient infrastructure and resources.

The results of the second round highlight differences between the possibilities and constraints for integration of DL and AI into the STEAM and CS subjects. While some interviewees shared their attempts to incorporate DL and AI into STEAM, the interviews revealed significantly greater obstacles in doing so compared to CS. The collated statements highlight crucial factors for successful integration into STEAM subjects: identifying suitable application areas in the subjects, equipping teachers with the necessary knowledge, and providing curricular space for teaching these topics.

## 4. Discussion of key findings across the trainings

The diverse composition of the participants as well as training formats and content coupled with a small sample size in the second round posed significant challenges for result analysis and interpretation. Only 12 out of 20 participants in the STEAM

trainings had no CS background (10 participants were in the STEAM Lithuanian training and 2 were in the STEAM Austrian training), making it difficult to evaluate the difficulty and complexity of DL and AI for STEAM teachers with no CS background. However, the third round introduces more STEAM trainings involving approximately 40 STEAM teachers, providing an opportunity to both broaden and validate the findings. This expansion will enhance our understanding of the training's impact and effectiveness, offering a more robust analysis that accounts for the diversity of participant experiences and training approaches.

Qualitative feedback revealed a general appreciation of the training as a valuable introduction to DL and AI. The quantitative data also showed some competence improvement as well as high suitability of the chosen training topics, reflecting the perceived relevance and suitability of the training content. Post-training assessments showed an increase in self-reported competences in DL and AI application. Knowledge tests further corroborated this improvement, displaying increased median scores and a reduced interquartile range in DL/AI understanding. In Berlin, the LLM analogue exercise was well-received, whereas digital exercise on LLM, ethics case studies, and Orange3 received lower ratings. The qualitative feedback clarified that the case studies on ethics require more refinement, albeit some participant highly appreciated their inclusion into the training. While CS participants were very interested in the planned digital exercise on ChatGPT (LLM) and Jupyter Notebooks, the exercise was cut short due to the time limit of the training. Similar to the first round of interventions, teachers highlighted the importance of adapting the training material to cater to various grade levels and the didactic needs of students (e.g., the analogue exercise on LLM would not suit more advanced students).

A recurrent theme across responses was the inadequacy of training duration. While the training of the CS teachers (2 hours 15 minutes) did not manage to allocate sufficient amount of time to the two important blocks: ethics and digital LLM exercise using Jupyter Notebooks, the longer STEAM training (4 hours) needs to take into the account the lacking CS background of some STEAM teachers. This is also apparent in

the periodic recommendations for future training to concentrate on different formats, like those designed for more lasting knowledge retention (such as training series) or modular learning (that can be more accessible for teachers via online learning). There is a need to identify suitable formats and incentives to address the needs for further training as well as challenges of finding time given work-related and private time demands of teachers.

Both CS and STEAM trainings provided participants with highly valuable content in both theoretical knowledge as well as pedagogical aspects (introducing materials to the teachers so that they could diver into the learning experience of students). Yet these single trainings were not sufficient for all of the participants to integrate the learned subjects into their teaching, especially if they never taught DL and AI before the training. Inadequate technical infrastructure (e.g., availability and maintenance of computers) and resources, alongside a lack of sufficient foundational knowledge, were identified as major barriers to the effective integration of DL and AI into teaching. In addition, STEAM teachers who require much more training on the basics of DL and AI compared to the CS teachers did not always see a clear connection of DL and AI to their subjects and have to prioritize their curriculum content, leaving no space and no incentives to integrate DL and AI subjects into their classes.

Despite greater obstacles of DL and AI integration for the STEAM teachers, they joined their CS colleagues in valuing the importance of DL and AI topics into teaching due to the importance of these topics in society and everyday life. However, echoing the results of the first round of interventions existing curriculum overload was cited as a significant challenge, necessitating the reprioritisation of topics for the successful integration of new subjects like DL and AI. Yet, integrating DL and AI into the framework curriculum does not automatically result in sufficient DL and AI expertise among teachers. Training teachers both via educational programs at universities (pre-service) as well as via professional development (in-service) is essential and cannot be accomplished via single training sessions.

# 5. Limitations

Some inherent limitations of the evaluation should be considered:

**Workshop Format Variation**: The three trainings (CS German, STEAM Austria, and STEAM Lithuania) differed in the composition of the participants, format, and content. Therefore, we cannot trace any impact on the training format as such.

**Sample Size and Composition**: The small number of participants from each location affects the generalizability of the findings. Expanding the sample size in future studies would contribute to a more robust and generalisable outcome. Only 12 out of 20 participants in the STEAM trainings had no CS background (10 participants were in the Lithuanian training).

**Instrument Validity and Reliability**: The quality of the survey and knowledge test instruments was not evaluated prior to their application, which could influence the reliability of the data collected. There is an ongoing effort to collect more data and evaluate the questions in future project rounds. All instruments were pre-tested with a small sample of respondents.

**Cross-cultural Challenges:** While comparing findings between the countries, one should be aware of the fact that differences could also be traced to cross-cultural differences affecting survey response styles, e.g., differences in acquiescence response styles (Rammstedt, Danner, and Bosnjak 2017).

**Differentiation between subjects:** In the interviews and in some survey items, it is not always clear, whether respondents mean their CS subject or their second STEAM sub-

ject. Although all the training in Berlin was conducted within the context of CS teachers with CS being in the centre of the discussion and the STEAM subjects were in the centre of the STEAM trainings, one must be careful interpreting the findings.

# 6. Conclusion

This evaluation report provides insights into the impact of the three designed trainings on DL and AI for secondary-level in-service CS and STEAM teachers across three European countries – Germany, Lithuania, and Austria. The evaluation of the training sessions on DL and AI has offered valuable insights, albeit with certain limitations and challenges. The diverse backgrounds of participants and the variance in training formats, coupled with a smaller sample size in the second round, presented complexities in result analysis. The third round's introduction of more STEAM trainings with a larger group of teachers is anticipated to provide a more comprehensive understanding of the training's impact and effectiveness.

Participants appreciated the training as an effective introduction to DL and AI, with quantitative data showing competence improvement. However, the need for refined training materials and longer, more diverse training formats was evident, particularly to accommodate the varying needs of CS and STEAM teachers.

Major barriers to integrating DL and AI into teaching included inadequate resources and a lack of foundational knowledge, especially among STEAM teachers. This pointed to the necessity of not only integrating these subjects into the curriculum framework, but most importantly, providing comprehensive, ongoing professional development to equip teachers effectively.

Disclaimer: Parts of this text could be generated or rephrased by ChatGPT, DeepL Write, LanguageTool, and Google Docs spell checking, but were carefully checked and revised by the authors.

# 7. References

Baskerville, Richard L., and A. Trevor Wood-Harper. 1996. "A Critical Perspective on Action Research as a Method for Information Systems Research." *J Inf Technol* 11 (3): 235–46. https://doi.org/10.1080/026839696345289.

Burns, Anne. 2010. *Doing Action Research in English Language Teaching: A Guide for Practitioners.* ESL & applied linguistics professional series. New York: Routledge.

Creswell, John W., and Vicki L. Plano Clark. 2018. *Designing and Conducting Mixed Methods Research.* Third edition. Thousand Oaks, California: SAGE.

European Commission / EACEA / Eurydice. "Informatics Education at School in Europe." https://redined.educacion.gob.es/xmlui/bitstream/handle/11162/235704/informatics.pdf?sequence=1. Accessed June 16, 2023.

Kuckartz, Udo, and Stefan Rädiker. 2020. *Fokussierte Interviewanalyse mit MAXQDA: Schritt für Schritt.* Springer eBook Collection. Wiesbaden: Springer VS.

Mayring, Philipp. 2010. *Qualitative Inhaltsanalyse: Grundlagen und Techniken.* 11., aktual. und überarb. Aufl. Pädagogik. Weinheim, Basel: Beltz.

OECD. 2023a. "Education GPS - Germany - Overview of the Education System (EAG 2022)." Accessed June 10, 2023. https://gpseducation.oecd.org/CountryProfile?plotter=h5&primaryCountry=DEU&treshold=5&topic=EO.

OECD. 2023b. "Education GPS - Lithuania - Overview of the Education System (EAG 2022)." Accessed June 16, 2023. https://gpseducation.oecd.org/CountryProfile?plotter=h5&primaryCountry=LTU&treshold=5&topic=EO.

OECD. 2023c. "Education GPS - Austria - Overview of the Education System (EAG 2022)." Accessed June 19, 2023. https://gpseducation.oecd.org/CountryProfile?plotter=h5&primaryCountry=AUT&treshold=5&topic=EO.

Rammstedt, Beatrice, Daniel Danner, and Michael Bosnjak. 2017. “Acquiescence Response Styles: A Multilevel Model Explaining Individual-Level and Country-Level Differences.” *Personality and Individual Differences* 107:190–94. https://doi.org/10.1016/j.paid.2016.11.038.

Ridsdale, Chantel, James Rothwell, Mike Smit, Michael Bliemel, Dean Irvine, Dan Kelley, Stan Matwin, Brad Wuetherick, and Hossam Ali-Hassan. 2015. “Strategies and Best Practices for Data Literacy Education Knowledge Synthesis Report.” Accessed July 16, 2023. https://www.researchgate.net/publication/284029915_Strategies_and_Best_Practices_for_Data_Literacy_Education_Knowledge_Synthesis_Report.

Schröder, Eike, Felix Suessenbach, and Mathias Winde. 2022. “Policy Paper: Informatikunterricht: Lückenhaft Und Unterbesetzt: Informatikunterricht in Deutschland – Ein Flickenteppich Auch Hinsichtlich Der Datenlage.” Accessed June 07, 2023. https://www.stifterverband.org/medien/informatikunterricht.

Schwarz, Richard, Lutz Hellmig, and Steffen Friedrich. 2022. “Informatik-Monitor.” https://informatik-monitor.de/fileadmin/GI/Projekte/Informatik-Monitor/Informatik-Monitor_2022/Informatik-Monitor_2022_2023_FINAL.pdf. Accessed June 07, 2023.

# Appendix