AI4ED

TOWARDS AN Artificial Intelligence (AI) DRIVEN EDUCATIONAL PROCESS INTEGRATING MODERN CAREERS IN THE EDUCATIONAL SYSTEM

Deliverable

D 3.1
-
AI4Ed Requirements on data semantics and data formats

Deliverable Lead: Uni Bremen
Deliverable due date: 31/07/23
Actual submission date: 27/07/23
Dissemination level: PU
Version: DRAFT

Document Control Page	
Title	AI4Ed Requirements on data semantics and data formats
Creator	University of Bremen
Description	Concrete application of deliverables of WP 2 for data collection, proceeding, and exploitation
Contributors	All partners
Creation date	21/07/2023
Type	Report
Language	English
Audience	☒ public ☐ confidential
Review status	☒ Draft ☐ Assigned Reviewer accepted ☐ Coordinator accepted
Action requested	☐ to be revised by the Assigned Reviewer ☒ for approval by the Project Coordinator ☒ for acknowledgement by Partners

Revision history

Version	Author(s)	Changes	Date
V0	Uni Bremen	Document creation	21/07/2023
V1	All	feedback	08/2023
V2	Uni Bremen	Updated draft	31/08/2023

Table of Contents

EXECUTIVE SUMMARY / ABSTRACT

Abstract	*This report focuses on recommendations for the collection of data and the internal flows of information and technologies. Data protection, data minimisation, storing and sharing of data and data semantics and formats have to be taken into account.*
	Report does not aim at an unified structure of data semantics and/or format; it hopes to ease the development, application and evaluation of learning environments, that have the potential to exploit the advantages of AI.
Keywords	AI, data format, data structure

.

1   Introduction

The project AI4Ed addresses a long-term strategic vision to provide personalised tutoring according to the student´s/apprentice`s needs and to prevent dropout via artificial intelligence (AI). This will be achieved by developing a theoretical framework in which pedagogy based key performance indicators (KPIs) are identified and integrated into AI systems. To assure a smooth collection, proceeding, calculation and evaluation of data, some AI-specific requirements on data semantics and data formats have to be met.

This document intends to give a concrete approach to legal, technical and general pre- and boundary conditions and visions for the implementation of the different learning environments on moodle (or comparable learning platforms) and how those could be supervised and exploited under the umbrella of the AI learning processes. It is also a summary of the potential data formats that can be processed by the chosen AI system, referring to personalised tutoring, active learning and dropout prevention.

2   Data protection

In the recent years, many laws and regulations, focussing on data protection in general and with respect to AI, have been published (cp. f. e. EU 2018) – and this process in ongoing.

They claim relevance on European, national, regional or institutional level and must be respected during all activities within AI4Ed project.

Partly they are contradictory, for sure they are subject of ongoing updates; for example the European Union announced in 06.2023: "As part of its digital strategy, the EU wants to regulate artificial intelligence (AI) to ensure better conditions for the development and use of this innovative technology" (EU 2023); whenever this regulation will be published.

Thus, a description of a detailed standard of data protection or allowed formats and the actual procedures would not be sustainable; please simply accept/follow the current regulations of the institutions where you aim at gathering data for AI4Ed project.

Regulations and procedures might differ largely; for example if ITB Uni Bremen wants to collect the same (from structural perspective) data in university lectures and in lessons of Vocational Education and Training (VET) schools:

A lecturer at the university has to ask for the permission of the data protection officer at the university and of the students, only. The collection of non-anonymous data for AI projects might be allowed.

A teacher at a VET-school has to ask for permission of the rector, the officer for data protection in the German federal state of Bremen, by the apprentices and, in case that they are minor; by their parents. The collection of non-anonymous data for AI projects is not an option at all.

3   Data minimisation

Clearly, the use of AI for drop-out prevention, personalised tutoring and active learning aims at identifying patterns of risk/supporting factors, that are not recognised by an experienced teacher/lecturer. But, nevertheless, data minimisation should be high on the agenda of all partners.

Please ask yourself in the course of AI4Ed project, whether the information that a concrete student/apprentice has a risk of failure is important – or that the relevant data would be to identify patterns that indicate this risk.

During the previous meetings it was decided, and in the previous reports (D2.1. and D2.2.) published, that AI4Ed developed an ambiguous arsenal of Key Performance Indicators (KPI). This does not imply that all partners have to apply all of them – in the contradictory; please respect the principle of data minimisation.

Do not collect data of which you do not have the hypothesis that it might help you/the AI to clarify variance of learning patterns of your students/apprentices.

For example, please reflect on the need of collecting data that refers to the teacher/lecturer.

Roughly written, data collected might be anonym, pseudo-anonym, or personal. Under the principle of data minimisation; anonymous data is favourable:

1. Anonymous data: Please ensure, that no options to identify a concrete learner are available. This includes: No chat, functional e-mail addresses to log-in ("student1/apprentice1") and institutional devices (no personal ip address stored).

2. Pseudo-anonym: Some personal data is/might be collected (sex/age/(ip addresses)/ students/apprentices write individual stuff in the chat) that allow identification with minor or major efforts.

3. Personal: Each person (student/apprentice or teacher) and its activity is detected and their activities are visible to everybody who has access to the learning platform/data storage system.

4   Storing and sharing of data

IT-Servers should be located in a secured place; for example in the Housing-Centre of your institution. Access to Housing-Centre should be controlled; for example via badges with Pin. Servers should be located in one section and only persons with servers in this section should have access to it. The whole Housing-Centre should be secured by an alarm-system. All servers should be secured by strong passwords, ssh-keys and all data should be stored with encryption, so that only the owner of the data has access to it. Clients should be secured by strong passwords, virus/malware protection and data-encryption.

If storing on your local server is not an option; store on another server within EU with highest security standards.

Share raw data with no-one but Alchemy. Assure that access to sensitive data by Alchemy is only possible via secure VPN-Tunnels.

5   Data semantics and formats

The experts for AI, Alchemy, inline with partner IMH, forwarded the following exemplarily table on data, that they can proceed:

Parameter	Type	Description
Gender	*Boolean*	The student's gender.
is_spanish	*Boolean*	If the student is Spanish.
score_avg	*Float*	The student's average grade in the formation.
Course	*Integer*	The course the student is currently in.
Ects	*Integer*	Number of credits obtained so far.
years_studying	*Integer*	Years that the student has been studying in the formation.
num_not_presented	*Integer*	Number of final exams not presented.
num_not_passed	*Integer*	Number of final exams not passed.
num_passed	*Integer*	Number of final exams passed.
num_b	*Integer*	Number of final exams with B grade.
num_a	*Integer*	Number of final exams with A grade.
num_announcements	*Integer*	Total number of announcements taken by the student.

Table 1. Exemplarily type of data that is processable by AI; Source: Alchemy/IMH

Additionally it was decided, that all learning activities should happen on moodle, as their API (Application Programming Interface) allows export of data of learning patterns via *.csv.

Please announce, if you use scales, whether they are metric, ordinal or nominal and, if using a Likert scale from 1-5, whether "5" is the highest confirmation or rejection.

o **Dataset**

Regarding data collection, the educational cycle is generally divided into three distinct parts: before the start of the course, during the course, and at the end of the course. Each of these stages has specific characteristics and objectives, and relevant information is collected in each one for the development and evaluation of the educational process.

These are the different stages:

- Before the start of the course
- During the course
- Upon finishing the course

▪ Before the start of the course

During this stage, students will fill out forms to provide personal information. This form might contain the following questions:

1. *Date of birth*
2. *Gender*
3. *What is your nationality?*
4. *Are you a native of the country where the course is being offered?*
5. *What is your current level of education?*
6. *What was your final grade at your current level of education?*
7. *When did you completed your last level of education?*
8. *What is your current employment status?*
9. *What is your area of study or expertise?*
10. *Which is your main motivation for taking the course?*
11. *What is the level of motivation you have for the training?*
12. *What are your expectations of the results you wish to achieve?*
13. *Are you afraid of failure or reluctant to follow the course comfortably?*
14. *Special conditions that we should be aware of?*
15. *Do you have availability to meet the course schedule?*
16. *Do you have any time constraints during the course period?*
17. *Do you think the course program is clear enough?*
18. *Do you have stable access to a digital device and the internet?*
19. *What is your skill level with computer tools?*
20. *Do you have specific technical knowledge required for the course you are enrolling in?*
21. *How do you get this specific technical knowledge?*
22. *What is your level of oral communication skills?*
23. *What is your level of written communication skills?*
24. *How would you rate your ability to work in a team?*

25. *How confident are you in your abilities to successfully complete the course?*

26. *How do you usually react when facing unpredictable situations?*

27. *How do you usually react when facing problems to solve?*

28. *What is the approach that most motivate you in the development of learning activities?*

29. *How do you describe yourself as a learner?*

▪ During the course

During the course, information related to students' performance will be collected. This information will be automatically gathered from the specific learning management systems of each partner.

From the learning management systems, the following data will be obtained:

- *Students*
- *Training programs*
- *Subjects*
- *Activities*
- *Contents*
- *Enrolments*
- *Consumed contents*
- *Navigation patterns*
- *Grades*

▪ Upon finishing the course

Upon finishing the course, during this final stage, partners chose from or adapt the following questions related to the course evaluation.

EXPECTATIONS

1. *Did the course meet your expectations?*

2. *How would you rate the relevance of the course to your personal/professional needs and goals?*

3. *Do you believe the course provided you with the necessary knowledge and skills to face challenges related to the topic?*

4. *What was the level of practical utility of the course in your personal or professional life?*

5. *What was the overall impact of the course on your personal or professional development?*

6. *Would you recommend this course to others interested in the topic?*

AUTOEVALUATION

7. *What level of commitment did you have during the course?*

8. *To what extent do you believe you dedicated enough time to complete the tasks and activities of the course?*

9. *To what extent do you believe you dedicated enough effort to complete the tasks and activities of the course?*

10. *How responsible do you consider yourself in meeting the deadlines and expectations of the course?*

11. *How would you rate your level of self-discipline during the course to maintain the pace of study and participation?*

12. *How committed were you to actively participate in the discussions and activities of the course?*

13. *Have you got enough time to finish the activities of the course?*

COURSE CONTENT

14. *Was the course content relevant and applicable to your needs or interests?*

15. *Was the course content well-structured and clearly organized?*

16. *Were the support materials (manuals, presentations, additional resources, etc.) useful for understanding the course content?*

17. *How often did you need to review the contents and learning resources?*

18. *Do you think that the learning resources and guidance allowed you to progress autonomously throughout the study?*

19. *Do you think that you needed more support from the tutor/trainer?*

20. *Did the assessment tools and methods seem appropriate to you?*

21. *Was the feedback from this assessment always fast and clear?*

22. *Did the feedback from this assessment facilitated your progression in learning?*

6   Conclusions

The application of AI within educational context is still a new and fascinating field. Different approaches, traditions and believes should not only be respected, but also appreciated.

Thus please understand the project AI4Ed as an experimentation field. If your findings are; for example, that some of the KPI that you have chosen are useless, or that AI support for "individual tutoring" is of added value within your setting, AI support for "dropout prevention" does not deliver any additional information at all – no problem; we are a "forward looking project".

7   References

EU   (2018):   Artificial   Intelligence   for   Europe.   https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=COM:2018:237:FIN; visited 21.7.2023.

EU (2023): EU AI Act: first regulation on artificial intelligence. Updated: 14-06-2023. https://www.europarl.europa.eu/news/en/headlines/society/20230601STO93804/eu-ai-act-first-regulation-on-artificial-intelligence; visited 21.07.2023.