

Template for AI characterisation

Case study: Duolingo for Schools

This case study is realised using the ready-to-use format template that was designed by LORIA¹ and Inria in the context of the AI4T project. It is based on the LORIA Report on template for analysing AI-related features in learning resources².

“Duolingo for Schools is a dashboard right inside of a teacher’s Duolingo account that allows them to create classrooms and assignments, and to follow students’ activity. We’re excited to work with educators to bring the world’s number one language learning app into classrooms, with features designed to maximize teacher effectiveness and student learning”.

Let’s have a look at how it works!

► For Duolingo for schools, we use both elements given about Duolingo and Duolingo for School as the second one is a layer (dashboard for teacher) over Duolingo.

Note

Please consider that the elements given in this template are relevant for one specific use scenario of Duolingo for school where:

1. teachers create accounts for pupils
2. with no elements of personal identification (Name, Surname, email address).

In other available scenarios (for example, a teacher uses pupils already existing accounts (ID/Email), some responses given in this template won’t be relevant.

¹ Loria (Lorraine Research Laboratory in Computer Science and its Applications) is part of a research unit (UMR 7503), common to [CNRS](#), the [University of Lorraine](#) and INRIA. It is a member of the AI4T consortium and brings its expertise in AI in education (and on learning analysis) to the AI4T Erasmus+ project.

² [Report on template for analyzing AI-related features in learning resources](#) - Jiajun PAN, Azim ROUSSANALY, Anne BOYER - AI4T Erasmus+ project, 2022.

Template for AI characterisation

Case study: Duolingo for Schools

Group 1: Usage Scenario and Decision level

This class is to clarify the purpose and users of an AIER and to help users to understand what their role is in using the Artificial Intelligence Educational Resources - AIER.


- Layer 1: Usage Scenario

- Who is the resource for?



- Learner-oriented AI,
- Instructor-oriented AI and
- Institutional system-oriented AI.

- What are the purposes of the resource?




Duolingo for schools “With Duolingo for Schools, teachers can create one (or more!) classrooms and either add students, or allow their students to join themselves. Teachers can manage student access to course content and track student progress throughout their learning journey by tracking student activity, including XP earned and time spent working on a course, as well as individual student level activities, such as lessons completed and scores earned”.

<https://duolingoschools.zendesk.com/hc/en-us/articles/6830454446093-What-is-Duolingo-for-Schools->

<https://blog.duolingo.com/duolingo-for-schools/>

Duolingo: Duolingo proposes adapted exercises and learning pathways to improve proficiency in many languages. It can be used to learn French.

 Duolingo for schools propose training in 4 of the 5 languages of the AI4T Consortium: DE, EN, FR, IT, DE

Usage scenarios layer describe the users and applications of the AIER, which is often the first information teachers want to know. AIER could be classified into 3 types:

- ◆Learner-oriented AI,
- ◆Instructor-oriented AI and
- ◆Institutional system-oriented AI.

There are many different usage scenarios based on the type of AIER. For example, learner-oriented AIER could be an intelligent computer system designed to help students learn a course, while a teacher-oriented AIER could be one that automatically generates questions and tests for teacher-prepared courseware for a particular course.

Template for AI characterisation

Case study: Duolingo for Schools

- Layer 2: Decision level
- What is the decision for the AI?

Duolingo for schools: Adapt learning pathway of students or students groups when teachers assign them tasks via a dashboard

Duolingo: AI algorithms help to provide, for example, the right exercise at the right moment while considering many educational concepts such as:

The Hermann Ebbinghaus “forgetting curve” (<https://blog.duolingo.com/how-we-learn-how-you-learn/>)

The “learning frontier” (pushing your learning to the next level, but only when you're ready for it) <https://blog.duolingo.com/keeping-you-at-the-frontier-of-learning-with-adaptive-lessons/>

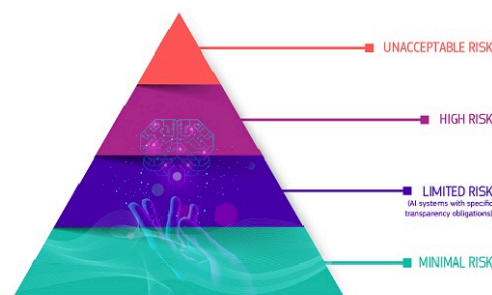
- How sensitive this decision is?

Duolingo for schools: As far as elements from the dashboard (skills /levels / lessons) are not directly transformed into grades or that Duolingo is not used for summative assessment, the risk can be considered as low.

Reminder:

This consideration should be seen in the context of the **Regulatory framework on AI** proposed by the EU, where all AI education systems are classified as high risk by default.

A risk-based approach



The Regulatory Framework defines 4 levels of risk in AI:

- Unacceptable risk
- High risk
- Limited risk
- Minimal or no risk

The **Regulatory framework proposal on artificial intelligence set up by the EU** classifies as high-risk AI systems used in: “initial or vocational training, that may determine the access to education and professional course of someone’s life (e.g. scoring of exams)”.

Template for AI characterisation

Case study: Duolingo for Schools

- Who bears the responsibility for the decision?

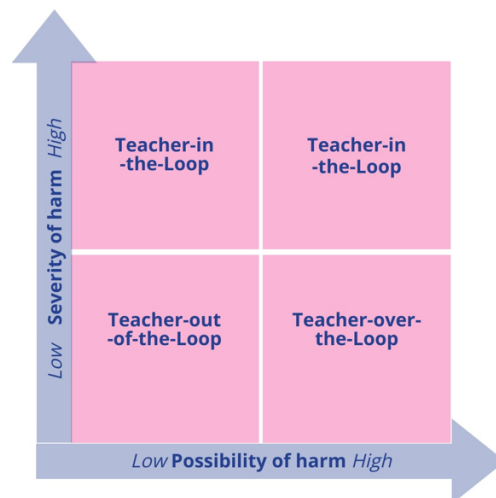
Duolingo for school: Decision to give a task or assessment to a learner is taken by the teacher.

Duolingo: Exercises and their layout is proposed by the AIER.

- How far is the teacher in the decision loop?

Duolingo for school: Teachers can propose two types of assignments to learners, the first is to collect experience points (or XP) and the second one is to level up a skill.

Duolingo: Once the decisions are made by teachers, it's Duolingo for schools which set up the actual detailed tasks to learners. It's **Decision automation** and Teachers in the precise case are out-of-the-loop.



- **Teacher-out-of-the-loop:** In a situation where there is a low probability and low severity of harm caused by, for example, an educational app that is used out of school, the educator's oversight is not required (Figure 1, bottom-left).

Decision level layer describes the proportion of the final outcome that is attributable to the faculty when teacher using AIER. Normally, the decision level depends on the output of the AIER. For example, for AI that uses scoring strategies to automatically assess student learning outcomes, the level of decision making depends on whether the scores are sent directly to the student or to the teacher and who makes the subsequent assessment results. Decision level helps users to know their right and role in using an AIER.


Template for AI characterisation

Case study: Duolingo for Schools

Group 2: Technology and Algorithm

This class is to clarify the approaches the AIER used to reach its purpose and to help teachers understand how it works.

- Layer 3: Technology
- What is the issue addressed by the AI?

 **Duolingo for Schools:** It's a **dashboard** right inside of a teacher's Duolingo account that allows them to create classrooms and assignments, and to follow students' activity.

<https://blog.duolingo.com/duolingo-for-schools/>

Duolingo: It's a language **learning app** which claims 500 Million users worldwide.

<https://www.duolingo.com/learn>

Technology layer describes the AI technologies used in the AIER. Considering that an AIER may be an educational platform, a standalone or online application, aso., several different technologies may be used in a single resource. The purpose of this layer is to clarify whether AI-based technologies are actually being used and to introduce users to what kind of AI technologies are being used.


The technology could be one of the following list:

- ◆ Automatic generation of educational content (Courses, texts and so on)
- ◆ Improved educational content
- ◆ AI assistance for teachers to provide personalized instruction for each student
- ◆ Fast feedback to students
- ◆ Assistance in monitoring students.
- ◆ Automatic assesses students' learning behaviour and learning routes
- ◆ Automatic records the student's learning process.
- ◆ Targeted improvement based on learning analysis
- ◆ Chatbot between teachers, students, parents and relevant groups
- ◆ Other technology for educational tasks.

Template for AI characterisation


Case study: Duolingo for Schools

- Layer 4: Algorithm
- Which family of approaches do the algorithms related?



- Knowledge-based Systems:
- Machine Learning: for the 2 found examples
- Deep Learning:

- What are the algorithms?



Duolingo:

Ex1. Sleeping recovering Bandit for notifications on Duolingo
<https://research.duolingo.com/papers/yancey.kdd20.pdf>

Ex2. Bayesian Knowledge training for knowledge tracing (estimating a student's current knowledge state and predicting his/her performance in future interactions, students can receive personalised learning materials)
<https://sharedtask.duolingo.com/papers/chen.slam18.pdf>

Algorithm layer describes the algorithms used in the AIER. Compared to the previous layer, this layer focuses more on the scientific level than on the technical level. We will declare the family of the algorithms from three common groups: 🌟 Knowledge-based Systems, 🌟 Machine Learning and 🌟 Deep Learning. Note that we are aware of the debate about whether deep learning is machine learning or not. However, we have separated them here because it is easier for teachers to understand. For each group, the algorithm could be selected from the following list:

- ◆ Knowledge-based Systems: Rule-based systems (or expert systems), Ontology, Semantic networks and so on.
- ◆ Machine Learning: Clustering, Approximate possibility, Regression Analysis, Representation and Dimensionality reduction, Active learning, Decision Trees, and so on.
- ◆ Deep Learning: Convolutional Networks (CNNs), Long Short-Term Memory Networks (LSTMs), Recurrent Neural Network (RNN), Generative Adversarial Networks (GANs).

Please note that we will only identify the algorithms used in this layer and briefly introduce the algorithms, without comparing and evaluating the algorithms used.


Template for AI characterisation

Case study: Duolingo for Schools

Group 3: Personal Data - Data security - Transparency

This class is to clarify the input data used by AIER and the secure handling of that data, to make it clear that AIER demonstrates the transparency of the internal mechanisms and to help users trust in the security of AIER.

- Layer 5: Personal data
- What kind of personal data is used?

 **Duolingo:** In order to propose personalised exercises and learning pathways to learners, there is a need for the AIER to follow learners.

In the context of Duolingo for schools, it's possible without **personal data** from students.

Duolingo for schools: allows the creation of profiles without name, surname or email addresses:

- Teachers can also create accounts for their students.
- There is no need to add name or forename or email address
- Example of registration made by a teacher in Duolingo for schools:

<https://drive.google.com/file/d/1a71TUmkdXjSr8fldU4RAe5BYdRE9UJpC/view?usp=sharing>

Note: Teachers have to have a Duolingo account (creation: username + email address)

<https://support.duolingo.com/hc/en-us/articles/204829870-How-do-I-create-an-account-and-or-login>

In this case the general data policy of Duolingo applies: <https://www.duolingo.com/privacy>

👉 Any user can access it's personal data and ask for them to be removed.


Personal data layer lists the personal data used in the AIER. Nowadays information security is increasingly valued. In addition to the data explicitly proposed for input in the AIER, external sources, such as shared social accounts, may collect additional personal information. The creation of this layer is necessary to enable users to clarify this information.

Template for AI characterisation


Case study: Duolingo for Schools

- Layer 6: Data security

- Is the personal data anonymous? If yes, how does it work?

 **Duolingo for schools:** In the usage scenario chosen for this example, students are registered by their teacher without the use of name, first name or email address. The data is therefore anonymous.

- What are the possibilities for outsiders to audit the resource?

 **Duolingo for schools & Duolingo:** As far as information was found on the subject, there is no possibility for an outsider to audit the resources
The Common Sense Privacy Program gave Duolingo an overall 63% rating in 2021
<https://privacy.commonsense.org/privacy-report/duolingo>

The Common-Sense Privacy program evaluates **privacy policies** so that parents and teachers can make smart choices about the learning tools they use with their children and students, and so that schools and districts can participate in improving the technology used in K–12 classrooms).

Data security layer specifies whether non-user visitors and external visitors to the AIER can access the data and how the data is to be handled securely. This layer is related to the previous one. Different personal data should be handled in different ways, such as anonymity, encryption, and denial of access.

◆ **Anonymization** means that the information recipient is completely unable to identify the data subject directly or briefly. There are five common data anonymization operations: generalisation, suppression, anatomization, permutation, and perturbation.

◆ **Encryption**, also known as **pseudonymization** means that the information receiver cannot identify the data subject directly, but the information can be de-anonymized by the data manager after cross-identification with other information.

◆ **Denial of access** is a simple and straightforward way to keep the information only in the information manager, without giving access to others.

In this layer, it also clarifies whether the data will be used by external visitors. Many software share data with their own data partners nowadays, and if this is the case, it can also lead to different issues regarding the security of the data.

Template for AI characterisation

Case study: Duolingo for Schools

- Layer 7: Transparency
- Is there a mechanism for explaining decisions? If yes, how does it work?

Duolingo for School / Duolingo

There is no mechanism to explain the many micro-decisions made by the AIER to provide tasks appropriate to each stage of the learning path in each language offered in Duolingo. Incidentally many scientific articles are published by the Duolingo teams in scientific publication to explain how some elements of the model work:

<https://research.duolingo.com/>

To complement this layer about transparency even if not about AI transparency, Duolingo offers learners the possibility to take the “Duolingo English Test”.

As learner can use the result of this test to claim their proficiency in English to register for Higher Education institutions, many info are published about it:

- **The Duolingo English Test Technical Manual:** it provides an overview of the design, development, administration, and scoring of the Duolingo English Test. Furthermore, the Technical Manual reports validity, reliability, and fairness evidence, as well as test-taker demographics and the statistical characteristics of the test. <https://duolingo-papers.s3.amazonaws.com/other/det-technical-manual-current.pdf>
- An external assessment of the reliability of the test, particularly in terms of fraud prevention, is carried out.

Transparency layer clarifies the level to understand the internal mechanism of the AIER. For example, in the educational tasks to assess students through their learning behaviors, the internal mechanism of this AIER is about which learning behaviours were selected and which parameters were used to generate the final report. If the final assessment report is given to the teacher that shows these parameters or if the AIER's introductory website explains the internal mechanism of this AIER, it will help teachers understand the possibilities, limitations, and risks of this AIER in education.

In addition to this, the transparency to teachers of the intermediate results generated by AIERs in the educational tasks is worth declaring at this layer. For example, an AIER delivering personalised instruction may have to first generate a profile of each student that it uses to recommend different educational content in the process of accomplishing this task. If the profile is available for teachers, it will help users understand and trust this AIER.