

Section 1 System Information

1.1 System's Use.

Purpose. Chatbot assessing responses to vocational training tasks.

Capability. Skill acquisition evaluation.

Domain. Education and vocational training.

AI User. Universities.

AI Subject. Students.

content generated by FillGen

1.2 Components. The chatbot uses NLP processing tools for advanced text analysis and language understanding, and APIs to connect with university systems like Learning Management Systems. Hosted on cloud computing platforms, it ensures efficient operation and data management. It also uses database systems to store student responses and evaluations. Monitoring is key for maintaining system performance and security. It includes tracking response accuracy and speed, ensuring user satisfaction, and adhering to data security regulations like GDPR. Additionally, a user feedback system gathers insights from universities and students, essential for ongoing refinement and effectiveness.

1.3 Data. The input data includes student responses in text or voice to training tasks, along with detailed task specifications. Training data comprises annotated student responses and curriculum guidelines, aligning the chatbot with educational standards. Validation data involves a diverse range of student responses and expert evaluations to test the chatbot's accuracy. Testing data includes unseen student responses, used to evaluate the model's generalization skills, with metrics like accuracy, precision, recall, and F1-score assessing real-world effectiveness. This data structure ensures the chatbot's robust response assessment in varied vocational scenarios.

1.4 Evaluation.

Evaluation at development stage. The chatbot AI's accuracy, response time, and user interaction capabilities were rigorously tested, with a focus on its consistency and scalability, although it faced challenges in fully understanding complex vocational contexts.

Evaluation at deployment stage. During deployment, the chatbot was closely monitored for real-time accuracy and efficiency in handling live student interactions and task specifications, ensuring stable performance under varying workloads.

Evaluation at use stage. In the usage phase, the chatbot's ongoing accuracy, response efficiency, and reliability were continually assessed in educational settings, with user feedback guiding improvements and addressing any emerging technical limitations.

1.5 Teams. The development team included software engineers, instructional designers, and machine learning experts, ensuring a diverse skill set. In terms of gender diversity, there were team members representing female and male perspectives.

Section 2 Risks

This system is high risk as per EU AI Act because it is used in the domain of 'Education and vocational training' (Annex III), specifically for the purpose of assessing students' responses to vocational training tasks. Systems used for assessing students in educational and vocational training institutions are classified as high risk.

2.1 Capability Risks

Privacy Breach. Students risk having their sensitive personal information exposed if the chatbot system fails to adequately secure data, leading to privacy violations and potential misuse of information. Universities face the risk of violating privacy rights if the chatbot AI mishandles personal data, necessitating stringent data security and explicit user consent to avoid legal issues and trust erosion (HR Article 12).

2.2 Human Interaction Risks

Misinterpretation and Misguided Learning. Students are at risk of receiving inaccurate feedback or misinterpretations of their responses due to limitations in the AI's understanding. This could lead to misguided learning paths or misconceptions in their vocational training.

Dependency and Skill Erosion. There's a risk that reliance on the AI for assessment and feedback could lead to a decline in critical thinking and self-assessment skills among students, as they might become overly dependent on automated feedback instead of developing self-evaluation capabilities.

2.3 Systemic Impact

Gender Equality and Bias. There's a potential risk of the AI system inadvertently promoting gender discrimination, challenging the goal of gender equality in education and possibly leading to biased outcomes (Goal 5).

Amplifying Inequalities. The deployment of the chatbot AI could risk amplifying educational inequalities (Goal 10). Biases or accessibility issues within the system may lead to discriminatory outcomes, undermining equitable education in institutions and widening societal educational disparities, particularly affecting marginalized communities.

Section 3 Mitigation Strategies

The AI system can become minimal risk as per EU AI Act if it is redesigned to serve as a supplementary tool for human educators, rather than independently assessing students' responses to vocational training tasks. The AI system could provide suggestions and insights to human educators who would make the final assessment decisions. This would reduce the direct impact of the system on the students' assessment and thus lower the risk.

3.1 Mitigations of the Capability Risks

Privacy Breach. The system will incorporate explicit consent for data collection, anonymization to protect student identities, and robust security measures to ensure data privacy and security (Article 12).

3.2 Mitigations of the Human Interaction Risks

Misinterpretation and Misguided Learning. Incorporate regular reviews by educational experts to ensure that the chatbot's interpretations and guidance align with accurate and current educational standards. This helps prevent misinterpretations that could lead to misguided learning.

Dependency and Skill Erosion. Combine the use of the chatbot with traditional teaching methods and encourage critical thinking exercises. This approach reduces dependency on the AI for learning and assessment, helping to maintain and enhance students' critical thinking and self-evaluation skills.

3.3 Mitigations of the Systemic Impact

Gender Equality and Bias. The system will be trained on a gender-neutral and diverse dataset and undergo regular bias audits to ensure fair and unbiased assessments for all students (Goals 5 and 10).

Amplifying Inequalities. To mitigate this risk, it is crucial to collaborate with educational experts and stakeholders. It is essential to design the system for universal accessibility and train it with diverse datasets to minimize biases, while continuously monitoring its impact across different student groups to ensure equitable educational outcomes.

Section 4 Benefits

4.1 Human Interaction Benefits

Enhanced Educational Quality and Accessibility. It offers a platform that aids students in acquiring essential skills for employment and entrepreneurship (Goal 4), contributes to students' rights to work, offering multiple pathways to employment and fair remuneration (Article 23), and contributes to the full development of human personality (Article 26).

Freedom of Expression. By facilitating an environment where students can freely express their ideas and thoughts during vocational tasks, the system promotes the exchange of information and ideas, aligning with the right to freedom of expression (Article 19).

4.2 Systemic Impact Benefits

Economic Growth and Employment Opportunities. By fostering technological upgrading and innovation, it aids in developing productive activities, job creation, and ensuring full and productive employment (Goal 8). The system also plays a role in decreasing the number of youths not engaged in employment, education, or training.

Advancing Inclusivity and Sustainable Development. The system significantly contributes to reducing societal inequalities (Goal 10), and promoting sustainable, inclusive societies (Goal 16) by offering unbiased, accessible vocational training opportunities to all students, irrespective of their background. This not only fosters transparency and accountability in educational institutions but also supports sustainable development by optimizing resource use and promoting digital learning environments.