# Practical Interview Protocol – Erasmus Plus Project

Juan Carlos Nieves, Andrea Aler Tubella

Department of Computing Science, Umeå University

Email – juan.carlos.nieves@umu.se, andrea.aler@umu.se

## Abstract

With the goal of exploring the state-of-the-art of Trustworthy AI in Higher Education, we developed this interview protocol. The specific goals of this protocol were to obtain expert feedback on the following topics: 1. General awareness of the Guidelines amongst stakeholders in HE 2. Inclusion of the Requirements in current educational programs 3. Current educational practices for Trustworthy AI (topics, learning outcomes, evaluation) 4. Incentives to facilitate the inclusion of Trustworthy AI topics in HE 5. Risks and opportunities Partners from ALLAI, Universidad de Alcalá, Maynooth University and Umeå Universitet followed a training session in order to unify how the interviews were conducted. Interviewees were therefore asked the same questions in the same manner, allowing to contrast answers in a qualitative analysis.

## Purpose of the Protocol

This protocol is meant as a guideline for interviewers. For the purposes of this document, "you" can refer to the reader (if read within a statement) or the interviewee (if read within a question).

What would be the best-case scenario? In the best-case scenario, this project helps to improve and redesign education programs in Higher Education (HE) in the field/scope of Trustworthy AI.

As a side-effect, the interviews may help the participating entities may get a better view of the aims of the Trustworthy AI guidelines and to reflect on how these affect or are aligned with their own views and education programs.

Our goal is to receive feedback about the following aspects:

1. What is the understanding of HLEG guidelines?

2. How useful is the assessment list for Trustworthy AI in education?

3. How relevant is it for Higher Education?

4. How clear is it for Higher Education?

5. How precise is it for Higher Education?

6. How complete is it for Higher Education?

7. Which issues are already covered by existing HE programs or courses?

8. Which steps would be needed to follow to introduce Trustworthy AI education in HE programs and courses?

## Agenda

Subject to change, to accommodate cultural requirements:
09:00-09:30 · Introductions, including presentation of the Trustworthy AI guidelines

09:30-10:00 ・ HE program (or course) case introduction by the organisation
10:00-10.30 ・ General Perspectives and Questions
10.30-11:00 ・ Questions of the Assessment List
11:00-11:30 ・ Ordering of the Requirements
11.30-12.00 ・ Questions for Specific Areas of the Assessment List
12:00-12:15 ・ Closing remarks

# A.　Introduction

Introduction of the people involved in the meeting, project - background, consent issues, description of process, and follow up steps, etc. Ensure that interview can be (voice) recorded.

# B.　Introduction of the Purpose of the Interview

1. Slide deck

2. Ensure that interview can be (voice) recorded. Make it clear that none other will have access to the recordings and that they will be deleted upon the completion of the project report.

3. Ensure that it is understood it is not about the performance or vision of the entity but about the suitability of the trustworthy guidelines to improve Higher Education.

4. Make it clear that the individuals will not be noted by name anywhere. Any information, e.g. their role or location will only be used in an aggregated manner.

5. Determine whether they would like to list their AI assessment activities as part of the final report, or whether that is confidential.

6. State the agenda for the day.

# C.　Education case (Education program or a course)

Discuss the higher education program (or course) with them. Allow them to present the education case. Make it clear that the scenario is meant to provide contextual information.
Possible topics to discuss/ask:

1. Learning outcomes.

2. Learning outcomes vs the seven requirements of the HLEG guidelines.

3. Teaching material.

4. Examination methods.

5. Heterogeneity of the students.

6. Employability of students.

# D.　General Perspectives

This section is a generalised discussion of the HLEG guidelines and its assessment of AI systems during their development, deployment, and use.

1. How would you describe the current status of "trustworthy" AI in connection with higher education in general? (e.g. national education strategies, practices in current education at your organisation.)

2.  Can you say something about the strategy of your organisation has for AI education development? (Purpose, development, administration, recent initiatives)

3.  Which of the HLEG requirements are you already teaching in your education program? Do you teach other issues related to trustworthy AI?

4.  In which education cycle should trustworthy AI education need to start?

5.  What resources should be available for trustworthy education in HE?

6.  Are there any resources that you already use for teaching aspects related to trustworthy AI?

7.  What are the minimum incentives that should be there for promoting trustworthy AI in HE?

8.  What risks and opportunities do you associate with trustworthy AI in HE?

9.  How could HE benefit from trustworthy AI?

# E.     Questions on the Assessment List

This section aims to ask questions on the HLEG assessment list. Try to keep the discussion within reasonable time limits.

1.  In which language did you read the Guidelines and the Assessment List? (Should be asked prior to the interview, but again during)

2.  In overall terms, is the assessment list useful for education purposes? Why/Why not?

3.  Is it beneficial or not to make it part of an actual HE course? If so, in which form? If not, why?

4.  What type of support do you need to teach the Guidelines?

# F.     Ordering of the Requirements

In this section, request the participants to rank the 7 Requirements (Transparency, Accountability,...) in order of their application/importance (with 1 being the highest).

   Make it clear that the ordering is in terms of significance to their education given the education case and within the context of this interview. You may use the printout cards and/or remind them of the 7 requirements.

   If multiple persons/roles are taking part in the interview, you may record any notable disagreements, but only one order is permitted, i.e. the organisation's position.

1.  Interpretation (their own words) of each of the 7 key requirements.

2.  Which TAIG requirements of the assessment list are relevant? Why/why not? In which order? Make a ranked list.

3.  Why this order and why some requirements are considered less or not relevant?

4.  Are there requirements in the TAIG not relevant? Why?

# G.     Questions for Specific Areas of the Assessment List

Take the two extremes (i.e. the highest and lowest priority) from the list produced in the previous section of the interview.

1.  Which aspects are already considered in their education?

2.  Why is/isn't the requirement relevant for your education?

3.  Evidence of addressing it:

   (a)  What are the learning outcomes related to the requirement?

   (b)  What are the evaluation methods related to the learning outcomes?

4.  Do you think this requirement is clearly outlined in the Assessment List? Could you tell me how you interpret it?

5.  Which questions around this requirement are the most valuable ones for trustworthy AI education?

## H.    Closing remarks

A quick wrap-up of the interview.

1.  Is there anything else you would like to add/ask?

2.  What was most positive from this interview?

## The 7 Requirements

1.  Human agency and oversight

   (a)  Fundamental rights:

   (b)  Human agency:

   (c)  Human oversight:

2.  Technical robustness and safety

   (a)  Resilience to attack and security:

   (b)  Fallback plan and general safety:

   (c)  Accuracy

   (d)  Reliability and reproducibility:

3.  Privacy and data governance

   (a)  Respect for privacy and data Protection:

   (b)  Quality and integrity of data:

   (c)  Access to data:

4.  Transparency

   (a)  Traceability:

   (b)  Explainability:

   (c)  Communication:

5.  Diversity, non-discrimination and fairness

   (a)  Unfair bias avoidance:

   (b)  Accessibility and universal design:

   (c)  Stakeholder participation:

6.  Societal and environmental well-being

   (a)  Sustainable and environmentally friendly AI:

   (b)  Social impact:

   (c)  Society and democracy:

7.  Accountability

(a) Auditability:

(b) Minimising and reporting negative Impact:

(c) Documenting trade-offs:

(d) Ability to redress: