



Trustworthy **AI**

Marco para la Educación Superior en IA Fiable V1.0

Septiembre 2021



Co-funded by the
Erasmus+ Programme
of the European Union

Índice

Índice	2
1. Visión general	4
2. Métodos.....	4
2.1. Revisión de literatura.....	4
2.2. Entrevistas a expertos.....	5
3. Requisitos éticos	5
3.1. Situación actual.....	6
3.2. Requisitos individuales y su importancia en la educación.....	7
2.3. Recomendaciones.....	8
4. Objetivos de aprendizaje y evaluación	9
4.1. Objetivos de aprendizaje	9
4.2. Objetivos de aprendizaje por requisito	10
4.3. Evaluación.....	11
4.4. Recommendations.....	11
5. Recursos para la educación	13
5.1. Estado del arte.....	13
5.2. Necesidades.....	13
6. Estrategias, incentivos, riesgos y oportunidades.....	14
6.1. Estrategias e incentivos	14
6.2. Riesgos y oportunidades.....	15
6.3. Recomendaciones.....	15
7. Referencias	16
Appendix A. Interview profiles	19
Appendix B. Practical Interview Protocol – Erasmus Plus Project.....	21
Appendix C. Literature Review	26

Nota de copyright:

Este material se presenta para garantizar la difusión oportuna de trabajos académicos y técnicos. Los derechos de autor y todos los derechos sobre los mismos pertenecen a los autores o a otros titulares de derechos de autor. Se espera que toda persona que copie esta información se adhiera a los términos y restricciones invocados por los derechos de autor de cada autor. En la mayoría de los casos, estas obras no pueden volver a publicarse sin el permiso explícito del titular de los derechos de autor.

Esta obra está bajo una Licencia Creative Commons Atribución-NoComercial-SinDerivadas 4.0 Internacional.



Titulares del producto intelectual: Andrea Aler Tubella, Juan Carlos Nieves, Umeå University, Sweden

Entrevistas realizadas por:

Andrea Aler Tubella, Umeå University, Sweden
Marçal Mora-Cantallops, Universidad de Alcalá, Spain
Douglas Cirqueira, Maynooth University, Ireland
Noah Schöppel, ALLAI, Netherlands
Christofer Talvitie, ALLAI, Netherlands

Plantilla: Orla Casey, Momentum

Revisiones del documento: all Trustworthy AI partners

1. Visión general

Este documento se ha desarrollado en el contexto del proyecto Erasmus+ " Trustworthy AI " con el objetivo de facilitar la introducción de las Directrices Éticas del Grupo de Expertos de Alto Nivel para una IA Fiable (EUCommission, 2018) en todas las disciplinas de la educación superior. Este objetivo está alineado con la estrategia digital de la UE, que enfatiza la necesidad de educar profesionales que puedan "dar forma a la tecnología de una manera que respete los valores europeos" (UE, 2020). Con este propósito, las Directrices Éticas del Grupo de Expertos de Alto Nivel para una IA Fiable describen los requisitos necesarios para un desarrollo responsable y fiable. El objetivo de este proyecto es, por tanto, utilizar estas directrices como un punto de partida para la introducción de competencias éticas y socio-jurídicas en los temas de Educación Superior relacionados con la IA.

El objetivo principal de este marco es describir los principios y estrategias de aprendizaje a seguir para desarrollar las competencias de los estudiantes. Los hallazgos se presentan en forma de **recomendaciones para educadores**, **necesidades en materiales educativos** e **incentivos de políticas** que responden a las siguientes preguntas:

1. ¿Qué estrategias se necesitan para introducir eficazmente los requisitos éticos del Grupo de expertos de alto nivel en educación superior?
2. ¿Qué competencias y resultados de aprendizaje relacionados con la IA fiable deberían desarrollar los estudiantes en educación superior? ¿Cómo evaluarlos?
3. ¿Qué recursos pueden facilitar la introducción de la educación de una IA fiable?
4. ¿Qué riesgos y oportunidades presenta la introducción de la educación de una IA fiable? ¿Qué incentivos políticos pueden facilitar estas oportunidades?

El Marco que presentamos se basa en el **análisis de 11 entrevistas a expertos en 5 países**, así como en un estudio de la literatura en forma de **revisión sistemática de la literatura**.

A lo largo de este documento, nos referiremos al Grupo de expertos de alto nivel como "HLEG" (de sus siglas en inglés), a las Directrices sobre una IA Fiable como "las Directrices", a la Lista de evaluación asociada a las Directrices como "Lista de evaluación" y a los 7 requisitos éticos descritos en las Directrices como "los requisitos". La educación superior a veces se abrevia como "ES".

2. Métodos

2.1. Revisión de literatura

Un enfoque clave de las Directrices es incorporar perspectivas éticas y socio-jurídicas en el desarrollo y uso de la IA. Sin embargo, las leyes, las normas sociales y las dimensiones éticas son muy contextuales (Turiel, 2001). Por lo tanto, para comprender el estado del arte en la educación superior, se llevó a cabo una revisión sistemática de la literatura (Anexo C) con el objetivo de analizar el desarrollo y la evaluación de las competencias éticas en la Educación Superior en una variedad de campos. Las principales preguntas para el análisis son:

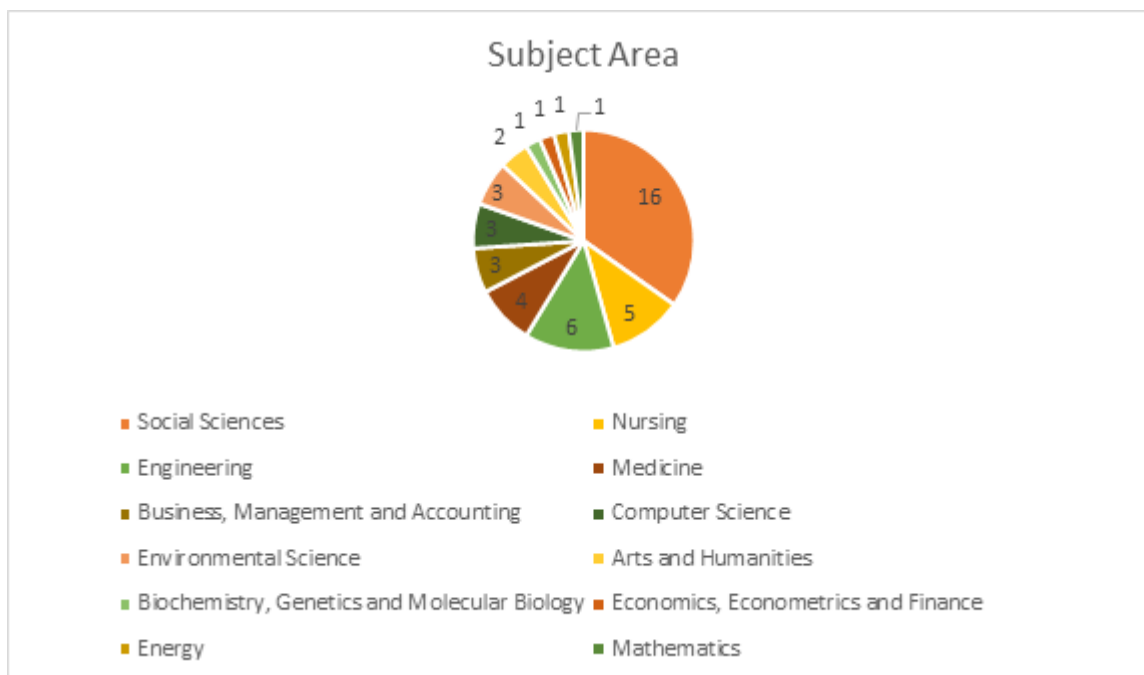
1. ¿Qué competencias deben desarrollar los estudiantes de ES para dominar los aspectos éticos y socio-jurídicos?
2. ¿Cómo enseñar y evaluar estas competencias en el contexto de la ES?

Realizamos la búsqueda de literatura en la plataforma Scopus, para obtener resultados en una variedad de áreas temáticas. Los artículos se identificaron mediante los siguientes términos de búsqueda que aparecen en el título, resumen o palabras clave:

"ethics AND teaching AND "higher education" AND (competence OR competency OR skills)".

Limitamos nuestra búsqueda a publicaciones a partir de 2015 y recuperamos un total de 50 publicaciones. Nos centramos en resultados de investigación originales, por lo que excluimos 1 libro, 1 editorial, 1 resumen extendido y 6 artículos de revisión (revisiones de artículos o revisiones de planes de estudio). Finalmente, se excluyeron manualmente 13 artículos cuyos temas no comprendían la enseñanza de competencias relacionadas con la ética a

estudiantes de Educación Superior (la mayoría de los artículos descartados abordaban la ética de la enseñanza, las competencias para los docentes o no estaban en el ámbito de la ES). Finalmente, no se pudieron recuperar 4 artículos, dejando un total de 24 artículos para análisis. Los artículos analizados cubren una amplia variedad de áreas temáticas. Según la afiliación del autor, se cubren 17 países, así como 12 áreas temáticas indexadas por Scopus.



2.2. Entrevistas a expertos

Con el objetivo de explorar el estado del arte de la IA Fiable en la educación superior, Umeå Universitet desarrolló un protocolo de entrevista (Apéndice B). Los objetivos específicos de este protocolo fueron **obtener comentarios de expertos** sobre los siguientes temas:

1. Conocimiento general de las Directrices entre las partes interesadas de la ES.
2. Inclusión de los Requisitos en los programas educativos vigentes.
3. Prácticas educativas actuales para la IA Fiable (temas, resultados de aprendizaje, evaluación).
4. Incentivos para facilitar la inclusión de temas de la IA Fiable en la ES
5. Riesgos y oportunidades.

Socios de ALLAI, Universidad de Alcalá, Maynooth University y Umeå Universitet siguieron una sesión de formación para unificar cómo se llevaron a cabo las entrevistas. Por lo tanto, a los entrevistados se les hicieron las mismas preguntas de la misma manera, lo que permitió contrastar las respuestas en un análisis cualitativo.

Un total de 11 entrevistados fueron seleccionados por su participación en la Educación Superior, a nivel de desarrollo de políticas, coordinación de programas o enseñanza. Las entrevistas se llevaron a cabo durante un periodo de 6 semanas. Los expertos entrevistados, con afiliaciones en 5 países, contribuyeron casos de uso en los campos de medicina, derecho, informática y ciencias sociales (Apéndice A). Las respuestas de los entrevistados impulsan las recomendaciones de este Marco y ilustran el estado actual de la IA Fiable en la educación superior.

3. Requisitos éticos

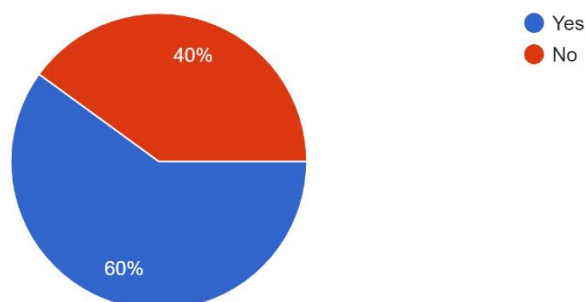
La Comisión Europea ha destacado transformar los requisitos de la lista de evaluación de las directrices éticas en competencias específicas debe ser el siguiente paso en la creación de un “ecosistema de confianza” para el florecimiento de la IA europea (UE, 2020). Por lo tanto, el primer enfoque de este marco es una inmersión en profundidad en los requisitos éticos de la lista de evaluación. Primero analizamos cómo se incluyen actualmente los requisitos de HLEG en la educación, y luego nos enfocamos en los requisitos individuales para comprender su

importancia percibida y las preguntas educativas clave que los rodean. A partir de este análisis, identificamos varias recomendaciones para involucrar eficazmente la lista de evaluación y los requisitos de HLEG en la educación superior.

3.1. Situación actual

Are the HLEG requirements currently included in any form in the course/program?

10 responses

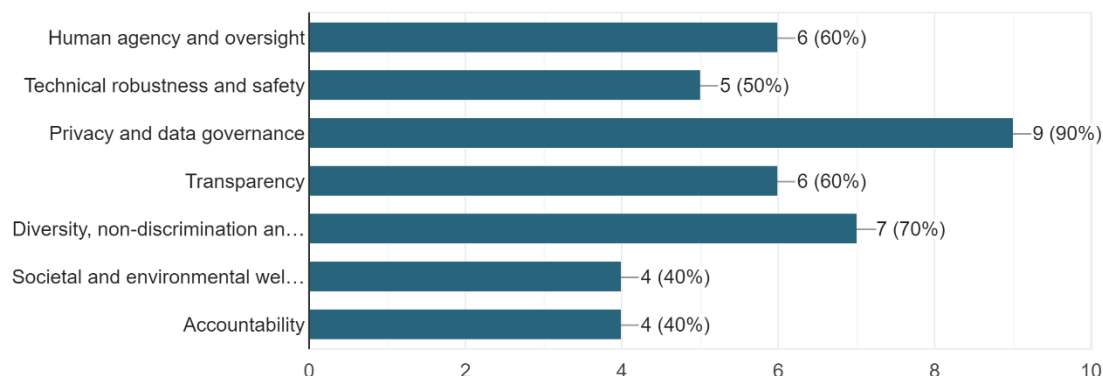


En cuanto a la inclusión de los requisitos éticos del HLEG en los cursos y programas vigentes, **el 60% de los expertos entrevistados afirman que ya están incluidos en su caso educativo**. Una tendencia común en las entrevistas a expertos (7 de 11) es que, **si bien en la educación se cubren diferentes requisitos, no se relacionan explícitamente con la IA ni con las pautas del HLEG**. Muchos aseguran que los temas relacionados con la confiabilidad se abordan debido a su relevancia, pero que la relación con los requisitos del HLEG suele ser implícita en lugar de un esfuerzo deliberado.

Las directrices y la lista de evaluación del HLEG se valoran por la claridad que aportan al especificar diferentes requisitos. Es posible que estos documentos en su conjunto no sean adecuados para su inclusión directa en los cursos, pero el enfoque humano es fundamental y siempre debe mantenerse.

Todos los entrevistados valoran las directrices y la lista de evaluación en su forma actual por establecer requisitos claros y aportar claridad a su significado. Sin embargo, su inclusión en la educación plantea algunos desafíos: los encuestados plantean que la extensión y la naturaleza técnica de los documentos **no es adecuada para todas las disciplinas y niveles educativos**. Además, falta una perspectiva sobre cómo se aplica cada requisito a diferentes disciplinas. El desafío de la traducción de diferentes términos técnicos puede traer diferentes perspectivas dependiendo del idioma de las directrices que se esté estudiando. Una observación frecuente que hacen los expertos es que los **diferentes cursos pueden abordar solo algunos de los requisitos**, por lo que no se examinan las directrices en su conjunto, sino que se centran en unos pocos requisitos específicos relevantes. En general, varios encuestados señalan que el aspecto clave de la Lista de evaluación es **el enfoque en el ser humano detrás del sistema** y enfatizan el valor de transmitir a los estudiantes que la responsabilidad y las obligaciones éticas del desarrollo de la IA recaen en aquellos involucrados en el proceso.

Summary: Which of the HLEG requirements are already being taught in their education program?
10 responses



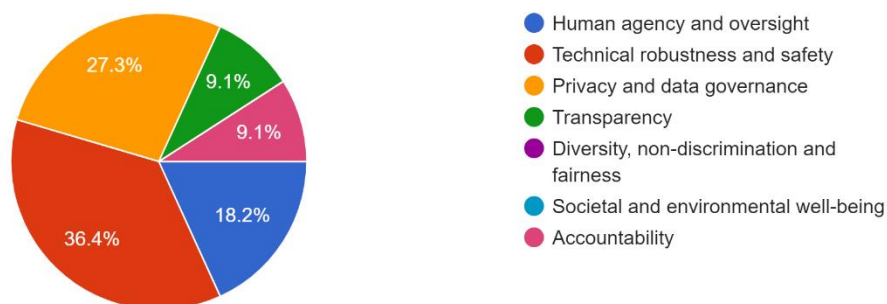
Si se analizan los requisitos de forma individual, destaca claramente en términos de su prevalencia en los programas educativos actuales **“Privacidad y Gobernanza de Datos”**, que actualmente se incluye en el **90% de los programas educativos planteados por los entrevistados**. Siguiendo la idea de lo implícito, los entrevistados afirman que este requisito se aborda a través de la lente de las buenas prácticas de datos, el GDPR o la ética, más que como un enfoque específico en el propio requisito. En el otro extremo del espectro, **“Bienestar social y ambiental”** y **“Rendición de cuentas”** son los menos incluidos, apareciendo en el 40% de los programas de los entrevistados.

Aunque los requisitos del HLEG se incluyen actualmente en los programas educativos, ya que se cruzan con otros temas, a menudo no se abordan a través del enfoque de las Directrices.

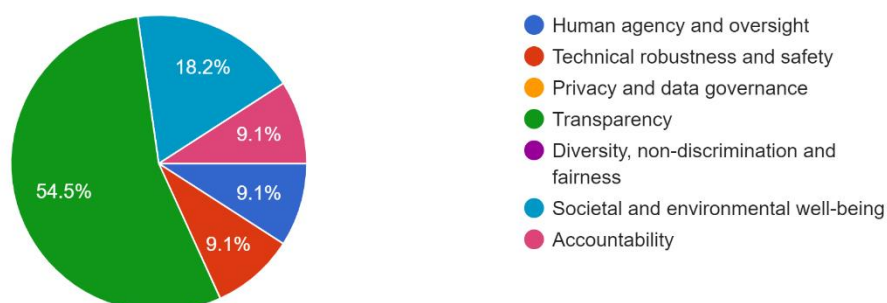
3.2. Requisitos individuales y su importancia en la educación

Un tema clave en las entrevistas a los expertos es una evaluación de los requisitos en términos de su importancia en la educación. En particular, se pidió a los entrevistados que ordenarán los requisitos según su relevancia para su caso educativo. Existe un consenso significativo entre los expertos (100% de los encuestados) de que **todos los requisitos son relevantes**, pero que **su importancia para cada curso varía según el tema y el área**. Por esta razón, no hubo un acuerdo significativo en el orden de los requisitos: cada caso educativo generó diferentes ordenes según el área de aplicación y los temas abordados en el curso.

Highest ranked requirement
11 responses



Lowest ranked requirement
11 responses



El consenso es que todos los requisitos son relevantes para la educación. La importancia relativa de cada requisito depende en gran medida del tema y el área de aplicación del curso.

A pesar de la gran variación en el orden de los principios, el principio de “Transparencia” destaca como el menos relevante, siendo clasificado de esta manera en el 54,5% de los casos de educación. Sin embargo, las razones de esta calificación son muy dispares: algunos expertos creen que el principio de transparencia está incluido en otros requisitos, mientras que otros creen que otros requisitos utilizan conceptos más básicos que son más fáciles para los principiantes, y aún otros consideran el resto de los requisitos como más fundamentales.

2.3. Recomendaciones

El estado del arte indica que la educación actual no se enfoca en las Directrices del HLEG, sino que se abordan diferentes aspectos de la IA Fiable cuando coinciden con contenido relevante para cada curso. Este hallazgo apunta a un vacío en la ES, tanto para los educadores como para los estudiantes, en términos de conocimiento de las directrices y su contenido. En este sentido, es evidente que se necesita un doble enfoque, dirigido tanto a los educadores como a los estudiantes.

En primer lugar, **existe una necesidad de formación** y disponibilidad de materiales para que los educadores de diferentes disciplinas conozcan y aprendan sobre la IA Fiable desde la perspectiva de las directrices. Muchos entrevistados (5 de 11) destacan esta necesidad, subrayando que una barrera para incorporar educación en IA Fiable es la falta de experiencia de los educadores, ya sea por la novedad del tema, su interdisciplinariedad o por falta de tiempo para familiarizarse con estos aspectos.

En segundo lugar, muchos entrevistados mencionan que la relevancia de los requisitos incumbe a una variedad de disciplinas y temas. En este sentido, en lugar de tener un solo curso centrado en las Directrices, la mayoría consideraría más natural incluir la enseñanza de los Requisitos en cursos ya existentes. Aunque ninguno de los requisitos se considera más importante que los demás, muchos de los entrevistados destacan que cada requisito puede ser relevante dependiendo del curso y la disciplina. Por eso, **aclarar la importancia de cada requisito para diferentes temáticas** es un paso necesario para conseguir hacer explícita la inclusión de las Directrices en la educación.

En general, la evaluación de los requisitos HLEG en términos de su importancia en la educación brinda las siguientes recomendaciones para los educadores:

- R1. Incluir explícitamente** los requisitos del HLEG en **los cursos existentes** cuando sea relevante.
- R2. Acercar el contenido del curso y los requisitos** del HLEG explicitando qué requisitos se abordan en cada curso y cómo.

4. Objetivos de aprendizaje y evaluación

4.1. Objetivos de aprendizaje

Al evaluar las competencias relacionadas con la incorporación de dimensiones éticas y sociales en la educación superior en todas las disciplinas, los resultados de nuestro análisis de la literatura indican un énfasis en la doble-competencia (Kim Brown et al., 2019; Noah & Aziz, 2020; Trobec & Starcic, 2015; Zamora -Polo & Sánchez-Martín, 2019): el desarrollo de competencias técnicas junto con la capacidad de comprender y actuar de acuerdo con expectativas éticas y sociales. Aunque hay una notable ausencia de literatura sobre objetivos de aprendizaje específicos, en nuestro análisis sobresalen tres objetivos de aprendizaje para demostrar competencias sociales y éticas (ver Apéndice C):

- **Apreciación/sensibilidad ética:** Identificar y comprender las dimensiones éticas y morales de una situación.
- **Análisis ético:** deliberar sobre posibles acciones, cómo se relacionan con pautas éticas y códigos de conducta, y sus posibles consecuencias.
- **Toma de decisiones éticas / Ética aplicada:** Seleccionar e implementar un curso de acción en respuesta al razonamiento ético.

Este análisis se corresponde perfectamente con el análisis de temarios, donde se ha descubierto que los objetivos más comunes en la enseñanza de la ética de la tecnología son variaciones de las habilidades de “reconocer / debatir / razonar” (Fiesler, Garrett y Beard, 2020). Adaptando estos objetivos de aprendizaje a los requisitos del HLEG, surgen los siguientes objetivos:

LO1. Apreciación: Identificar la aplicabilidad del requisito en diferentes contextos y sus diferentes dimensiones para diferentes actores.

LO2. Análisis: Deliberar sobre posibles implementaciones del requisito, cómo se relacionan con requerimientos éticos y códigos de conducta, y sus posibles consecuencias.

LO3. Aplicación: Seleccionar e implementar técnicamente una solución en respuesta al análisis sobre el requerimiento.

La combinación de estos objetivos de aprendizaje indicaría un alto nivel de competencia de los estudiantes en tres facetas diferentes: identificación, razonamiento e implementación.

4.2. Objetivos de aprendizaje por requisito

En las preguntas sobre objetivos de aprendizaje concretos para cada requisito, las entrevistas a expertos ilustran cómo se aborda actualmente esta educación. Se pidió a los entrevistados que dieran un ejemplo de cómo presentan ciertos requisitos en su caso educativo, así como qué enseñanzas sobre cada requisito consideran más valiosas para la ES. Lo resumimos en el siguiente desglose:

- **Acción y supervisión humanas:**
Las preguntas que surgen en torno a este requisito giran alrededor de 3 temas: **control humano, marcos legales e inclusión**. En primer lugar, reconocer e implementar el nivel apropiado de control humano para garantizar la voluntad y la supervisión humana. En segundo lugar, comprender las limitaciones legales del área de aplicación (por ejemplo, en el ámbito médico). Por último, adoptar una perspectiva sólida basada en los derechos humanos y garantizar inclusividad para toda la sociedad.
- **Solidez técnica y seguridad:**
Los entrevistados piden que la educación sobre este requisito sea específica a la IA, **con los conceptos de precisión y confiabilidad** en primer plano. Además, algunos encuestados enfatizan la importancia de enseñar **como conseguir un equilibrio entre la solidez técnica y las limitaciones éticas**.
- **Rendición de cuentas:**
Las preguntas sobre este requisito giran en torno a **la auditoría y el mantenimiento de registros, los marcos legales para la responsabilidad y la demostración de la minimización de los efectos negativos**. Esta discusión incluye un enfoque sobre cómo auditar sistemas inteligentes y cómo garantizar trazabilidad para que los sistemas puedan auditarse de manera eficaz. Además, se hace un llamamiento a enseñar a los estudiantes de las áreas técnicas los aspectos legales relacionados con la responsabilidad y la demostración de la debida diligencia en la mitigación de efectos negativos.
- **Gestión de la privacidad y de los datos:**
Las preguntas que se plantean sobre la “gestión de la privacidad y de los datos” son todas en torno a los datos, específicamente **su calidad, procedencia y preservación de la privacidad**. Las preguntas educativas mencionadas son cómo recopilar y reconocer datos de calidad, mantener la privacidad y evitar sesgos en los datos y los modelos creados a partir de ellos. Además, se hace hincapié en los requisitos legales para la recopilación y el almacenamiento de datos (por ejemplo, GDPR).
- **Transparencia:**
Las preguntas educativas para la “transparencia” son muy multifacéticas, pero existe un consenso significativo entre los encuestados. Una primera pregunta implica educar a los estudiantes sobre **como reconocer los sistemas transparentes** y en qué se diferencian de los opacos. Además, muchos entrevistados relacionan la transparencia con la explicabilidad y plantean la cuestión de educar a los estudiantes para que puedan explicar las decisiones o el razonamiento del sistema y proporcionar **las habilidades para desarrollar una IA Fiable**. Una pregunta adicional relaciona la transparencia con la trazabilidad: educar a los desarrolladores sobre cómo **documentar y exponer los datos, los procesos y las decisiones tomadas en el proceso de diseño**. Finalmente, hay un último tema que gira en torno a **la transparencia del procesamiento de datos**: ser explícito sobre cómo se recopilan los datos y qué se hace con ellos.
- **Bienestar social y ambiental ; Diversidad, no discriminación y equidad:**
Las preguntas sobre estos requisitos coinciden de manera significativa, y comprenden el tema de la experiencia interdisciplinaria para abordar la dificultad de definir o cuantificar el bienestar o la diversidad. Además, algunos expertos apuntan a la idea de aprender a realizar o leer evaluaciones de impacto para evaluar los efectos de los sistemas.

En general, las preguntas planteadas sobre cada requisito, aunque específicas a cada uno, tienen dos hilos en común: reconocimiento e implementación. Por un lado, existe un fuerte empuje en enseñar a los estudiantes cómo **reconocer si se está cumpliendo un requisito**. Por otro lado, muchas preguntas giran en torno a **métodos técnicos para el desarrollo de la IA Fiable**, p. Ej. métodos de mantenimiento de registros, métodos de recopilación de datos que preservan la privacidad, métodos de explicabilidad.

Las preguntas sobre los requisitos individuales revelan dos necesidades de competencia paralelas: reconocimiento e implementación.

Además, se preguntó a los entrevistados cómo se abordaban estos requisitos actualmente en su caso de educación. Curiosamente, los conceptos de IA Fiable a menudo se introducen en cursos técnicos, como resultado de una aplicación directa de la tecnología a casos de la vida real. Esto sugiere que las preocupaciones sociales juegan un papel importante en lo que se enseña, ya que **a los educadores les resulta más fácil cerrar la brecha en los conceptos de la IA Fiable cuando pueden relacionarlos con problemas famosos**.

4.3. Evaluación

A lo largo del análisis de la literatura y las entrevistas a expertos, existe una falta notable de metodologías consistentes para evaluar competencias no-técnicas como la conciencia ética y social o la comprensión y aplicación de pautas y códigos de conducta. Los métodos de evaluación sobresalientes en el análisis de la literatura se basan principalmente en la autoevaluación, p. Ej. (Ibáñez-Carrasco, Worthington, Rourke and Hastings, 2020; Mulot-Bausière et al., 2016) – o en la evaluación de los educadores sin criterios de examinación explícitos - p. Ej. (DeSimone, 2019; Lapuzina, Lisachuk y Romanov, 2018). Por otro lado, los entrevistados informan de que no llevan a cabo una evaluación explícita de las competencias relacionadas con la IA Fiable ni incluyen este tipo de evaluación como parte de la evaluación general de proyectos de programación.

La notable falta de enfoques estructurados para la evaluación constituye una barrera tanto para evaluar las habilidades de los estudiantes como para medir la efectividad de diferentes prácticas docentes. Por lo tanto, **son muy necesarias directrices pedagógicas sobre cómo realizar dicha evaluación**.

4.4. Recomendaciones

Tanto el análisis de la literatura como las entrevistas a expertos revelan la necesidad de dos niveles diferentes de competencia. El primero corresponde a entender cómo **reconocer si se está cumpliendo un requisito y cómo**. Esta competencia corresponde al LO1 tal como hemos identificado en el análisis de la literatura: la habilidad de comprender lo que *significa* un requisito en el contexto de una determinada aplicación. De hecho, este tipo de competencia es necesaria para todos los estudiantes, no solamente como alumnos sino también como ciudadanos, ya que permite identificar y adoptar tecnología Fiable. Además, proporciona un nivel de madurez inicial en términos de comprensión de los requisitos HLEG.

La segunda competencia identificada en todos los requisitos corresponde a los **métodos técnicos para el desarrollo de IA Fiable**. Existe consenso entre los entrevistados sobre la necesidad de enseñar métodos concretos de explicabilidad, trazabilidad, recopilación de datos, evaluación de impacto, etc. Esta necesidad se relaciona estrechamente con el LO2 y LO3 tal como se presentan en la literatura: conocer las herramientas técnicas disponibles es necesario para poder tomar una decisión informada e implementarla. Dado que las técnicas relevantes varían mucho según el tema y el área del curso, es particularmente importante incluir explícitamente en el plan de estudios qué temas y métodos se abordarán (Bates et al., 2020).

Finalmente, vale la pena observar que una forma popular de introducir conceptos de IA Fiable en el aula es discutir las preocupaciones sociales actuales con las aplicaciones de la tecnología estudiada en el curso. De hecho, 6 de cada 11 entrevistados creen que sería valioso **relacionar los requisitos establecidos por las directrices con términos más prácticos**, ya sea a través de ejemplos del mundo real, participación de la industria o herramientas concretas para experimentar con diferentes conceptos en clase.

Este análisis conduce a las siguientes recomendaciones sobre objetivos de aprendizaje:

R3. Establecer objetivos de aprendizaje claros que describan el nivel de competencia que se espera del estudiante. Los objetivos de aprendizaje recomendados para cada requisito individual son:

LO1. Apreciación: Identificar la aplicabilidad del requisito en diferentes contextos y sus diferentes dimensiones para diferentes actores.

LO2. Análisis: Deliberar sobre posibles implementaciones del requisito, cómo se relacionan con las pautas éticas y códigos de conducta, y sus posibles consecuencias.

LO3. Aplicación: Seleccionar e implementar técnicamente una solución en respuesta al análisis en cuanto al requerimiento.

R4. Incluir explícitamente **metodologías de desarrollo de IA fiable** en los planes de estudio (por ejemplo, procedimientos de mantenimiento de registros, métodos de recopilación de datos que preservan la privacidad, herramientas de explicabilidad).

R5. Relacionar los requisitos con los conocimientos prácticos impartidos en el curso. Ejemplos de temas por requisito incluyen:

Acción y supervisión humanas	Control humano adecuado, marcos legales, identificación y participación de las partes interesadas
Solidez técnica y seguridad	Precisión, fiabilidad, equilibrio entre requisitos técnicos y no técnicos.
Rendición de cuentas	Auditoría, mantenimiento de registros, marcos de responsabilidad, minimización de daños
Gestión de la privacidad y de los datos	Calidad de los datos, recopilación de datos, metodologías de privacidad, cumplimiento de GDPR
Transparencia	Transparencia de procesos, explicabilidad
Bienestar social y ambiental; Diversidad, no discriminación y equidad	Colaboración interdisciplinaria, evaluación de impacto

5. Recursos para la educación

5.1. Estado del arte

De manera uniforme en todas las entrevistas, los expertos mencionan que **no utilizan ningún recurso específico relacionado con la IA Fiable**. Más bien, algunos mencionan el uso de ejemplos de actualidad, casos de estudio y literatura relevante. Esta falta de recursos se enfatiza en varias de las entrevistas con un énfasis adicional en la falta de capacitación y tiempo para familiarizarse con la IA Fiable y encontrar los recursos didácticos disponibles.

De manera similar a los resultados de la entrevista, la revisión de la literatura revela que las habilidades de razonamiento ético y moral a menudo se enseñan a través de **métodos dirigidos por los estudiantes** que se centran en fomentar la reflexión y el debate entre los estudiantes: caso de estudio (Lapuzina et al., 2018), juego de roles (Trobec & Starcic, 2015), debate (K. Brown et al., 2019), aprendizaje experiencial (Ibáñez-Carrasco et al., 2020). Esta observación se alinea con los hallazgos de otras revisiones de la literatura, que enfatizan la prevalencia de juegos, juegos de roles y casos estudio en la educación en ingeniería y ciencias de la computación (Hoffmann & Cross, 2021). Aunque está muy extendido, sin embargo, existe disensión en la literatura, donde algunos abogan por una capacitación más formal en el tema, p. Ej. filosofía moral, en contraste con las actividades dirigidas por estudiantes (Aközer & Aközer, 2017).

Las estrategias de enseñanza más utilizadas para enseñar aspectos de IA Fiable influyen en el tipo de recursos disponibles actualmente. De hecho, diversos organismos han desarrollado **casos de estudio** disponibles abiertamente sobre la ética de la IA, como la Princeton University¹, Santa Clara University², University of Washington³ and UNESCO⁴.

5.2. Necesidades

Cuando se preguntó a los entrevistados sobre qué tipo de recursos serían útiles para integrar la IA Fiable en la educación superior, surgieron varios temas. En primer lugar, 5 de los 11 entrevistados coincidieron en preguntar por casos de uso. Curiosamente, hubo un consenso significativo sobre el tipo de **casos de uso** que se consideran necesarios: deben ser realistas e implementables. De hecho, el uso de casos reales traídos directamente de la industria que imitan situaciones en las que los estudiantes graduados pueden encontrarse se considera importante para la utilidad de estos escenarios. En contraste con la literatura, donde los casos de uso se usan a menudo para la reflexión y el debate, varios entrevistados sugirieron que los casos de uso deberían usarse para la exploración práctica, donde pueden implementar y “jugar” con diferentes soluciones.

Otra mención frecuente es la necesidad de material para ayudar en la evaluación, es decir, **ejercicios o tareas** con una guía de calificación que se pueda utilizar directamente para evaluar a los estudiantes. De hecho, varios entrevistados compartieron la dificultad de evaluar el conocimiento de conceptos abstractos. Un comentario final compartido entre los entrevistados fue la necesidad de **recursos para los profesores**. Esta necesidad se enfatizó particularmente en relación con la interdisciplinariedad necesaria para incumplir todos los aspectos de los requisitos. Por lo tanto, los entrevistados solicitaron recursos que podrían ayudarlos a conocer los diferentes requisitos y comprender cómo encajarlos en su plan de estudios. Estos incluyeron material introductorio sobre las Directrices, programas de estudios o directrices de evaluación que los educadores podrían utilizar para informar el diseño de sus propios cursos. Además, para asegurar una visión múltiple de las Directrices, se sugirió proporcionar orientación sobre qué tipo de expertos podrían proporcionar información sobre los diferentes temas, o proporcionar conferencias grabadas sobre diferentes aspectos de los Requisitos por parte de académicos en diferentes campos y la industria.

Este análisis conduce a las siguientes **necesidades identificadas**:

N1. Hay necesidad de **casos de uso reales** en los que los estudiantes puedan desarrollar conocimientos prácticos.

N2. Es necesario desarrollar **ejercicios y tareas** relacionados con cada requisito, junto con esquemas de calificación para su evaluación.

N3. Existe la necesidad de **material de capacitación para los educadores** para ser informados sobre como añadir material sobre los Requisitos en sus cursos. Este material puede incluir material introductorio sobre las Directrices, programas de estudios, pautas de evaluación, guías sobre charlas invitadas o clases grabadas por expertos en diferentes disciplinas.

6. Estrategias, incentivos, riesgos y oportunidades

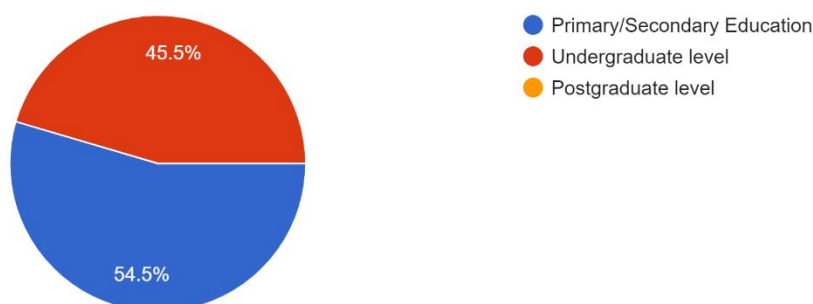
6.1. Estrategias e incentivos

De manera uniforme, 6 de cada 11 entrevistados afirman que **no conocen ninguna estrategia política específica para incluir aspectos de la IA Fiable en la educación**, ya sea a nivel de su institución o a nivel nacional. Simultáneamente, 5 de los 11 entrevistados mencionan que el tema de la **IA fiable está ganando importancia en su organización** y que están considerando activamente cómo incluirlo en sus programas. Este desajuste indica que, aunque se está introduciendo la IA fiable en la ES, el esfuerzo es impulsado principalmente por los propios educadores y no por las estrategias organizativas o nacionales. Este enfoque presenta el riesgo de un desajuste en las competencias entre los programas de diferentes instituciones de educación superior, ya que la introducción de la IA Fiable en los programas educativos se lleva a cabo de forma independiente en lugar de dentro de una estrategia coordinada.

Cuando se les preguntó cuándo debería introducirse la IA Fiable en la educación, los encuestados se dividieron entre el nivel escolar (educación primaria o secundaria) y el nivel de pregrado. En cualquier caso, esto indica un consenso en que estos aspectos no deben relegarse a títulos de especialización a nivel de posgrado, sino que deben ser introducidos en un nivel competencial anterior más general.

Summary: In which education cycle should trustworthy AI education need to start?

11 responses



En cuanto a las necesidades de políticas y los incentivos para impulsar la introducción de la IA Fiable en la ES, los entrevistados comentaron una gran variedad de sugerencias. Un gran punto de consenso (5 de 11 entrevistados) es **la necesidad de invertir en experiencia en IA Fiable para que los educadores** estén equipados para enseñar estos temas: esto puede tomar la forma de invertir en capacitación multidisciplinaria o impulsar la contratación de expertos en aspectos de IA Fiable para participar en la educación. Esta idea se alinea con los comentarios de los entrevistados sobre la falta de tiempo para familiarizarse con los temas a fin de estar preparados para introducirlos en el aula.

Varios entrevistados (3 de 11) mencionan la importancia de permitir flexibilidad en la estructura de titulaciones para permitir la inclusión de temas interdisciplinarios más amplios. Mencionan que las políticas actuales limitan estrictamente los objetivos de aprendizaje de los diferentes programas y dejan poco espacio para la colaboración e interdiscipliniedad entre departamentos. Por el contrario, la IA Fiable se considera un tema que se beneficiaría de la exposición de los estudiantes a diferentes disciplinas, lo que exige **incentivos políticos que fomenten el aprendizaje interdisciplinario**. Estos pensamientos se alinean con los recientes llamados a una educación transversal que permita la interdiscipliniedad al considerar la ética en la tecnología (Raji, Scheuerman y Amironesei, 2021).

Otro factor mencionado es la participación de la industria, tanto en la provisión de tópicos de casos de uso como en su participación en la empleabilidad de los estudiantes con conocimientos de IA Fiable. Además, también se mencionó el papel de la financiación académica, pidiendo a las agencias de financiación que consideren los aspectos de la IA Fiable antes de conceder la financiación.

6.2. Riesgos y oportunidades

Cuando se les preguntó a los entrevistados acerca de los riesgos, hubo un consenso significativo (6 de 11) al mencionar que existe el riesgo de introducir la IA Fiable en la ES antes de que las instituciones puedan prepararse, es decir, antes de que haya suficiente experiencia en el tema para poder instruir de manera competente. Además, otro riesgo relevante mencionado es que es importante que los estudiantes de todas las disciplinas puedan aprender sobre la IA Fiable. Si bien parece que está comenzando a ser un enfoque en STEM, hubo algunas preocupaciones de que otras disciplinas podrían no estar expuestas al tema en HE.

En términos de oportunidades, muchos entrevistados (5 de 11) enfatizaron que los aspectos de la IA Fiable son importantes para los estudiantes no solo como futuros profesionales, sino también como ciudadanos. En este sentido, enfatizaron los beneficios de formar una generación de profesionales que posea conocimientos interdisciplinarios y pueda comunicarse con profesionales de otras disciplinas en los términos de la IA Fiable.

6.3. Recomendaciones

En vista del consenso expresado en las entrevistas, identificamos las siguientes recomendaciones de políticas para incentivar la introducción de IA Fiable en los planes de estudios de ES:

- P1.** Coordinar la introducción de la IA fiable en los planes de estudio a **través de estrategias nacionales de educación**, asegurando una adopción uniforme.
- P2.** Incentivar a las instituciones de educación superior para que **obtengan la experiencia relevante** necesaria para instruir la IA fiable, tanto invirtiendo recursos en capacitación para educadores como contratando expertos.
- P3.** Incentivar la **colaboración interdisciplinaria en educación** valorándola en el plan de estudios e introduciendo créditos para ella.

7. Referencias

- Aközer, M., & Aközer, E. (2017). Ethics Teaching in Higher Education for Principled Reasoning: A Gateway for Reconciling Scientific Practice with Ethical Deliberation. *Science and Engineering Ethics*, 23(3), 825–860. <https://doi.org/10.1007/s11948-016-9813-y>
- Bates, J., Cameron, D., Checco, A., Clough, P., Hopfgartner, F., Mazumdar, S., ... de la Vega de León, A. (2020). Integrating fate/critical data studies into data science curricula: Where are we going and how do we get there? In *FAT* 2020 - Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency* (pp. 425–435). <https://doi.org/10.1145/3351095.3372832>
- Brown, K., Connelly, S., Lovelock, B., Mainvil, L., Mather, D., Roberts, H., ... Shephard, K. (2019). Do we teach our students to share and to care? *Research in Post-Compulsory Education*, 24(4), 462–481. <https://doi.org/10.1080/13596748.2019.1654693>
- Brown, Kim, Connelly, S., Lovelock, B., Mainvil, L., Mather, D., Roberts, H., ... Shephard, K. (2019). Do we teach our students to share and to care? *Research in Post-Compulsory Education*, 24(4), 462–481. <https://doi.org/10.1080/13596748.2019.1654693>
- DeSimone, B. B. (2019). Curriculum Redesign to Build the Moral Courage Values of Accelerated Bachelor’s Degree Nursing Students. *SAGE Open Nursing*, 5. <https://doi.org/10.1177/2377960819827086>
- EU. (2020). WHITE PAPER On Artificial Intelligence - A European approach to excellence and trust).
- EUCommission. (2018). ETHICS GUIDELINES FOR TRUSTWORTHY AI. High-Level Expert Group on Artificial Intelligence, (December). Retrieved from <https://ec.europa.eu/digital-single-market/en/news/ethics-guidelines-trustworthy-ai>
- Fiesler, C., Garrett, N., & Beard, N. (2020). What Do We Teach When We Teach Tech Ethics? A Syllabi Analysis. <https://doi.org/10.1145/3328778.3366825>
- Hoffmann, A. L., & Cross, K. A. (2021). Teaching Data Ethics: Foundations and Possibilities from Engineering and Computer Science Ethics Education. University of Washington Research Works Archive. APA. Retrieved from <http://hdl.handle.net/1773/46921>.
- Ibáñez-Carrasco, F., Worthington, C., Rourke, S., & Hastings, C. (2020). Universities without walls: A blended delivery approach to training the next generation of hiv researchers in canada. *International Journal of Environmental Research and Public Health*, 17(12), 1–12. <https://doi.org/10.3390/ijerph17124265>
- Lapuzina, O., Lisachuk, L., & Romanov, Y. (2018). Professional Ethics as an Important Part of Engineer Training in Technical Higher Education Institutions. *The New Educational Review*, 54(4), 110–121. <https://doi.org/10.15804/tner.2018.54.4.09>
- Mulot-Bausière, M., Gallé-Gaudin, C., Montaz, L., Burucoa, B., Mallet, D., & Denis-Delpierre, N. (2016). Formation des internes en médecine de la douleur et médecine palliative : bilan et suggestions des étudiants. *Médecine Palliative : Soins de Support - Accompagnement - Éthique*, 15(3), 143–150. <https://doi.org/10.1016/j.medpal.2015.12.002>
- Noah, J. B., & Aziz, A. B. A. (2020). A case study on the development of soft skills among TESL graduates in a University. *Universal Journal of Educational Research*, 8(10), 4610–4617. <https://doi.org/10.13189/ujer.2020.081029>
- Raji, I. D., Scheuerman, M. K., & Amironesei, R. (2021). “you can’t sit with us”: Exclusionary pedagogy in AI ethics education. In *FACCT 2021 - Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency* (pp. 515–525). New York, NY, USA: ACM. <https://doi.org/10.1145/3442188.3445914>
- Trobec, I., & Starcic, A. I. (2015). Developing nursing ethical competences online versus in the traditional classroom. *Nursing Ethics*, 22(3), 352–366. <https://doi.org/10.1177/0969733014533241>
- Turiel, E. (2001). *The Culture of Morality. The Culture of Morality*. Cambridge University Press. <https://doi.org/10.1017/cbo9780511613500>
- Zamora-Polo, F., & Sánchez-Martín, J. (2019). Teaching for a better world. Sustainability and Sustainable Development Goals in the construction of a change-maker university. *Sustainability (Switzerland)*, 11(15). <https://doi.org/10.3390/su11154224>

Appendix A.

Interview profiles

Appendix A. Interview profiles

A total of 11 interviewees were selected for their involvement in HE, whether through governance, program management or teaching.

This appendix contains an anonymised breakdown of their profiles and the use cases they brought to ground the interview on a specific topic.

Table 1: Interviewees' Geographical location as given by affiliation (note that some interviewees have several affiliations).

Organisational affiliation	Number of interviewees
Croatia	1
Ireland	3
Netherlands	2
Spain	3
Sweden	2
EU	1

Table 2: Interviewees' profile.

Position	Number of interviewees
Program coordinator	4
Professor	4
Lecturer	4
Policy maker	1

Table 3: Disciplines of the use cases selected by the interviewees.

Discipline of the use case	Number of interviewees
Computer Science	6
Medicine	1
Law	1
Social Sciences	1
Media Studies	1
Multidiscipline	1

Appendix B.

Interview Protocol

Appendix B. 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

Copyright © 2021 by Andrea Aler Tubella, Juan Carlos Nieves

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:

Appendix C.

Literature Review

Appendix C. Literature Review

1. Methodology

The goal of the systematic review is to analyse the relevant literature in order to answer the following questions:

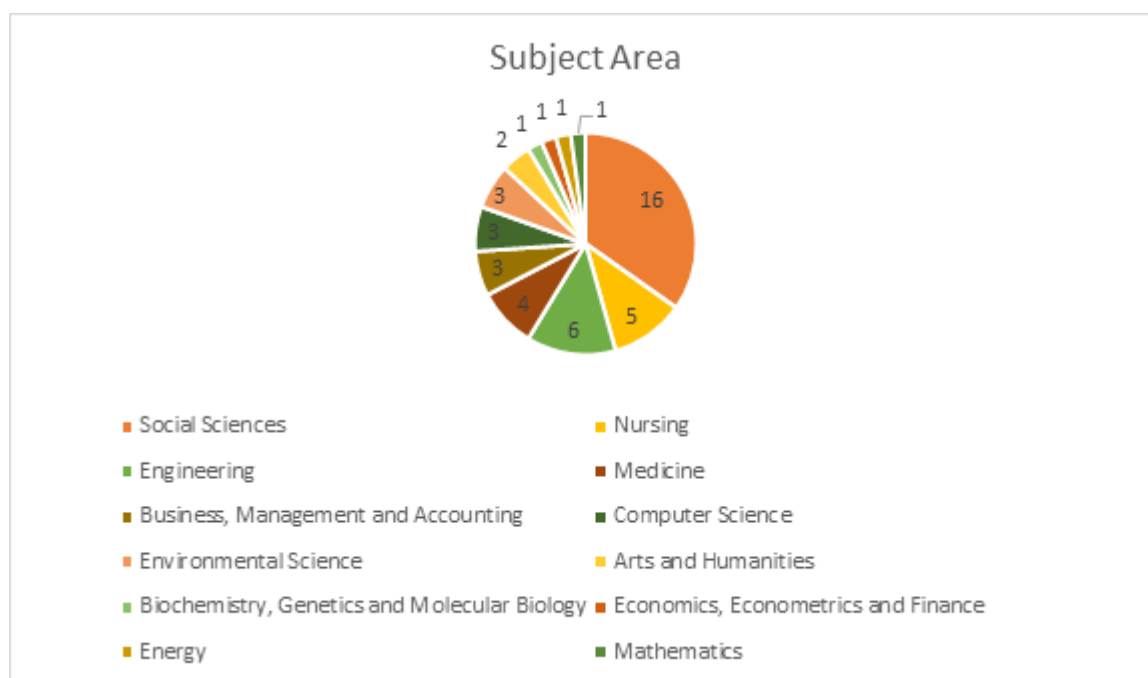
1. What competences and learning objectives are identified when teaching ethical aspects in HE?
2. How are these competences taught and evaluated?

We conducted the literature search on Scopus, to obtain results in a variety of disciplines. We used the following search terms, to be found in title, abstract or key:

ethics AND teaching AND "higher education" AND (competence OR competency OR skills)

We limited our search to publications from 2015 onwards and retrieved a total of 50 publications on 09/02/2021 at 15:13. We focused on individual research output, so excluded 1 book, 1 editorial, 1 extended abstract and 6 review articles (either paper reviews or curricula reviews). 4 papers were not accessible at the time of analysis. Finally, we manually excluded 8 papers whose abstract did not mention anything related to teaching skills related to ethics and 5 papers were removed upon further reading for lack of relevance (either not focused on HE or not focused on teaching aspects related to ethics). The final output is **24 papers** which we analysed.

The papers analyse cover a wide variety of subject areas. Based on author affiliation 17 countries are covered, as well as 12 subject areas as indexed by Scopus.



2. Overview

Publications identifying specific competences are few, although many mention that explicitly identifying competences is a pressing educational need. On the other hand, most publications propose teaching methods, with a strong focus on learning with a social component of debate and participation between students. For this reason, non-traditional teaching methods like case-studies and role-playing seem to be often proposed and studied. Much of the literature consists on exposing or evaluating how certain teaching practices were incorporated to teach ethics in specific

degrees or modules. Much of the literature emphasises the importance of incorporating different dimensions of ethics into their education. In particular, professional ethics as it refers to codes of conduct is mentioned often.

When assessing competencies related to incorporating ethical and social dimensions into Higher Education in all disciplines, results of our literature review indicate an emphasis on dual competence (Kim Brown et al., 2019; Noah & Aziz, 2020; Trobec & Starcic, 2015; Zamora-Polo & Sánchez-Martín, 2019): developing technical competence alongside the ability to understand and act according to ethical and social expectations. Although discussion on *specific* learning outcomes is notably absent, three learning goals are prevalent for demonstrating mastery of social and ethical competencies:

- Ethical appreciation/sensitivity: Identifying and understanding the ethical and moral dimensions of a situation.
- Ethical analysis: Deliberating about actions, how they relate to ethical guidelines and codes of conduct, and their possible consequences.
- Ethical decision-making/Applied ethics: Selecting and implementing a course of action in response to ethical reasoning.

Some examples of how these competences are identified can be found in the following table:

Competence	Quotes
Ethical appreciation	"When solving an ethical situation, the awareness and respect of ethical principles are not enough, the individual's role and ability to perceive the ethical dimension are also important." (Trobec & Starcic, 2015) "using technology to assess sociopolitical, ethical, and historical issues related to nursing practice" (DeSimone, 2019) "capacity to universalize principles of action, as a prerequisite for principled reasoning" (Aközer & Aközer, 2017)
Ethical analysis	"the student analyzes the pro and con viewpoints of an ethical question" (DeSimone, 2019) "reflective awareness of questions about both the "good life" and "right action" as objects of principled reasoning." (Aközer & Aközer, 2017)
Ethical decision-making	"tutors also perceived technical skills and the application skills of problem-solving and critical thinking, as important" (Dean et al., 2020) "Practice in the art of ethical decision making is the best way for ensuring that the graduates of educational leadership programs have the ethical muscle to make decisions that are moral" (Jones et al., 2020) "making and defending ethical decisions related to health-care issues" (DeSimone, 2019)

These findings squarely align with syllabi analysis, where it has been found that the most common sought outcomes for teaching Tech Ethics are variations on "recognize/critique/reason" (Fiesler, Garrett, & Beard, 2020).

The literature review reveals that ethical and moral reasoning skills are often taught through student-led methods focused on encouraging reflection and debate amongst students: case studies (Lapuzina et al., 2018), role playing (Trobec & Starcic, 2015), debate (K. Brown et al., 2019), experiential learning (Ibáñez-Carrasco et al., 2020), etc. This observation aligns with findings from other literature reviews, which emphasise the prevalence of games, role playing and case studies in Engineering and Computer Science Education (Hoffmann & Cross, 2021). Although wide-spread, there is however dissent in the literature, where some advocate for more formal training in e.g. moral philosophy, in contrast to student-led activities (Aközer & Aközer, 2017).

Across the literature there is a noticeable lack of consistent methodologies for assessing soft competences such as ethical and social awareness or understanding and application of guidelines and codes of conduct. The assessment

methods uncovered in the literature review mostly rely on self-assessment – e.g. (Ibáñez-Carrasco, Worthington, Rourke, & Hastings, 2020; Mulot-Bausière et al., 2016) – or experts' perception of student's knowledge without explicit grading criteria – e.g. (DeSimone, 2019; Lapuzina, Lisachuk, & Romanov, 2018).

3. References

- Mackenzie, M. L. (2015). *Educating ethical leaders for the information society: Adopting babies from business*. *Advances in Librarianship* (Vol. 39). <https://doi.org/10.1108/S0065-283020150000039010>
- Miñano, R., Aller, C. F., Anguera, Á., & Portillo, E. (2015). Introducing ethical, social and environmental issues in ICT engineering degrees. *Journal of Technology and Science Education*, 5(4), 272–285. <https://doi.org/10.3926/jotse.203>
- Trobec, I., & Starcic, A. I. (2015). Developing nursing ethical competences online versus in the traditional classroom. *Nursing Ethics*, 22(3), 352–366. <https://doi.org/10.1177/0969733014533241>
- Biasutti, M., De Baz, T., & Alshawa, H. (2016). Assessing the infusion of sustainability principles into university curricula. *Journal of Teacher Education for Sustainability*, 18(2), 21–40. <https://doi.org/10.1515/jtes-2016-0012>
- Mulot-Bausière, M., Gallé-Gaudin, C., Montaz, L., Burucoa, B., Mallet, D., & Denis-Delpierre, N. (2016). Residents' training in pain and palliative medicine: Students' assessment | Formation des internes en médecine de la douleur et médecine palliative: Bilan et suggestions des étudiants. *Medecine Palliative*, 15(3), 143–150. <https://doi.org/10.1016/j.medpal.2015.12.002>
- Gómez, V., & Royo, P. (2015). Ethical self-discovery and deliberation: Towards a teaching model of ethics under the competences model | Autodescubrimiento ético y deliberación: Hacia un modelo de enseñanza de la ética en el modelo por competencias. *Estudios Pedagógicos*, 41(2), 345–358. <https://doi.org/10.4067/s0718-07052015000200020>
- Sánchez-Martín, J., Zamora-Polo, F., Moreno-Losada, J., & Parejo-Ayuso, J. P. (2017). Innovative education tools for developing ethical skills in university science lessons. The case of the moral cross dilemma. *Ramon Llull Journal of Applied Ethics*, (8), 225–245.
- Galanina, E., Dulzon, A., & Schwab, A. (2015). Forming engineers' sociocultural competence: Engineering ethics at tomsk polytechnic university. In *IOP Conference Series: Materials Science and Engineering* (Vol. 93). <https://doi.org/10.1088/1757-899X/93/1/012078>
- Gokdas, I., & Torun, F. (2017). Examining the impact of instructional technology and material design courses on technopedagogical education competency acquisition according to different variables. *Kuram ve Uygulamada Egitim Bilimleri*, 17(5), 1733–1758. <https://doi.org/10.12738/estp.2017.5.0322>
- DeSimone, B. B. (2019). Curriculum Redesign to Build the Moral Courage Values of Accelerated Bachelor's Degree Nursing Students. *SAGE Open Nursing*, 5. <https://doi.org/10.1177/2377960819827086>
- Fernandez, F. J. I., & Martinez-Canton, A. (2019). Ethics and Transversal Citizenship in the Teaching of Science and Engineering. In *2019 9th IEEE Integrated STEM Education Conference, ISEC 2019* (pp. 104–110). <https://doi.org/10.1109/ISECon.2019.8881965>
- Lapuzina, O., Romanov, Y., & Lisachuk, L. (2018). Professional ethics as an important part of engineer training in technical higher education institutions. *New Educational Review*, 54(4), 110–121. <https://doi.org/10.15804/ner.2018.54.4.09>
- Rameli, M. R. M., Bunyamin, M. A. H., Siang, T. J., Hassan, Z., Mokhtar, M., Ahmad, J., & Jambari, H. (2018). Item analysis on the effects of study visit programme in cultivating students' soft skills: A case study. *International Journal of Engineering and Technology(UAE)*, 7(2), 117–120. <https://doi.org/10.14419/ijet.v7i2.10.10968>
- Riedel, A., & Giese, C. (2019). Development of ethical competence in (future) nursing programs—reification and steps as a foundation for curricular development | Ethikkompetenzentwicklung in der (zukünftigen) pflegeberuflichen Qualifizierung – Konkretion und Stufung als Grundlegung für . *Ethik in Der Medizin*, 31(1), 61–79. <https://doi.org/10.1007/s00481-018-00515-0>

- Aközer, M., & Aközer, E. (2017). Ethics Teaching in Higher Education for Principled Reasoning: A Gateway for Reconciling Scientific Practice with Ethical Deliberation. *Science and Engineering Ethics*, 23(3), 825–860. <https://doi.org/10.1007/s11948-016-9813-y>
- Oliphant, T., & Brundin, M. R. (2019). Conflicting values: An exploration of the tensions between learning analytics and academic librarianship. *Library Trends*, 68(1), 5–23. <https://doi.org/10.1353/lib.2019.0028>
- Brown, K., Connelly, S., Lovelock, B., Mainvil, L., Mather, D., Roberts, H., ... Shephard, K. (2019). Do we teach our students to share and to care? *Research in Post-Compulsory Education*, 24(4), 462–481. <https://doi.org/10.1080/13596748.2019.1654693>
- Jones, K. D., Ransom, H., & Chambers, C. R. (2020). Teaching Ethics in Educational Leadership Using the Values–Issues–Action (VIA) Model. *Journal of Research on Leadership Education*, 15(2), 150–163. <https://doi.org/10.1177/1942775119838297>
- Bates, J., Cameron, D., Checco, A., Clough, P., Hopfgartner, F., Mazumdar, S., ... de la Vega de León, A. (2020). Integrating fate/critical data studies into data science curricula: Where are we going and how do we get there? In *FAT* 2020 - Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency* (pp. 425–435). <https://doi.org/10.1145/3351095.3372832>
- Dean, B. A., Perkiss, S., Simic Misic, M., & Luzia, K. (2020). Transforming accounting curricula to enhance integrative learning. *Accounting and Finance*, 60(3), 2301–2338. <https://doi.org/10.1111/acfi.12363>
- Zamora-Polo, F., & Sánchez-Martín, J. (2019). Teaching for a better world. Sustainability and Sustainable Development Goals in the construction of a change-maker university. *Sustainability (Switzerland)*, 11(15). <https://doi.org/10.3390/su11154224>
- Ibáñez-Carrasco, F., Worthington, C., Rourke, S., & Hastings, C. (2020). Universities without walls: A blended delivery approach to training the next generation of hiv researchers in canada. *International Journal of Environmental Research and Public Health*, 17(12), 1–12. <https://doi.org/10.3390/ijerph17124265>
- Sahin, Y. G., & Celikkan, U. (2020). Information technology asymmetry and gaps between higher education institutions and industry. *Journal of Information Technology Education: Research*, 19, 339–365. <https://doi.org/10.28945/4553>
- Noah, J. B., & Aziz, A. B. A. (2020). A case study on the development of soft skills among TESL graduates in a University. *Universal Journal of Educational Research*, 8(10), 4610–4617. <https://doi.org/10.13189/ujer.2020.081029>